

What did the 2026 CDISC AI winners actually automate?

C1 · 30 minutes · CDISC × AI × SDTM/ADaM/SAP/TFL

Why selected

CDISC has published its 2026 AI Innovation Challenge winners across synthetic SDTM/ADaM generation, SAP generation and TFL generation. The winning approaches are standards-native workflows, making this a direct snapshot of where clinical-programming automation is moving.

Reading order

- 0–4 min · Map the three 2026 use cases.
- 4–13 min · Focus on SAP and TFL winners: inputs, outputs and traceability.
- 13–19 min · Ask why synthetic standards-native datasets matter for automation testing.
- 19–24 min · Compare with CDISC TechniCon protocol-to-CSR pipelines.
- 24–25 min · Separate automatable execution from judgment-heavy SP work.
- 25–30 min · Design a minimal AI-assisted TFL pipeline.

Brief background

The three use cases form a simple chain: data → plan → report. Syn2Real generates synthetic SDTM/ADaM from USDM definitions; Smart ClinSAP spans protocol interpretation toward executable SAS/R; PAIR uses SAPs, ADaM specifications and TFL shells to produce reporting packages.

The repeated principle is standards-native traceability. A production workflow should preserve the path from study intent and metadata through analysis data and generated code to the reported result.

TechniCon demonstrated the same direction: connected standards-based pipelines using USDM, SDTM, ADaM, Analysis/Derivation Concepts and ARS/ARD from protocol through reporting.

As code generation becomes cheaper, SP value shifts upward toward ambiguity resolution, derivation design, edge-case validation, change control and ownership of the evidence path.

Key vocabulary

Term	中文	Meaning
standards-native	标准原生	Designed around formal standards as core machine-readable inputs, not added after generation.
synthetic data	合成数据	Artificially generated data designed to reproduce useful structural or statistical properties without copying real participant records.
traceable linkage	可追溯关联	An explicit connection from an output back to its source input, metadata, or rule.
executable code	可执行代码	Program code that can be run directly rather than remaining narrative specification.
reporting package	报告包	The coordinated set of analysis programs, outputs, metadata, and supporting artifacts used for reporting.
agentic workflow	智能体工作流	A workflow in which AI agents plan or execute specialized steps using tools and intermediate results.
machine-readable	机器可读	Represented in structured form that software can interpret without relying on prose reading.
interoperability	互操作性	The ability of systems and standards to exchange and use information consistently.
lineage graph	血缘图/谱系图	A graph connecting results and artifacts to upstream data, rules, and source documents.
automation testing	自动化测试	Systematic testing of automated workflows, often using repeatable inputs and expected outputs.
analysis concept	分析概念	A structured representation of the scientific/statistical analysis intent.
derivation concept	衍生概念	A reusable structured description of how an analysis variable or result is derived.
governance layer	治理层	Controls that enforce traceability, standards compliance, review, and accountability.
human-in-the-loop	人在回路	A design where people review, approve, or intervene at defined workflow points.

Term	中文	Meaning
end-to-end traceability	端到端可追溯性	The ability to follow a result continuously from protocol intent through data and analysis to final reporting.

Useful phrases

1. the winning systems treat standards as inputs to automation rather than documentation produced afterward.
2. traceability is strongest when it is generated with the workflow instead of reconstructed after the fact.
3. synthetic data can make automation testing repeatable without exposing participant records.
4. a SAP generator becomes more useful when its outputs are connected to executable code and analysis metadata.
5. an agentic TFL workflow should preserve the path from shell and specification to code and result.
6. machine-readable metadata reduces the amount of study logic hidden in prose.
7. the automation boundary should be explicit before the agent starts generating artifacts.
8. standards-native design can reduce translation loss between clinical intent and implementation.
9. human review remains most valuable at ambiguity, exception, and release boundaries.
10. the real productivity gain comes from eliminating repeated handoffs while preserving evidence.

Comprehension questions

- What are the three 2026 CDISC AI Innovation Challenge use cases?
- Why is traceable linkage important for synthetic SDTM/ADaM?
- How is standards-connected SAP generation different from prose drafting?
- What inputs does CDISC say the winning TFL workflow uses?
- Which SP skills become more valuable when code generation becomes cheaper?

Retelling prompts

30 sec · Summarize the challenge as data → plan → report.

45 sec · Explain why traceability matters more than a correct-looking TFL.

60 sec · Explain how the SP role changes when SAS/R generation becomes automated.

5-minute speaking/output task

Design one minimal AI-assisted TFL pipeline. Define authoritative inputs, agent-generated work, deterministic execution/checks, SAP-to-cell traceability, and the human release boundary.

One sentence to keep

As clinical-programming automation improves, the scarce skill shifts from producing code to designing explicit rules, validating exceptions, and preserving the evidence path from clinical intent to reported result.

Sources

<https://www.cdisc.org/events/ai-innovation-challenge>

<https://www.cdisc.org/ja/node/9911>