

PP33- Reproducibility, Lineage, and Control: Building a Traceability Framework for the Modern Clinical Trials

As clinical trials become more complex, global, and data-intensive, the margin for ambiguity in how data is collected, transformed, and analyzed has vanished. Regulatory expectations are rising (from ICH E6(R3) to 21 CFR Part 11 and EMA’s emphasis on data integrity) and with them, the need for version control, audit trails, full traceability, reproducibility, and consistent handling of both blinded and unblinded data with granular access control across the entire clinical data lifecycle has become mission-critical. This poster presents a comprehensive, future-ready framework for establishing robust traceability from source data to submission deliverables. It details how to implement systematic lineage tracking across clinical systems: from EDCs and lab feeds through to SDTM transformations, ADaM datasets, TLF outputs, and final regulatory submission packages.

KEY ELEMENTS OF MODERN STATISTICAL COMPUTING ENVIRONMENT

Version Control

- Version control for all CSR programming artifacts—data, programs, and outputs—is essential. It enables reproducibility. It ensures data integrity. It creates an immutable and auditable history.
- Tools like Git and SVN work well for code. They do not work well for large clinical datasets such as SDTM or ADaM. They are inefficient and require heavy customization or manual effort.
- Versioning only the code is not enough. Regulators expect full version control of datasets, metadata, outputs, and all analytical artifacts.

Access Governance

- A permanent, immutable record of all access permissions is an absolute must for regulatory compliance and to demonstrate oversight and governance of the clinical trial analysis process.
- The system must create an auditable history of policy changes:
 - Who had access, when?
 - Who authorized the access?
 - What were they authorized to access?
- Solution should manage the versioning of the access control policy that surrounds the data and log the granular events.

Audit Trails

- An audit trail is another essential requirement for a regulated solution. It is a chronological, computer-generated record of all events, actions, and changes. It provides definitive proof of what happened and when.
- Audit trails must be immutable (WORM). Once recorded, entries cannot be changed or deleted. This prevents tampering and protects data integrity.
- They must be time-stamped. Every entry needs a consistent, tamper-proof date and time.
- They must be attributable. Each action must be linked to a specific user or system process.

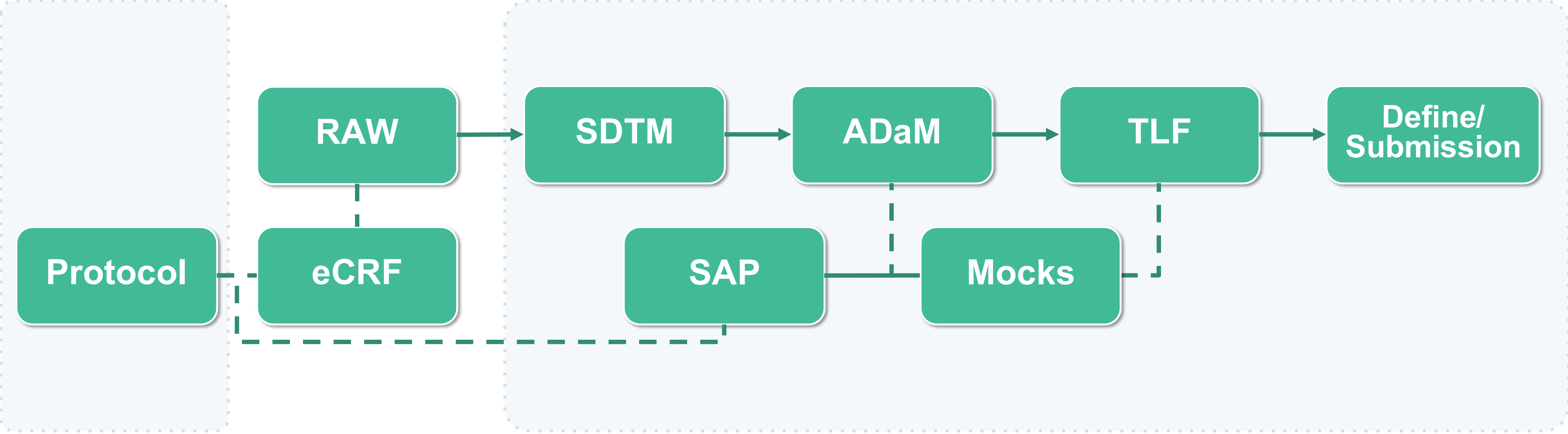
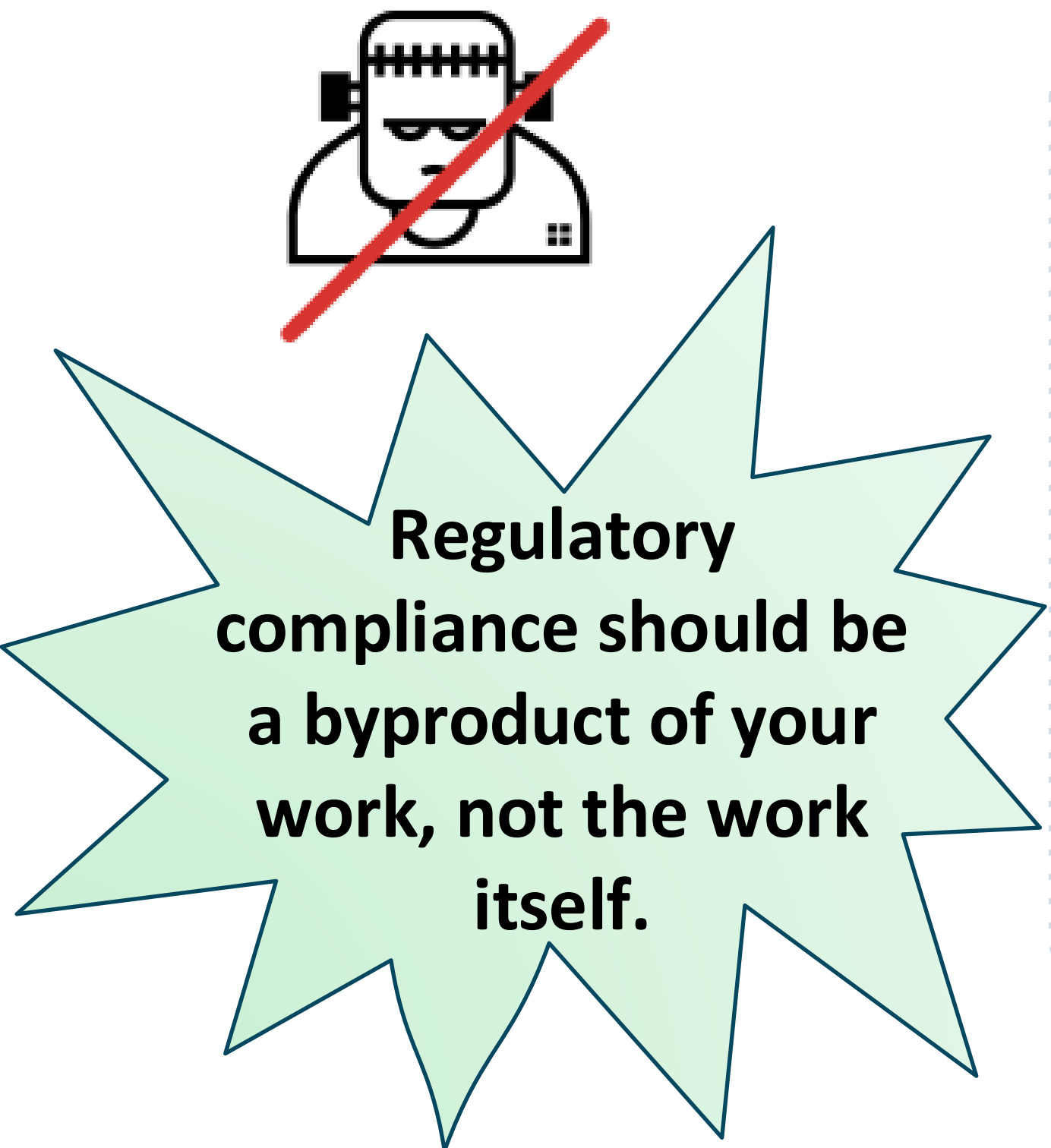
Traceability & Lineage

- A verifiable, end-to-end relationship between the data, programs, and output is a must. It should connect every output directly back through the analysis program, the specific ADaM dataset version, the source SDTM data version, and ultimately, to the raw patient data point and the protocol.
- It functions as the immutable metadata linkage across all steps of the analytical process, enabling the sponsor and regulatory reviewer to answer, “How was this result produced?” with a definitive and reproducible answer at any point in time.

Polyglot, AI, LLM, & Agentic AI

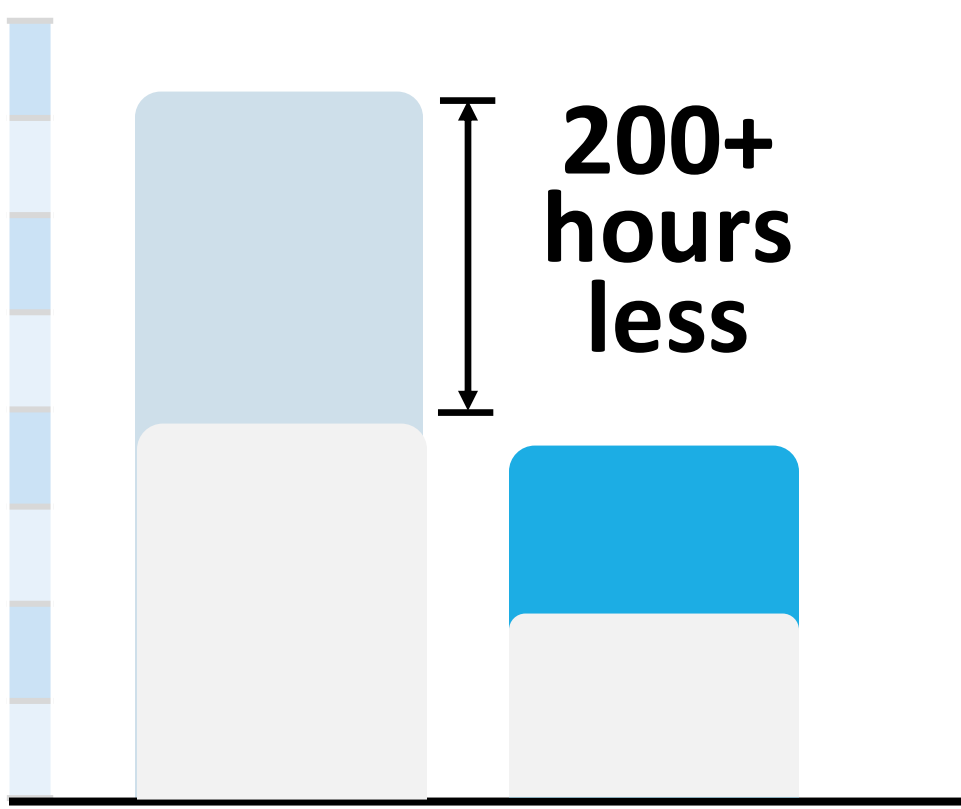
- Simultaneously have multiple versions of open-source computes (e.g., of R) within the same controlled environment
- Manage R or Python packages and ensure validated use through package server. Ensure vulnerability checks of Packages
- Create and preserve any number of reproducible runtime environments required for maintaining regulatory submission compliance
- Create and publish interactive apps (e.g. RShiny and Python Streamlit) with granular access control to the underlying data.
- Shared infrastructure that provides biostatisticians with a personal workspace for programs and results, a dedicated environment to install needed packages, and read-only access to clinical data.
- Use of familiar IDEs

As regulatory expectations become more rigorous, a single, end-to-end solution is essential

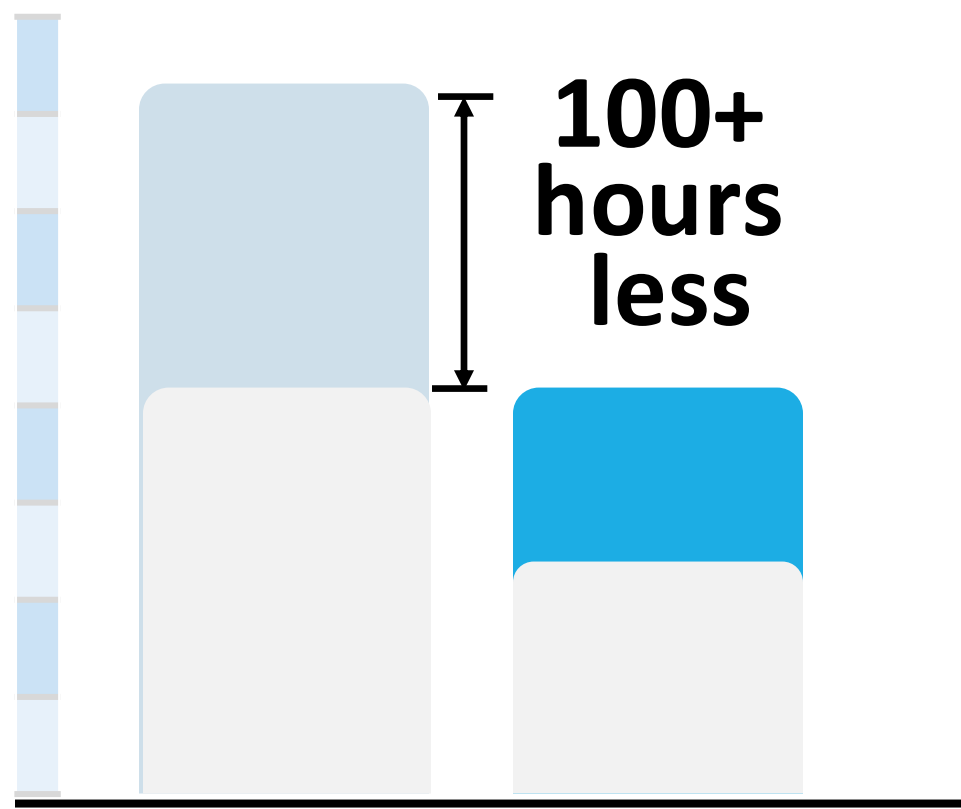


WHAT BUSINESS EFFICIENCIES ARE ACHIEVABLE BY ADOPTING A MODERN SOLUTION?

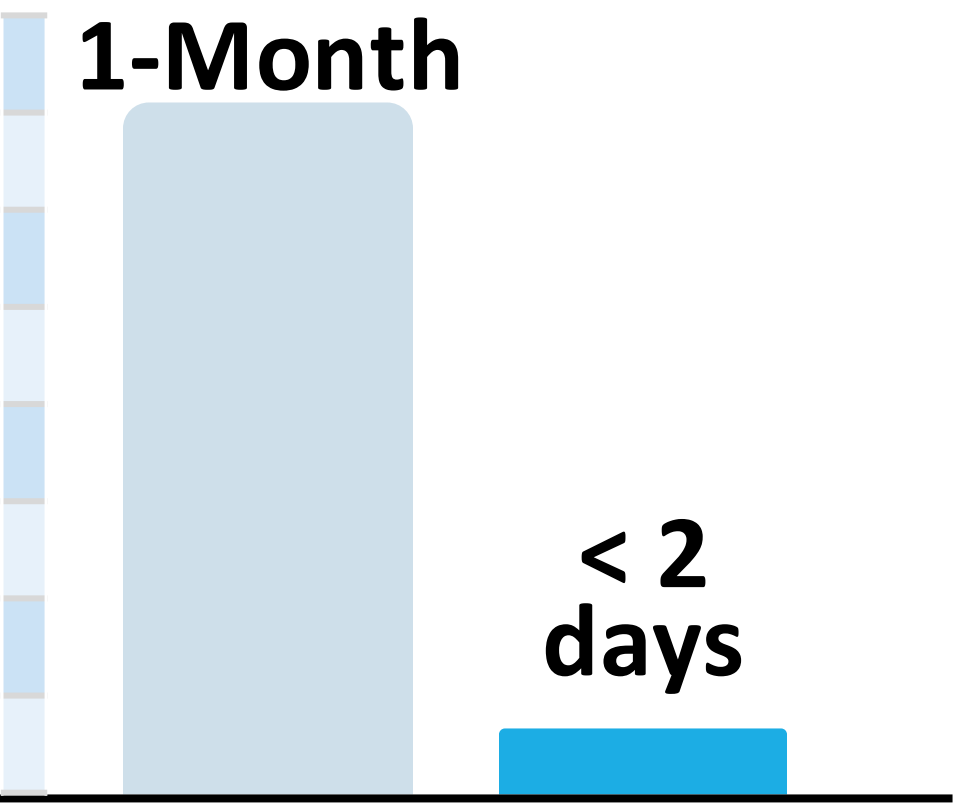
Clinical Trial Analysis and Reporting Key Performance Indicators (KPIs) must focus on business deliverables and critical activities.



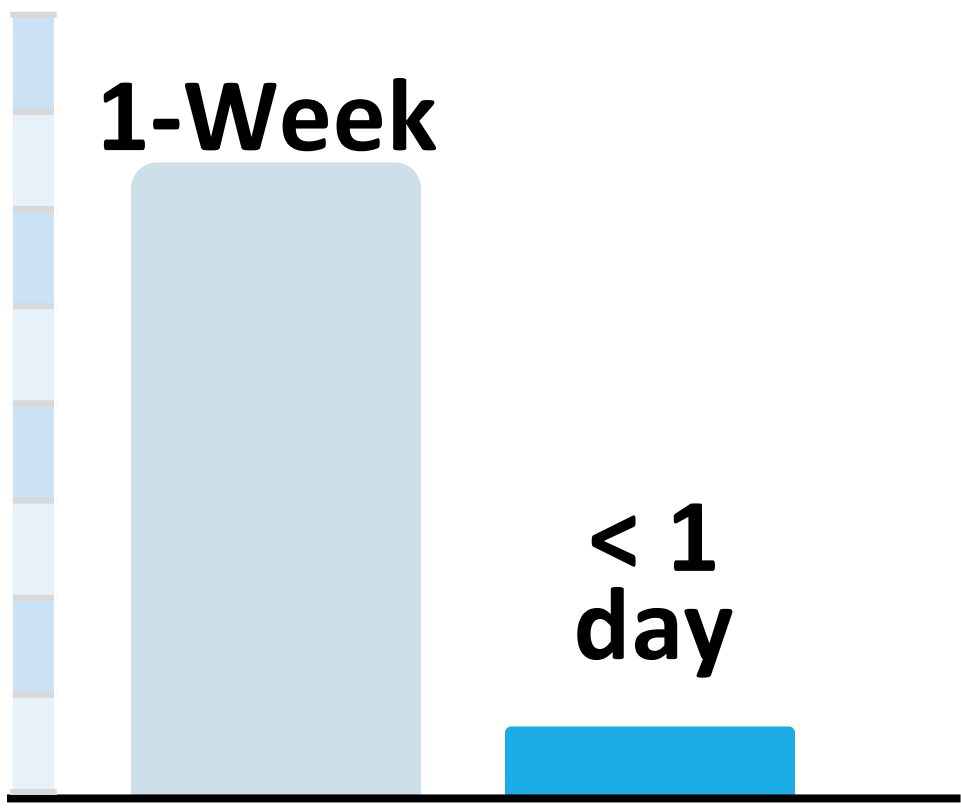
Programmer Productivity: Code Review & Debugging Effort Per Study Per Programmer



Answering Regulatory Health Queries Such as ‘who, what and when’....



Delivery of Final TLFs for a CSR



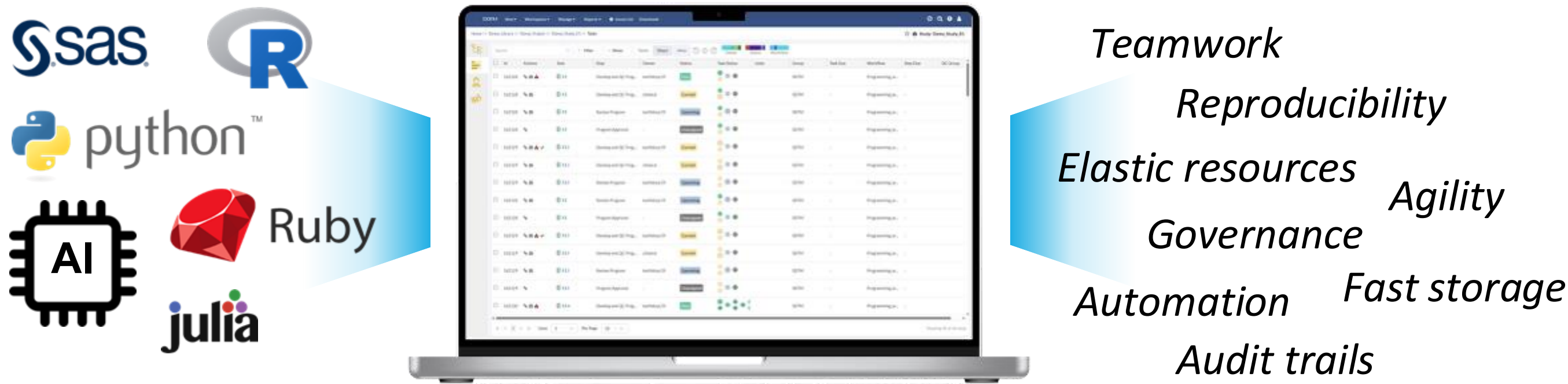
Delivery of Top-Line TFL Results

Other Example KPIs:

- SDTM/ADaM and TLF Cycle Time:** Days from when SDTM is finalized to ADaM completion, and days from ADaM completion to TLF package delivery
- TLF Iteration Burden per Study:** Total # of iteration QC/loops across all TLFs in a study
- QC Pass Rate:** Percent of SDTM/ADaM datasets & TLFs that pass QC on first attempt

WHAT SHOULD A MODERN SOLUTION HAVE?

- Single comprehensive application** with built-in version control, access control, compute management, traceability, reproducibility, and comprehensive audit trails.
- Polyglot programming:** Seamless integration and execution of SAS, R, Python, R shiny, and Agentic AI apps within the same GxP compliant ecosystem
- Automated governance:** Built-in 21 CFR Part 11, GxP, and EU Annex 11 compliance
- End-to-end traceability:** Complete visibility from data acquisition to submission, including data provenance, blinding/unblinding, and chain of custody.
- Reproducibility and transparency:** Every dataset, program, and output can be reproduced with a full audit trail for verification and compliance.
- Integrated AI and automation:** Agentic AI capabilities for intelligent code review, dataset validation, and report generation
- Visualization and control:** Visual tracking of study progress with one-click validation summary reports for faster business delivery such as topline report, CSR, ISS, ISE etc



ONE single comprehensive application that is GxP, 21CFR11, and EUAnnex11 compliant with audit trails, granular access control, version control, chain of custody, traceability, and reproducibility for regulated polyglot execution, and high-performance storage

