Automation of CDISC Data Review remains no longer optional: It's A Critical Need of the Hour

Author: Mrityunjay Kumar
Ephicacy Lifescience Analytics, Bengaluru

Meet the Speaker

Mrityunjay Kumar
Title: Associate Director (Statistical Programming)
Company: Ephicacy Lifescience Analytics

He leads a team of experienced programmers that includes SAS & R team and works for Medical Affairs and Market Access, Oncology, Hematology etc.

With over 19+ years of experience in Clinical Statistical Programming including SAS, R, and Python, he enjoys working on innovative solutions and expertise building for clinical data science.

He holds Master of Science (M.Sc.) degree in Bioinformatics from Sri Ramchandra University, Chennai (India) in 2006 and a certificate in Future Leadership Program from National University Singapore (NUS) in 2022.

Agenda

- Why Data Review Matters
- Data Review Pain Points
- Why need for Automation ?
- Automation Architecture
  - Automation of CDISC Review: 5 Layers
  - Automated Review: QC & Validation
  - Case Study Example
- Tools & Technologies
- Implementation Strategy
- Future Perspective
- Key Takeaways
- Q&A

Why Data Review Matters ?

Data Review Pain Points

Manual Review Limitations

- Repetitive checks (e.g., variable consistency across datasets)
- Reviewer fatigue → high risk of missed discrepancies
- Dependency on individual expertise
- Lack of audit trail for decisions

Real-Life Examples

- Mismatch between SDTM and ADaM population flags
- Inconsistent treatment variables across datasets
- TLF numbers not matching ADaM outputs
- Late discovery of issues → rework → submission delays

We are spending ~70% time finding issues, not solving them

Why need for Automation ?

Trial Growth

Trial complexity increasing

Trial size increasing

Number of trials increasing

Data Explosion

RWD, biomarkers, devices

Multi-source data integration

Increasing Pressure on Clinical Trials & Data Review

Regulatory Pressure

Traceability

Reproducibility

Accelerated approvals

Operational Challenges

Checklists based repetitive tasks

Pressure on Profitability

Economic conditions

Automation Architecture

What Does ‘Automation of CDISC Review’ Really Mean?

Layer 1: Structural Checks
- ✓ Dataset structure (SDTM/ADaM compliance)
- ✓ Variable attributes (type, length, label)
- ✓ Controlled terminology

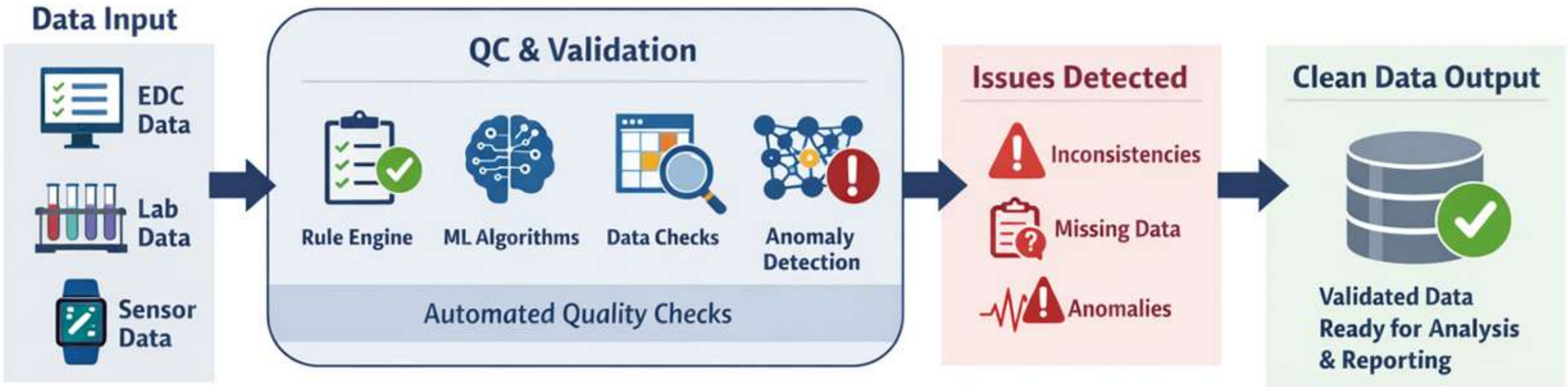
Layer 2: Cross-Dataset Consistency
- ✓ SDTM vs ADaM mapping validation
- ✓ Population flag consistency (SAFFL, ITTFL)
- ✓ Treatment alignment

Layer 3: Data Integrity Checks
- ✓ Missing values
- ✓ Outliers
- ✓ Logical inconsistencies (e.g., end date < start date)

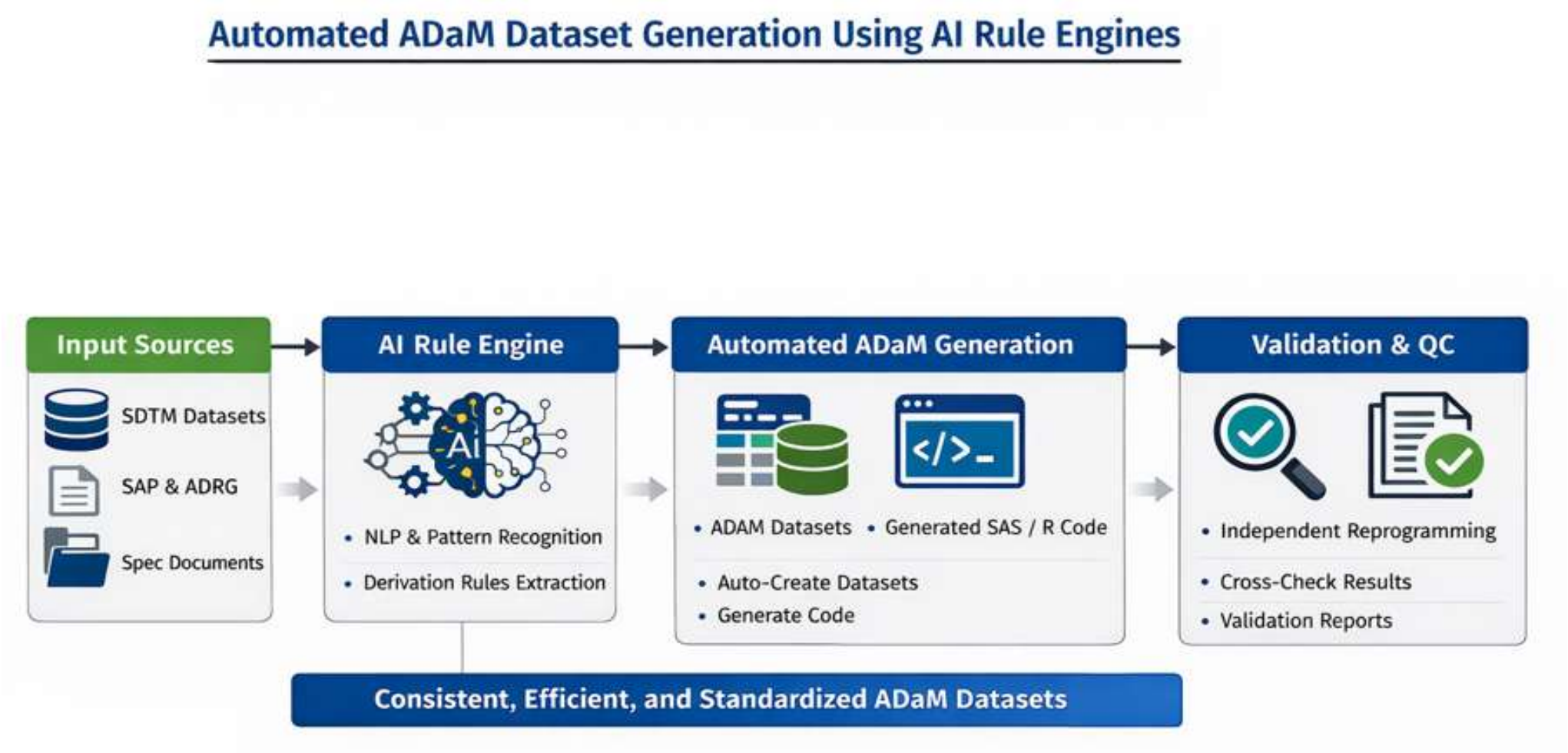
Layer 4: TLF Validation
- ✓ Table numbers vs ADaM datasets
- ✓ Derived variables validation
- ✓ Reproducibility of outputs

Layer 5: Metadata-Driven Review
- ✓ Using define.xml / spec-driven checks
- ✓ Traceability from SDTM → ADaM → TLF

Automated Review: QC and Validation



Case Study Example



Tools & Technologies

Common Tools

- **SAS** (validation macros, PROC COMPARE)
- **R** (data validation frameworks, tidyverse)
- **Python** (automation pipelines, ML checks)

Industry Tools

- Pinnacle 21 (compliance)
- Open-source validation frameworks
- Custom in-house solutions

Evolve Beyond Single Tools: Embrace Hybrid Ecosystems

Implementation Strategy

1. Identify high-impact repetitive checks
2. Build reusable validation rules
3. Standardize metadata inputs
4. Create centralized validation repository
5. Integrate into programming workflow

Start small, scale fast - but standardize early
Integrated Tech, Flexible Work: The New Standard

Future Perspective

Future Outlook

- AI-driven review assistants
- Real-time validation during programming
- Integration with submission systems
- Federated learning for cross-study insights

Challenges

- No specific guidance for AI-assisted programming by FDA/EMA
- AI Hallucination Risk
- CDISC TA specific domain knowledge gap
- Evolving regulatory framework

Key Takeaways

The future of clinical trials lies in smarter, reproducible, and integrated statistical programming. Organizations must embrace innovation to remain agile and regulatory-ready. The journey begins with upgrading skills, systems, and mindset.

In today's clinical landscape, **automation** is not about efficiency. It's about **credibility**, **compliance**, and **speed of drug outreach** to patients.

1. Manual review alone is not scalable
2. Automation improves quality and efficiency
3. Metadata-driven frameworks enable standardization
4. Visualization enhances reviewer productivity
5. AI will shape the future of clinical data review

References

Next-Gen Statistical Programming in Clinical Trials: Driving Faster, Reliable, and Smarter Outcomes - *Mrityunjay Kumar, Ephicacy Lifescience Analytics; Shashikant Kumar, Ephicacy Consulting Group* https://phuse.s3.eu-central-1.amazonaws.com/Archive/2026/Connect/APAC/Hyderabad/PAP_PD05.pdf

https://www.cdisc.org/sites/default/files/2022-11/2022-US-Interchange-Automating%20Efficiency.pdf

https://www.ijrti.org/papers/IJRTI2505298.pdf

CDISC ARS User Guide version 1.0: Analysis Results Standard User Guide v1.0 - Analysis Results Standard Document – Wiki

PharmaSUG 2025 - Paper AI-349 Harnessing AI & CDISC ARS for Effortless Statistical Reporting & CSR Writing

CDISC 360i - https://www.cdisc.org/sites/default/files/2021-08/Automation_of_SDTM_and_ADaM_Generation_and_Artifacts_using_CDISC_360_Enriched_Metadata.pdf

THANK YOU