Open-Source Frameworks for Scalable Data Quality Orchestration in Clinical Data Pipelines

Pavan Kumar Reddy Sakam, Senior Clinical Programmer, AstraZeneca, Bengaluru, India
Vinodh Kumar Jestadi, Associate Director – Clinical Programming, AstraZeneca, Bengaluru, India

ABSTRACT

High-quality data is essential for trustworthy decision-making in clinical research. Clinical trial data today flows through increasingly complex pipelines, drawing inputs from multiple systems such as electronic data capture platforms, laboratories, and real-world data sources. While data volumes and velocity have increased, data quality practices often remain fragmented, manual, and tightly coupled to legacy systems. This disconnect results in rework, delays, inconsistent validation outcomes, and challenges during audits or inspections.

This paper presents a practical, open-source–driven approach to orchestrating data quality checks across clinical data pipelines. By separating validation logic from underlying systems, using metadata to drive rule execution, and automating execution through a controlled orchestration layer, data quality checks can be executed consistently, transparently, and at scale. A case study demonstrates how demographic and adverse event data can be validated dynamically using configurable rules and automated scheduling. The approach improves operational efficiency, strengthens governance, and supports inspection readiness, while remaining accessible to both technical and non-technical stakeholders.

1. Introduction

Clinical research relies on the integrity, consistency, and traceability of data to support critical decisions across study conduct, safety evaluation, and regulatory submissions. As clinical trials evolve, data pipelines have become increasingly complex, integrating data from multiple source systems such as electronic data capture platforms, laboratory systems, and external vendors. This complexity has increased both the volume of data and the operational burden associated with ensuring data quality.

Despite advances in data processing technologies, data quality practices often remain fragmented and heavily manual. Validation logic is frequently embedded within individual programs or tools, making it difficult to reuse, scale, or consistently govern across studies. As a result, teams experience rework, delays, inconsistent validation outcomes, and challenges when responding to audit or inspection queries.

There is a growing need for a data quality approach that is not only technically robust, but also transparent, repeatable, and understandable across a wide range of stakeholders, including data managers, quality leads, medical reviewers, and programmers. In regulated environments, this need extends beyond detecting issues to demonstrating how and when data quality checks were executed, and why outcomes can be trusted.

This paper presents a practical, open-source–driven framework for orchestrating data quality checks across clinical data pipelines. The approach focuses on separating validation logic from execution, driving rules through metadata, and automating controlled execution through an orchestration layer. By doing so, data quality checks can be executed consistently, transparently, and on a scale.

Scope clarification:

The framework described in this paper is currently implemented for source clinical data, such as data originating from EDC and upstream systems, prior to SDTM and ADaM transformations. While SDTM and ADaM validation remain critical downstream activities, they are intentionally out of scope for the current implementation. However, the architectural principles described—metadata-driven rules, controlled orchestration, and audit-ready outputs—are designed to be extensible and can be scaled to support SDTM and ADaM datasets in the future.

A proof-of-concept case study demonstrates how source-level demographic and adverse event data can be validated dynamically using configurable rules and automated execution. The objective is not to replace existing validation standards, but to establish a governed, scalable foundation for data quality execution that improves operational efficiency, strengthens inspection readiness, and supports collaboration across technical and non-technical stakeholders.

2. Clinical Data Pipelines: A Business and Quality Perspective

Clinical data pipelines form the backbone of modern clinical research operations. They represent the sequence of processes through which data moves from source systems to downstream analysis and reporting. While the underlying technologies and platforms may vary across organizations, the business and quality expectations from these pipelines remain consistent.

At a high level, stakeholders expect that data flowing through these pipelines is reliable, traceable, and fit for purpose at every stage of the study lifecycle. Achieving this consistently requires alignment between operational execution, quality controls, and governance.

2.1 Business Expectations from Clinical Data Pipelines

From a business perspective, clinical data pipelines are expected to support timely decision-making while minimizing operational risk and rework. Data must be available when needed, with sufficient confidence to support study oversight, safety evaluations, and submission timelines.

Key business expectations include:

- **Timeliness**: Data should move efficiently from source systems to downstream consumers without unnecessary delays caused by manual validation or rework.
- **Consistency**: Validation outcomes should be predictable and repeatable across studies, domains, and data refresh cycles.
- **Operational efficiency**: Data quality activities should reduce manual effort rather than introduce additional operational burden.
- **Transparency**: Stakeholders should be able to easily understand what checks were executed, when they ran, and what the outcomes were.
- **Scalability**: As study portfolios grow and data sources diversify, data quality processes should scale without requiring proportional increases in effort.

When these expectations are not met, teams often experience downstream consequences such as delayed reviews, duplicated effort across functions, and increased pressure during critical milestones. In this context, data quality is not merely a technical concern—it directly impacts delivery timelines, resource utilization, and organizational confidence in the data.

2.2 Data Quality Expectations in Regulated Environments

In regulated clinical research environments, data quality expectations extend beyond detecting discrepancies. Organizations must also demonstrate that data quality processes are controlled, traceable, and consistently applied.

Regulatory and quality-focused expectations typically include:

- **Repeatability**: Validation checks should produce consistent outcomes when run under the same conditions.
- Traceability: Each validation result should be traceable to a defined rule, execution time, and dataset version.
- **Explainability**: Outcomes should be interpretable not only by programmers, but also by data managers, quality reviewers, and auditors.
- **Governance**: Changes to validation logic should be controlled, versioned, and auditable.
- **Inspection readiness**: Evidence of data quality execution should be readily available to support audits and inspections.

Traditional approaches often struggle to meet these expectations when validation logic is embedded within individual programs or executed manually. In such cases, demonstrating consistency, control, and traceability becomes difficult, particularly when responding to inspection questions that span multiple studies or time points.

Bridging the gap between business efficiency and regulatory rigor requires a data quality approach that is both operationally practical and governance focused. This sets the foundation for adopting orchestration-based data quality execution, which is explored in the subsequent sections

3. Why Data Quality Remains a Bottleneck

Data quality challenges in clinical research persist not because validation checks are absent, but because of how those checks are implemented, executed, and governed. In many organizations, data quality practices have evolved organically over time, resulting in approaches that are effective in isolation but difficult to scale or standardize.

A common limitation is that validation logic is tightly embedded within individual programs, scripts, or systems. These checks are often written by different programmers using varying assumptions and styles, even when validating similar concepts across studies or domains. As a result, the same rule may be implemented differently in different contexts, leading to inconsistent outcomes and reduced confidence in validation results.

Manual execution remains another significant bottleneck. When data changes—due to refreshes, corrections, or upstream updates—validation checks often need to be re-run manually. This increases operational effort and introduces the risk of missed or skipped checks, particularly during high-pressure timelines. Visibility into whether checks have run, when they ran, and what failed is often limited to individual logs or emails, making cross-functional oversight difficult.

From a governance perspective, traditional approaches struggle to provide traceability and explainability. When validation outcomes are questioned during internal reviews or inspections, teams must often reconstruct execution details by reviewing code, rerunning programs, or manually compiling evidence. This reactive effort consumes time and increases the risk of inconsistent explanations.

These challenges are amplified in environments with multiple data sources, frequent data refreshes, and parallel study timelines. As portfolios scale, the operational cost of maintaining manual or loosely governed validation processes grows disproportionately.

Ultimately, the bottleneck is not the complexity of the data itself, but the lack of a coordinated mechanism to execute data quality checks in a controlled, repeatable, and transparent manner. Addressing this gap requires moving beyond isolated validation scripts toward an orchestrated approach that treats data quality as an integral part of the data pipeline lifecycle rather than a standalone activity.

4. Rethinking Data Quality Through Orchestration

To overcome the limitations of traditional data quality practices, data quality must be treated as a coordinated and governed process rather than a collection of isolated checks. This requires shifting focus from how individual validations are written to how and when validations are executed within the clinical data pipeline.

Data quality orchestration provides this shift by introducing a controlled mechanism that determines the execution of validation checks across datasets, studies, and data refresh cycles. Instead of relying on manual triggers or ad hoc execution, orchestration embeds data quality directly into the lifecycle of the data pipeline.

At the core of this approach are several guiding principles:

- **Separation of concerns**: Validation rules are defined independently of the systems or programs that execute them. This reduces coupling, improves reuse, and simplifies governance.
- **Metadata-driven configuration**: Validation logic, parameters, and thresholds are expressed through structured metadata rather than hard-coded logic. This enables consistent application of rules and simplifies updates.
- **Automated and controlled execution**: Validation checks are executed automatically at predefined points in the data lifecycle, reducing manual intervention and the risk of missed checks.
- **Standardized outputs**: Results are produced in a consistent, interpretable format that supports both operational review and audit readiness.

By orchestrating data quality execution, organizations gain predictability and transparency. Stakeholders can clearly understand when checks ran, what data was evaluated, and what outcomes were produced. This also enables validation activities to scale as data volumes and study portfolios increase, without a corresponding increase in manual effort.

Importantly, orchestration does not replace existing validation standards or domain expertise. Instead, it provides a framework that ensures those standards are applied consistently and traceably across the data pipeline. Validation logic remains driven by clinical and regulatory requirements, while orchestration ensures disciplined execution.

This orchestration-based approach lays the foundation for a flexible and extensible data quality framework. It supports current source-level validation needs while enabling future expansion to downstream datasets as organizational maturity and requirements evolve. The following section describes the conceptual architecture that operationalizes these principles.

5. Conceptual Architecture of the Orchestrated Data Quality Framework

The orchestrated data quality framework is designed as a layered architecture that supports flexibility, scalability, and governance while remaining understandable to both technical and non-technical stakeholders. Each layer has a distinct responsibility, enabling separation of concerns and controlled execution of data quality checks.

The current implementation focuses on source clinical data prior to SDTM and ADaM transformations. However, the architectural principles are intentionally designed to support future extension to downstream datasets.

5.1 Data Sources and Scope of Application

The framework operates on clinical data originating from upstream source systems, such as electronic data capture platforms, laboratory systems, and external vendor feeds. These datasets represent the earliest point at which data quality issues can be detected and corrected with minimal downstream impact.

By applying validation checks at the source data level, organizations can:

- Identify issues closer to data entry or ingestion
- Reduce propagation of errors to downstream transformations
- Support earlier review by data management and medical teams

While SDTM and ADaM datasets remain critical for submission and analysis, they are outside the scope of the current implementation. The focus on source data allows for a controlled proof-of-concept while establishing a foundation that can be expanded as governance and operational readiness evolve.

5.2 Metadata-Driven Rule Definitions

Validation rules within the framework are defined using structured metadata rather than being embedded directly in code. Metadata specifies what to check, which datasets or variables are involved, and how outcomes should be evaluated.

This approach enables:

- Consistent interpretation of rules across studies and domains
- Simplified maintenance and controlled updates to validation logic
- Clear traceability between rule definitions and execution outcomes

Because rules are metadata-driven, updates to validation requirements can be implemented without redeploying code, reducing operational risk and turnaround time.

5.3 Execution and Orchestration Layers

The execution layer is responsible for applying validation rules to the target datasets in a consistent and repeatable manner. It evaluates data against the metadata-defined rules and produces standardized results.

Above this, the orchestration layer controls when and under what conditions validations are executed. It manages scheduling, dependencies, and execution order, ensuring that checks are run at appropriate points in the data lifecycle. This controlled execution reduces reliance on manual triggers and provides predictable validation behavior across data refresh cycles.

Together, these layers ensure that data quality checks are not only accurate, but also reliably executed and governed.

5.4 Audit-Ready Outputs

A key outcome of the framework is the production of standardized, audit-ready outputs. Each validation run produces clear PASS or FAIL results, along with contextual information such as execution time and rule reference.

These outputs enable:

- Rapid review by operational teams
- Transparent communication of issues across functions
- Readily available evidence to support audits and inspections

By standardizing validation outputs, the framework improves explainability and reduces the effort required to demonstrate data quality execution to internal and external stakeholders.

6. Role of Apache Airflow as the Orchestration Layer

Within the proposed framework, Apache Airflow serves as the orchestration layer responsible for coordinating the execution of data quality validations. While Airflow is a technical tool, its role can be understood in simple terms: it ensures that the right checks run at the right time, in the correct sequence, and that the outcomes are recorded in a controlled manner.

From a data quality and governance perspective, orchestration is essential. Validation checks must not only exist, but must also be executed consistently across data refresh cycles, studies, and domains. Airflow provides a centralized mechanism to manage this execution without relying on manual intervention.

Key contributions of Airflow within the framework include:

- **Controlled scheduling**: Validation checks can be triggered automatically based on predefined schedules or data availability, reducing the risk of missed or delayed executions.
- **Dependency management**: Checks can be executed in a defined order, ensuring that prerequisite data preparation steps are completed before validation begins.
- **Execution traceability**: Each validation run is logged with execution status and timestamps, providing clear visibility into what ran and when.
- **Operational visibility**: Centralized monitoring allows teams to quickly identify failures and assess their impact without manually reviewing individual logs or scripts.

Importantly, Airflow does not define the validation logic itself. Validation rules remain metadata-driven and independent of the orchestration tool. This separation ensures that changes to validation requirements do not require changes to the orchestration framework, preserving flexibility and governance.

From a business and compliance standpoint, the use of an orchestration layer strengthens confidence in data quality execution. It reduces operational risk by enforcing consistent execution, improves transparency across stakeholders, and supports inspection readiness by providing traceable evidence of validation activities.

By embedding orchestration into the data quality framework, organizations move from reactive validation toward a disciplined, lifecycle-driven approach. This creates a reliable foundation for scalable data quality execution as data volumes, sources, and regulatory expectations continue to evolve.

7. Case Study: Metadata-Driven Validations on Source Data

To demonstrate the practical application of the orchestrated data quality framework, a proof-of-concept case study was implemented using representative source clinical datasets. The objective was to validate that metadata-driven rules, when combined with controlled orchestration, could execute data quality checks consistently and transparently without embedding logic directly into programs.

7.1 Use Case Overview

The case study focused on source-level clinical data commonly reviewed early in the study lifecycle, specifically demographic (DM) and adverse event (AE) datasets. These domains were selected because they are widely used, involve both single-dataset and cross-dataset relationships, and are reviewed by multiple stakeholder groups.

Validation rules were defined through metadata, specifying:

- Target datasets and variables
- Rule intent (e.g., completeness, consistency, logical relationships)
- Evaluation criteria for pass or fail outcomes

Examples of validation scenarios included checks within a dataset (such as mandatory field population) as well as checks spanning multiple datasets (such as consistency of subject identifiers across DM and AE). By defining these rules through metadata, the same validation logic could be reused across data refreshes without modifying execution code.

The scope of the case study was intentionally limited to source clinical data, prior to any SDTM or ADaM transformations, to allow focused evaluation of orchestration and governance capabilities.

7.2 Validation Execution and Outputs

Once validation rules were defined, execution was automated through the orchestration layer. Validations were triggered based on predefined conditions, ensuring consistent execution whenever source data was refreshed or made available for review.

Each execution produced standardized outputs summarizing validation results. These outputs included clear PASS or FAIL indicators for each rule, along with contextual information such as execution timing and rule reference. The standardized format allowed results to be easily reviewed by programmers while remaining interpretable for non-technical stakeholders such as data managers and quality reviewers.

When validation requirements changed, updates were made directly to the metadata definitions. No changes to execution logic or orchestration configuration were required. This demonstrated a key benefit of the approach: validation logic could evolve independently of the execution framework, reducing operational effort and minimizing the risk of unintended impacts.

Overall, the case study confirmed that orchestrated, metadata-driven data quality execution can improve consistency, transparency, and governance at the source-data level, while remaining adaptable to changing validation needs.

dm.csv   ae.csv   ex.csv   schema.yml.txt   clinical_study_schema_dq.py   schema_rules.yaml

clinical_schema_dq_20260106_1└8631.csv

8. Stakeholder Utilization and Benefits

An effective data quality framework must support a broad range of stakeholders involved in clinical study delivery. While the underlying execution may be technical, the outcomes and benefits must be accessible and meaningful to non-technical roles. The orchestrated, metadata-driven approach enables this by providing consistent, transparent, and interpretable validation outcomes.

8.1 Data Management and Clinical Monitoring Units

For Data Management and Clinical Monitoring Units, early visibility into data quality issues is critical. By executing validations at the source-data level and producing standardized outputs, the framework allows teams to identify and address issues closer to data entry or ingestion.

Key benefits include:

- Earlier detection of data quality issues
- Reduced downstream rework during later study stages
- Clear, repeatable validation outcomes that support review discussions

Because validation logic is metadata-driven and centrally governed, outcomes remain consistent across refresh cycles, improving confidence in review decisions.

8.2 Medical Reviewers

Medical reviewers rely on accurate and consistent data to assess subject safety and clinical events. In traditional workflows, inconsistent validation execution or delayed checks can obscure issues or require additional clarification.

The orchestrated approach provides:

- Predictable execution of checks relevant to medical review
- Clear indication of data issues affecting subject-level interpretation
- Improved trust in the data presented for review

Standardized outputs help medical reviewers focus on clinical relevance rather than questioning the reliability of underlying data.

8.3 Quality and Compliance Teams

From a quality and compliance perspective, the primary concern is not only whether issues were detected, but whether data quality processes are controlled, traceable, and auditable.

The framework supports these needs by:

- Providing traceable execution evidence for each validation run
- Ensuring validation logic changes are controlled through metadata
- Producing inspection-ready outputs that can be readily reviewed

This reduces the effort required to respond to audit or inspection queries and strengthens confidence in the overall data quality process.

8.4 Programming and Technical Teams

For programming and technical teams, the separation of validation logic from execution reduces duplication and maintenance effort. Programmers can focus on developing reusable validation logic, while orchestration ensures consistent execution.

Key advantages include:

- Reduced need for manual reruns and ad hoc scripts
- Simplified updates to validation requirements
- Improved collaboration through shared, governed rule definitions

By standardizing execution and outputs, the framework also improves communication between programmers and non-technical stakeholders.

9. Regulatory and GxP Considerations

In regulated clinical research environments, data quality processes must be demonstrable, controlled, and auditable. Regulatory expectations extend beyond identifying data issues to ensuring that validation activities are executed consistently and can be clearly explained during inspections or audits.

The orchestrated data quality framework supports these expectations by introducing structure and traceability into validation execution. Validation rules are defined through metadata, enabling controlled updates and clear documentation of validation intent. Each execution of a validation check is performed in a consistent manner and produces standardized outputs that capture execution context and outcomes.

From a GxP perspective, key considerations addressed by the framework include:

- **Traceability:** Validation results can be traced back to specific rule definitions and execution instances.
- **Consistency:** Automated execution reduces variability introduced by manual processes.
- **Transparency:** Standardized outputs allow reviewers to understand what was checked, when it was checked, and what the outcome was.
- **Governance:** Changes to validation logic can be managed through controlled updates to metadata, supporting change management practices.

The framework does not replace established validation standards or regulatory requirements. Instead, it provides a structured mechanism to consistently apply those standards within the data pipeline. By strengthening execution discipline and documentation, the approach supports inspection readiness and reduces the effort required to respond to regulatory queries.

Importantly, the use of open-source components within the framework does not diminish regulatory control. When implemented within a governed environment, open-source tools can support compliance objectives by enabling transparency, version control, and reproducibility. The emphasis remains on controlled usage, documented processes, and clear ownership of validation activities.

10. Limitations, Risks, and Future Scalability

While the orchestrated, metadata-driven framework demonstrates clear benefits for source-level data quality execution, it is important to acknowledge its current limitations, associated risks, and areas for future expansion. Addressing these considerations transparently strengthens governance and supports realistic adoption.

10.1 Current Scope Limitations

The current implementation is intentionally limited to **source clinical data**, such as data from EDC systems and upstream sources. Downstream datasets, including SDTM and ADaM, are not included in the present scope.
This limitation was a deliberate design choice to:

- Validate orchestration and governance concepts in a controlled setting
- Minimize complexity during initial adoption
- Focus on early-stage data quality where issues can be corrected with minimal downstream impact

While SDTM and ADaM validation remain essential, extending the framework to these datasets requires additional considerations, including standard-specific rules, transformation dependencies, and alignment with submission workflows.

10.2 Operational and Technical Risks

As with any orchestrated framework, certain operational and technical risks must be considered:

- **Adoption risk:** Successful use of the framework depends on stakeholder adoption, including agreement on metadata standards and governance processes.
- **Metadata quality risk:** Since validation behavior is driven by metadata, inaccuracies or poor governance of metadata definitions could impact validation outcomes.
- **Operational dependency:** Orchestration introduces reliance on scheduling and execution infrastructure, which must be monitored and maintained.
- **Change management:** Uncontrolled changes to validation rules or execution schedules could reduce confidence in outcomes if not governed appropriately.

These risks are not unique to orchestration-based approaches but must be actively managed to ensure long-term effectiveness.

10.3 Mitigation Strategies

Several mitigation strategies can reduce the impact of these risks:

- Establishing clear ownership and governance for validation metadata
- Implementing version control and review processes for rule changes
- Defining controlled execution schedules aligned with data refresh cycles
- Providing clear documentation and training for stakeholders

By embedding governance alongside technical implementation, the framework supports sustainable and compliant operation.

10.4 Future Scalability and Extension

The architectural principles described in this paper—separation of concerns, metadata-driven rules, and controlled orchestration—are designed to support future scalability. With appropriate governance and standardization, the framework can be extended to support SDTM and ADaM datasets, as well as additional data sources.
Future enhancements may include:

- Standard-specific rule libraries for SDTM and ADaM
- Expanded cross-dataset and cross-domain validations
- Integration with downstream analytics and reporting workflows

These extensions would allow organizations to progressively expand orchestration-based data quality execution across the full clinical data lifecycle while maintaining control and traceability.

11. Comparison with Existing Tools and Approaches

A variety of tools and frameworks are currently used to support data quality activities in clinical research. These range from embedded validation logic within statistical programs to standalone data quality platforms and rule-based engines. While each approach addresses specific needs, they often differ in how validation logic is managed, executed, and governed.

Traditional approaches typically involve validation checks written directly within domain-specific programs or scripts. While effective for individual use cases, these implementations are often tightly coupled to specific datasets or workflows. This coupling can make reuse difficult and introduce inconsistencies when similar checks are implemented independently across studies or teams.

Standalone data quality tools and platforms offer centralized rule management and reporting capabilities. However, these tools may operate independently of the data pipeline lifecycle, requiring manual coordination to ensure that

checks are executed at appropriate times. In addition, some solutions rely on proprietary configurations that can limit transparency and flexibility.
In contrast, the orchestrated, metadata-driven approach described in this paper emphasizes execution discipline and governance rather than tool-specific functionality. By separating validation logic from execution and embedding orchestration into the data pipeline, the framework ensures that data quality checks are executed consistently and predictably.
Key points of differentiation include:

- **Execution-first focus:** Orchestration ensures checks run reliably at defined points in the data lifecycle.
- **Metadata-driven flexibility:** Validation rules can evolve without modifying execution logic.
- **Transparency:** Open, readable configurations support explainability across stakeholders.
- **Governance alignment:** Execution evidence and traceability are built into the framework.

Rather than replacing existing tools, the framework can complement them by providing a controlled execution layer. Organizations may continue to use established validation standards and tools while leveraging orchestration to improve consistency, visibility, and governance.

12. Impact and Key Outcomes

The orchestrated, metadata-driven data quality framework demonstrated measurable benefits across operational, governance, and strategic dimensions. By embedding data quality execution into the data pipeline lifecycle, the approach addressed long-standing challenges associated with manual and fragmented validation practices.
From an operational perspective, automated and controlled execution reduced manual effort and rework associated with repeated data refreshes. Validation outcomes became predictable and repeatable, allowing teams to focus on issue resolution rather than execution logistics.
From a governance and compliance standpoint, standardized and traceable outputs strengthened inspection readiness. Validation activities could be clearly demonstrated through execution records, reducing the effort required to respond to audit or inspection queries.
Strategically, the framework established a scalable foundation for future data quality enhancements. By separating validation logic from execution and driving rules through metadata, the approach supports gradual expansion to additional datasets, domains, and standards as organizational maturity evolves.
Equally important, the framework improved collaboration across technical and non-technical stakeholders. Clear, interpretable outputs enabled shared understanding of data quality outcomes, strengthening confidence in the data used for review and decision-making.

13. Conclusion

As clinical data pipelines continue to grow in complexity, data quality practices must evolve from isolated, manually executed checks to coordinated and governed execution models. Traditional approaches, while effective in limited contexts, struggle to scale and provide the transparency and traceability expected in regulated environments.
This paper demonstrated how an orchestrated, metadata-driven approach can address these challenges by embedding data quality execution into the lifecycle of the clinical data pipeline. By separating validation logic from execution, automating controlled runs through orchestration, and producing standardized, audit-ready outputs, the framework improves consistency, operational efficiency, and inspection readiness.
Through a source-data–focused case study, the approach showed how validation rules can be executed predictably and transparently without increasing manual burden. Importantly, the framework supports collaboration across technical and non-technical stakeholders, enabling shared understanding of data quality outcomes.
While the current implementation is limited to source clinical data, the architectural principles described provide a scalable foundation for future expansion to downstream datasets, including SDTM and ADaM. By adopting orchestration-based execution and metadata-driven governance, organizations can strengthen trust in their data while preparing for evolving regulatory and operational demands.

14. CONTACT INFORMATION

Author Name: Pavan Kumar Reddy Sakam
Company: AstraZeneca
Email: sakampavankumar@gmail.com

Co Author Name: Vinodh Kumar Jestadi
Company: AstraZeneca
Email: vkjestadi@gmail.com