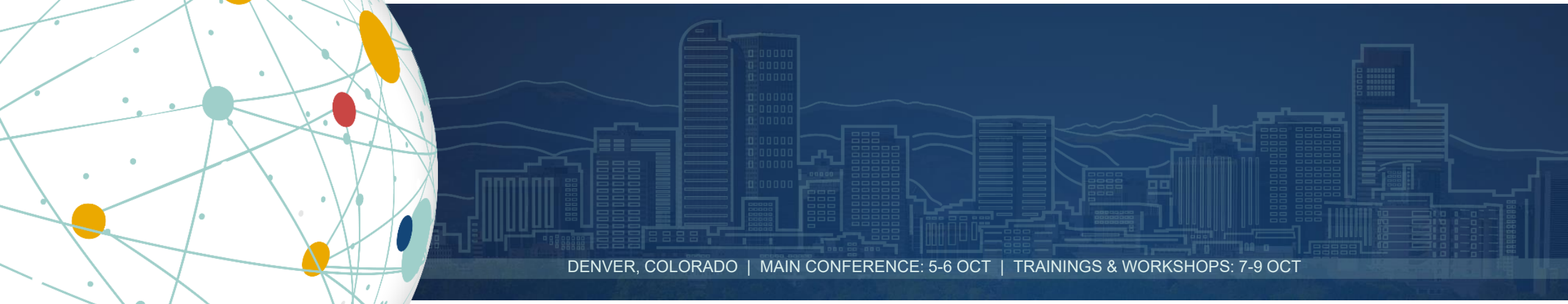


cdisc®

2026 US Interchange

THE FUTURE IS CONNECTED: STANDARDS AND AI POWERING DIGITAL TRANSFORMATION



Agentic AI Meets Data Simulation

W. Stetson Line
President, Clinventive

Stetson Line is President of Clinventive, LLC, and the creator of Verdatic, a digital twin data pipeline for clinical studies. His current work centers on agentic AI for clinical research: systems that import and curate standards, generate synthetic study data, and iteratively learn how to improve fitness for study rehearsal use-cases. A longtime CDISC volunteer, he founded and co-led the SDS Conformance Rules team, whose efforts laid the foundation for the present CORE rules catalog. He is actively seeking consensus for the defining principles and framework for assessing synthetic data quality.

Name : **Stetson Line**

Title : President

Company : Clinventive

Agenda

1. Defining synthetic data quality
2. Producing and optimizing synthetic data
3. Results and key take-aways

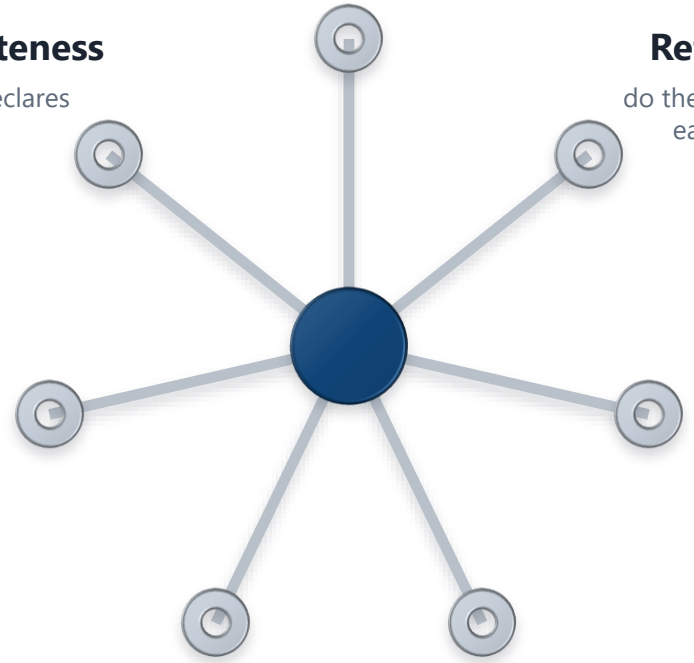
What defines synthetic data quality?

Proposal: Seven Assessment Axes

Structural conformance
does the data satisfy the published rules for the standard it claims?

Deliverable completeness
is everything the define declares actually delivered?

Referential integrity
do the records that should refer to each other actually do so?



Cross-domain consistency
do events, medications and findings agree with each other?

Temporal coherence
do dates and study days agree with the study's own calendar?

Longitudinal realism
do subjects change over time the way patients do?

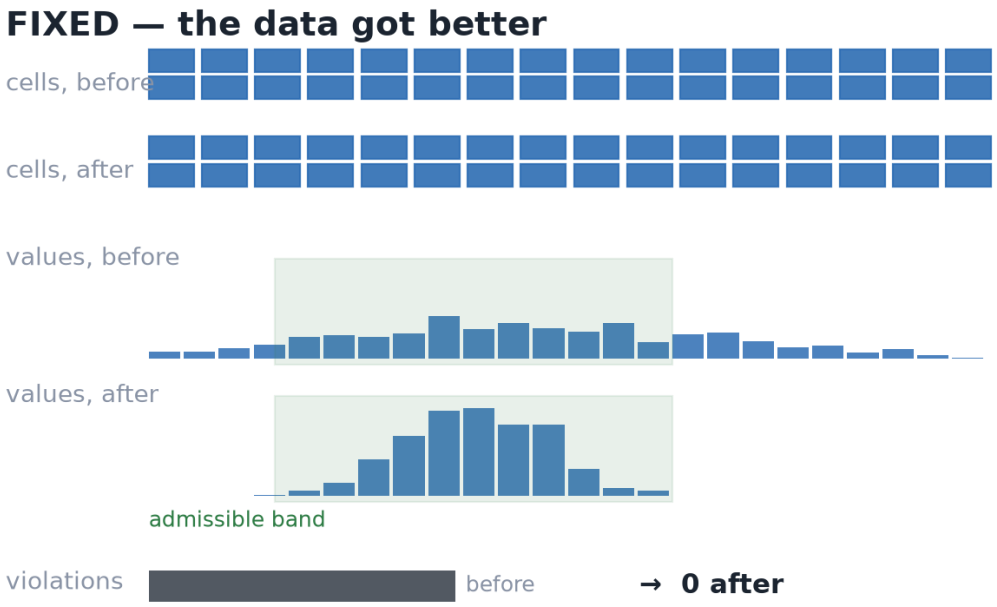
Clinical plausibility
are values and terms ones a clinician would accept?

- Each assessment reports its own number; no composite that could mask defects.
- A score counts only at constant or greater information content; no starving data to hide violations – see next slide.

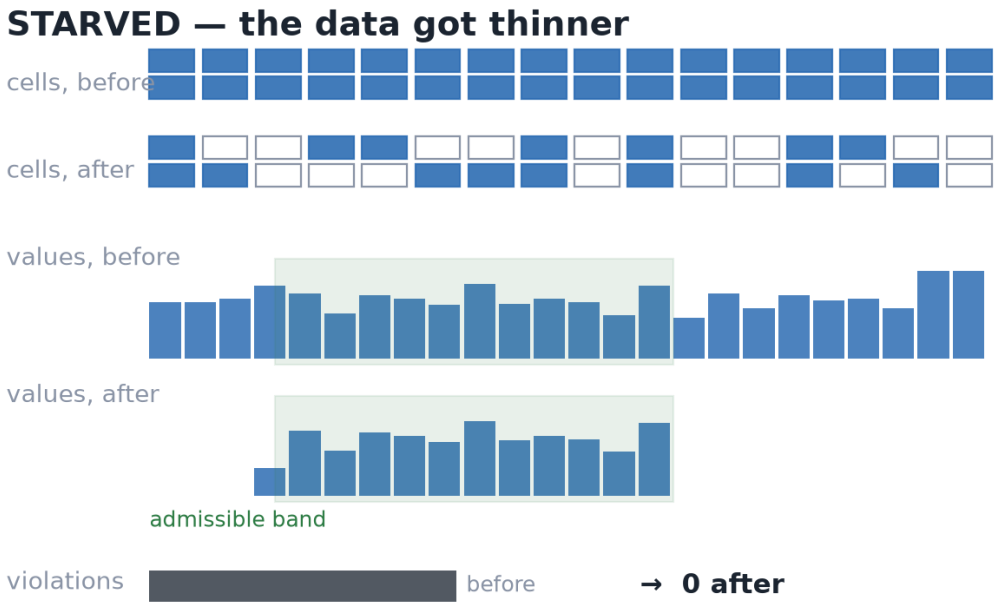
Synthetic Study Data Fidelity Benchmark, Methodology v0.1 draft, sections 3 and 4.

Why information content quantity matters: AI will game the system

- A violation count falls for two opposite reasons: the data gets better, or the data gets thinner.
- A field that stops carrying a value silences every check that first tests for the value's presence.



values moved inside the band; every cell still carries a value



values outside the band stopped being carried — what survives looks conformant; the empty cells are the price

Schematic — the mechanism, not a measurement. The study it was observed on is private; no figure from it appears here.

How can synthetic data be produced and optimized to meet that definition?

The rules that judge a dataset can also specify one

Read forward, the rule judges.

AESTDTC must be on or after RFSTDTC.

the same rule

Read backward, the rule specifies.

When generating an adverse event, draw its start date from the interval that begins at the subject's reference start date.

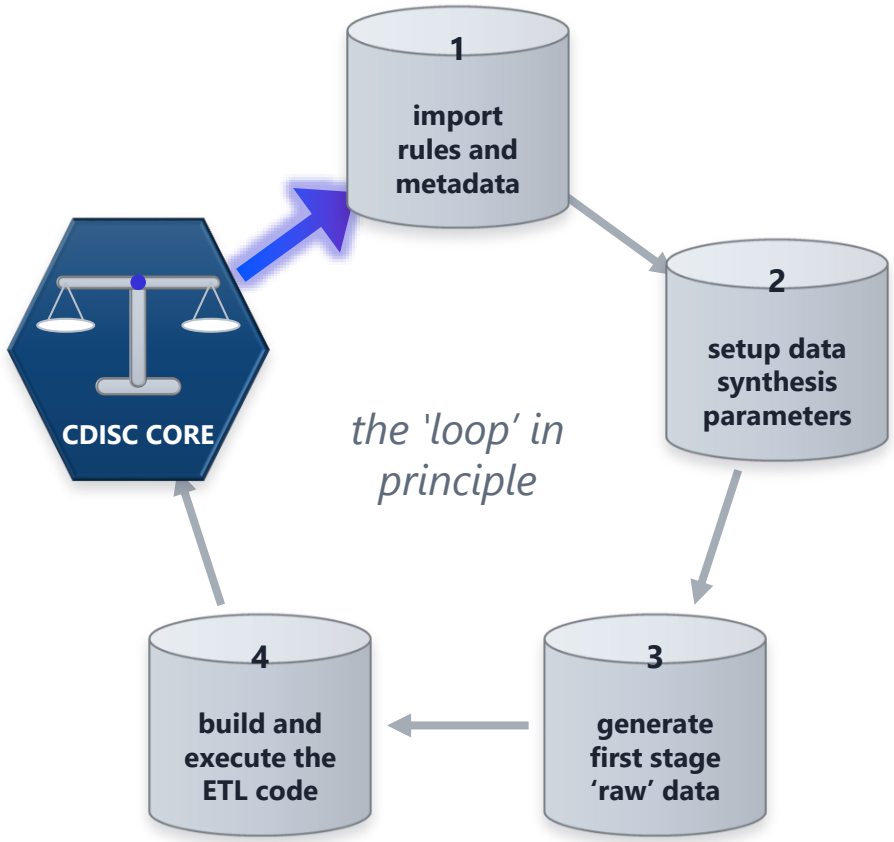
A study arrives with three rule sets: the published conformance rules for the standard, the sponsor's own edit checks, and the codelists and models that say which values are possible. All three can be read both ways.

published conformance rules

the sponsor's edit checks

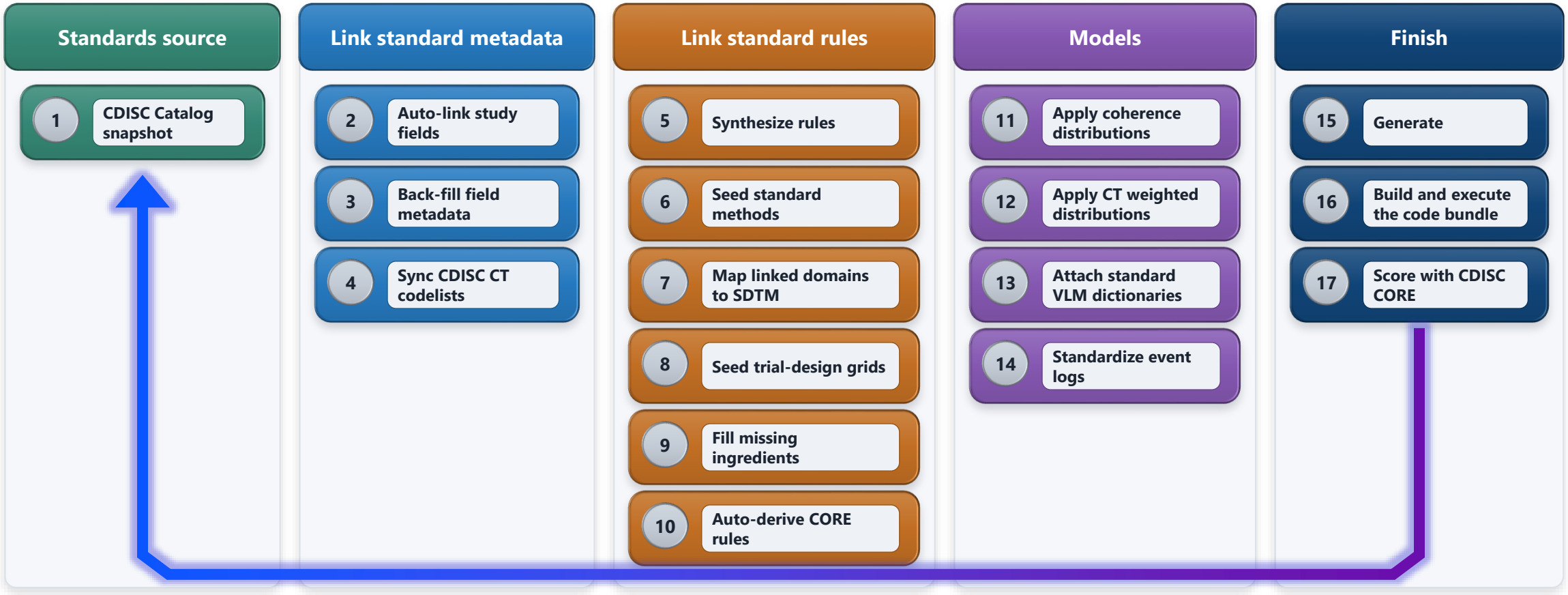
codelists and models

An agentic system that learns from rules to improve output quality



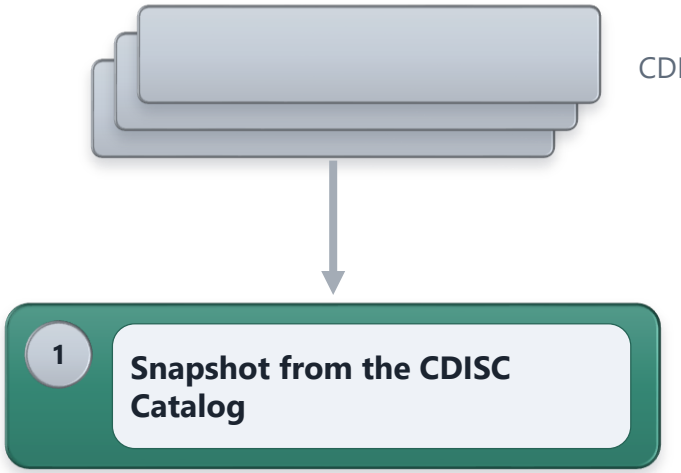
"Agentic" means exactly this: the system proposes each step from what it has read, acts on it, measures the result against the referee, and carries the result into the next pass. A human adjudicates agent decisions that fall short of confidence thresholds.

The example data synthesis engine is seventeen components in five sections, and each one primes or refines data optimization



Step 0: Import study design metadata e.g. Rave ALS, USDM, or other. This is outside the loop but a necessary pre-cursor.

Sourcing standards versioned metadata for project reference



Reads

the active CDISC catalog: dataset specializations, biomedical concepts, controlled terminology, and the CORE rule corpus for the declared standard and version. *Includes user-authored BCs/Specializations*

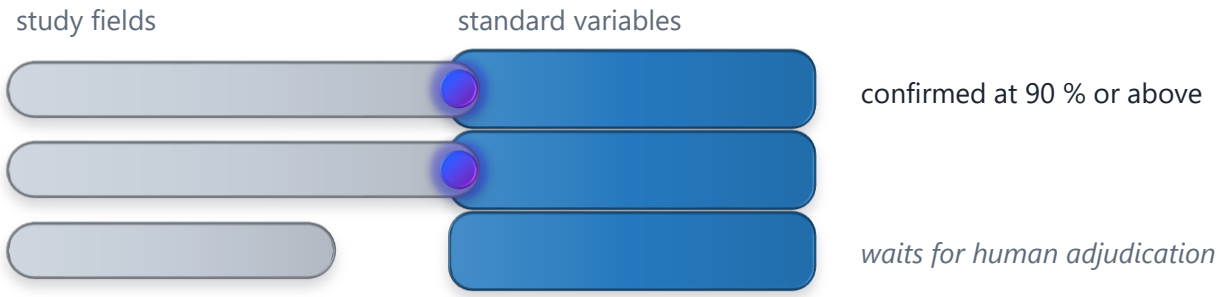
Produces

a project-scoped snapshot of the specializations the study needs, locked to one version.

Feeds the loop

every later decision resolves against this snapshot, so the rules the generator learns from and the rules CORE judges by are the same rules at the same version.

Matching study fields to standards to ‘backfill’ metadata and terminology



2 Auto-link study fields

Matches each study field to its biomedical concept and specialization, whatever the source (a CRF authoring system, USDM, or a standards scaffold), through a ladder of match methods; matches at 90 % confidence or above are confirmed automatically, the rest wait for a person.

3 Back-fill field metadata

Fills the field's role, conformance class, codelist and dictionary from the linked specification; existing hand-authored values are kept.

4 Sync CDISC CT codelists

Binds the study's dictionaries to the CDISC controlled terminology by name and adds the NCI concept identifiers and definitions to each entry.

After this section, every field knows which standard variable it is, what it may contain, and which terminology governs it.

Auto-link confirmation threshold 0.90, RaveAlsBcAutolinker.

Mining standards for rules and mappings, seeding code methods

5	**Synthesize rules**	Writes the generation and mapping configuration for each linked field from its specification.
6	**Seed standard methods**	Adds the standard derivation methods the mappings reference: subject elements from demographics, visits from the schedule, baseline flags, decodes.
7	**Map linked domains to SDTM**	Builds the mapping grid for every linked domain from its specialization: the plan the bundle is generated from.
8	**Seed trial-design grids**	Carries the trial design the generator emits (TV, TA, TE, TS) and the subject visits and elements (SV, SE) into the deliverable and enables the baseline-flag outputs every findings grid derives.
9	**Fill missing ingredients**	Lists the variables the CORE rules read that the study will not produce, each with a proposed default, its confidence and the rules it unlocks; the confident ones are accepted, the declared facts wait for a person.
10	**Auto-derive CORE rules**	Instantiates the CORE conformance rules for the study, field by field, and classifies each by how the generator can satisfy it.



rules



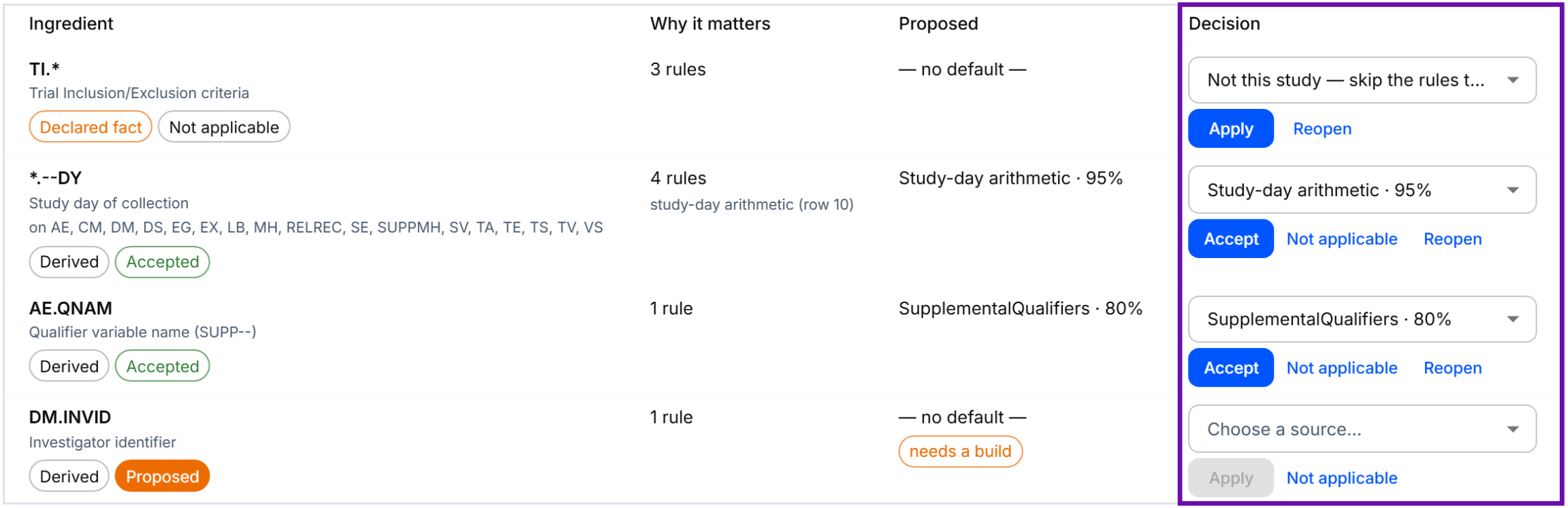
methods



mapping plan

After this section the generator has a rulebook written in its own terms, the methods to derive what the rules ask for, and the mapping plan that will turn its output into the submission datasets.

Missing Ingredients UI: The system proposes. A human decides.



Taken alone at ≥ 80 % confidence: 5 per study · marked not applicable to the study: 70 on VXH-301, 81 on VXH-501 · declared facts referred to a person: 4 per study, then 1 (the inclusion criteria) · invented: 0

One step-log row per decision — the audit trail the validation package already names. Counts from the ingredient pass on copies of both studies, 2026-09-12.

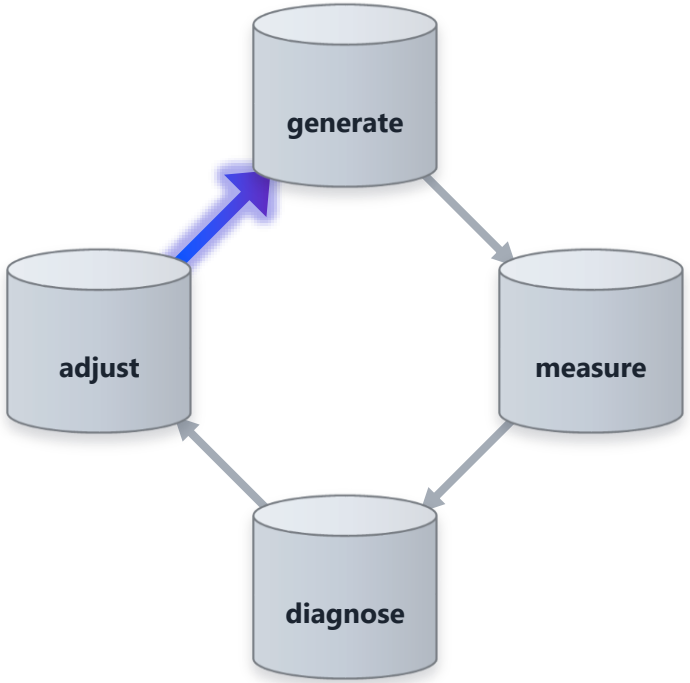
Setup → Fill missing ingredients, VXH-301, 2026-09-12 — one row asked a person, one it took at 95 %, one it refused to invent.

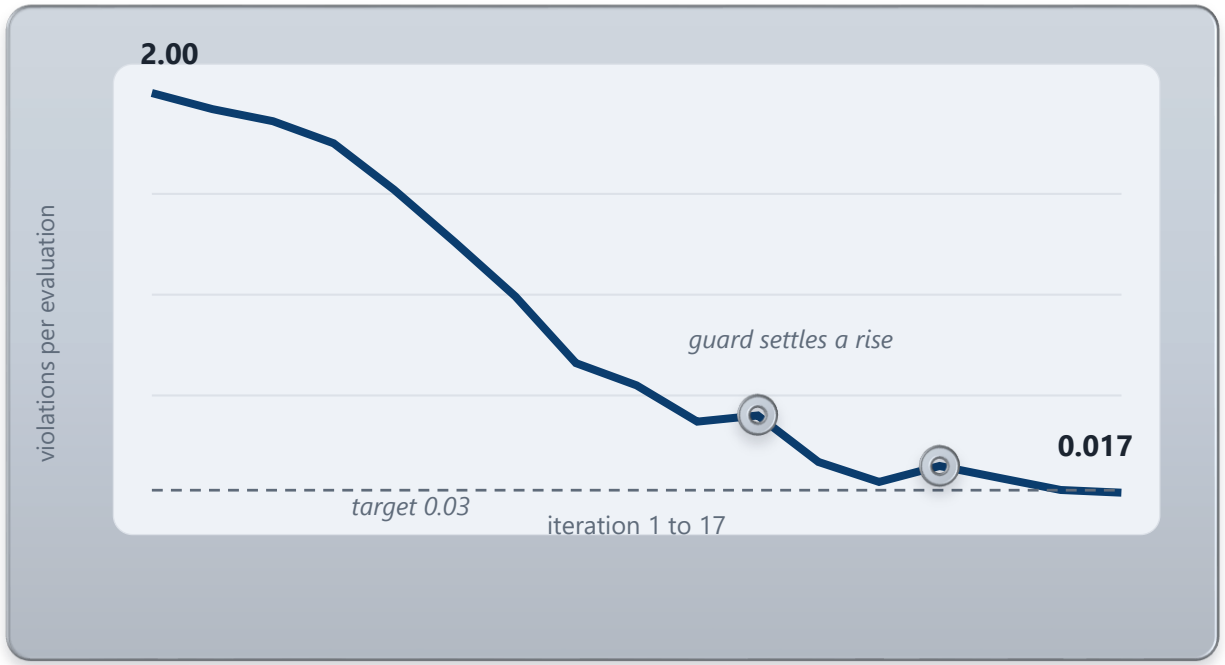
A rule is read backward into a generation constraint in one of four ways

1	**Solve**	a bound, an ordering, a drift limit or a conditional	becomes a distribution, a generation order or a fill rule, with the residual known before a row exists.
if not			
2	**Search**	a rule that relates records, with no closed form	is met by a measured search inside generation (slide 13).
if not			
3	**Interpret**	a rule written in prose	is read by a language model and used only above a confidence threshold.
if not			
4	**Refer**	a rule no generation parameter can satisfy	is handed to a person, with the reason.

On the sponsor-format demo study VXH-401, 33 of the sponsor's 52 edit checks were solved outright and 19 were referred; of the 784 CORE-derived rule instances the inline checker holds, 47 became constraints.

LLM individual rule optimizer iterates, diagnoses and adjusts until the rule violations are minimized to acceptable thresholds





On a fixed relational test case, violations per evaluation fell from 2.00 to 0.017 in seventeen iterations. On VXH-401 the loop confirmed convergence in a single pass, because the solve tier had already met the rules.

- The step size shrinks as the loop converges (learning-rate decay).
- A guard notices when the violation rate oscillates and settles it.
- The loop stops on the budget, on convergence, or when no progress is made within a patience window.

Fixed relational fixture, one captured run; VXH-401 run of record, 2026-09-13.

Seeding the synthetic data generator with realistic clinical models

11 Apply coherence distributions

Fits a distribution and a longitudinal coherence model to every numeric field with a range, so values are realistic and change over visits the way measurements do.

12 Apply CT weighted distributions

Gives coded fields realistic mixes (severity, outcome, action taken) and binds very large codelists to a clinical subset instead of sampling uniformly.

13 Attach standard VLM dictionaries

Binds value-level dictionaries so adverse events, medications and history are coded coherently across term, decode and body system.

14 Standardize event logs

Makes event forms produce distinct, coded events emitted once, rather than the same event repeated at every visit.



After this section the generator draws from models a clinician would recognize, inside the ranges and vocabularies the rules allow.

Generate ‘raw’ data, execute transformation, and score final datasets

15 Generate

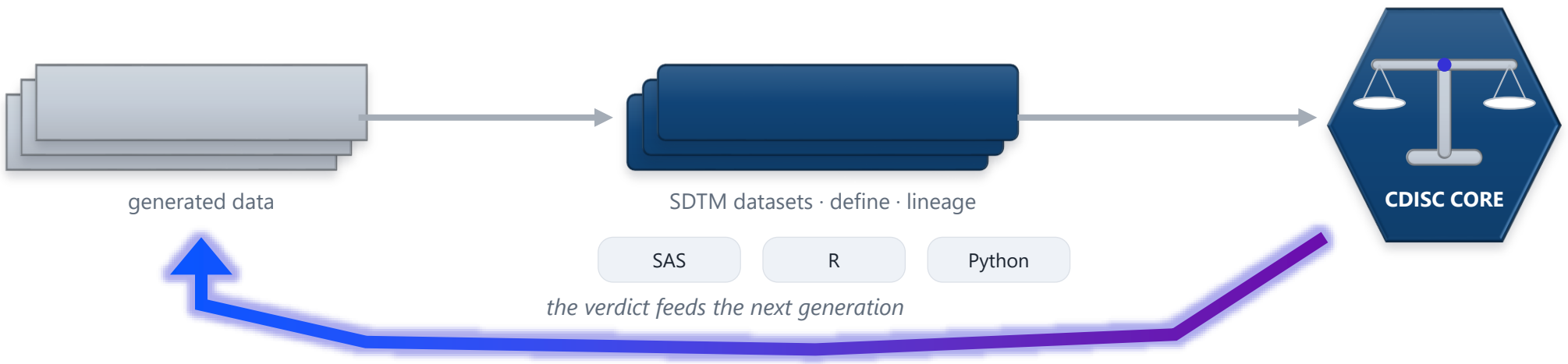
Produces the study's data from the plan: constraints from section 3, models from section 4, with the search loop of slide 13 inside.

16 Build and execute code bundle

Assembles the SDTM code bundle from the mapping plan (SAS, R or Python), runs it on the generated data, and writes the submission datasets, the define and the lineage.

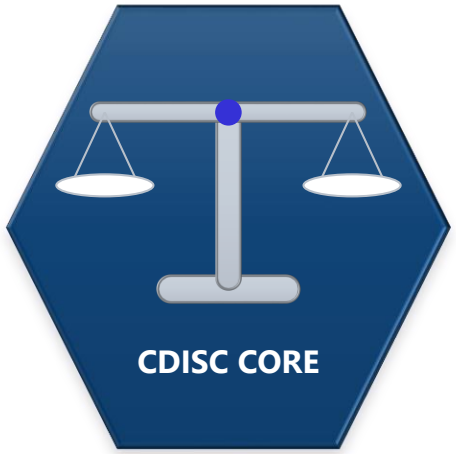
17 Score with CDISC CORE

Submits the executed bundle to the official engine and records the verdict for every rule.



The deliverable the referee judges is the same deliverable a sponsor would submit: the executed datasets, not the generator's working tables.

CDISC CORE is the final arbiter, and its verdict feeds the next generation



0.17.1 · SDTMIG 3.4 · 430 rules

CDISC's own open-source conformance engine, version-pinned, executed on the submission datasets the bundle produced.

passed

of 430

A passed rule confirms its constraint.

issue reported

of 430

An issue names the variable and the rule, and section 3 re-derives the constraint or section 4 adjusts the model.

not evaluable

of 430

A rule not evaluable names a column the deliverable does not yet carry, and the plan gains it.

into the next generation

The referee is the same engine a regulator's reviewer can run. Nothing in the loop grades its own homework.

Two studies: Optimizing via CORE feedback improves rule pass rate



Copies of VXH-301 and VXH-501; points 1 to 3 measured 2026-09-12 on CORE 0.16.0; points 4 and 5 re-measured 2026-09-13 on CORE 0.17.1 after fixes DF-135 to DF-139 (plan §22; VXH-301 CORE reports 54 and 55, bundle bd703bae…; VXH-501 reports 56 and 57, bundle 5e2d3b4e…).

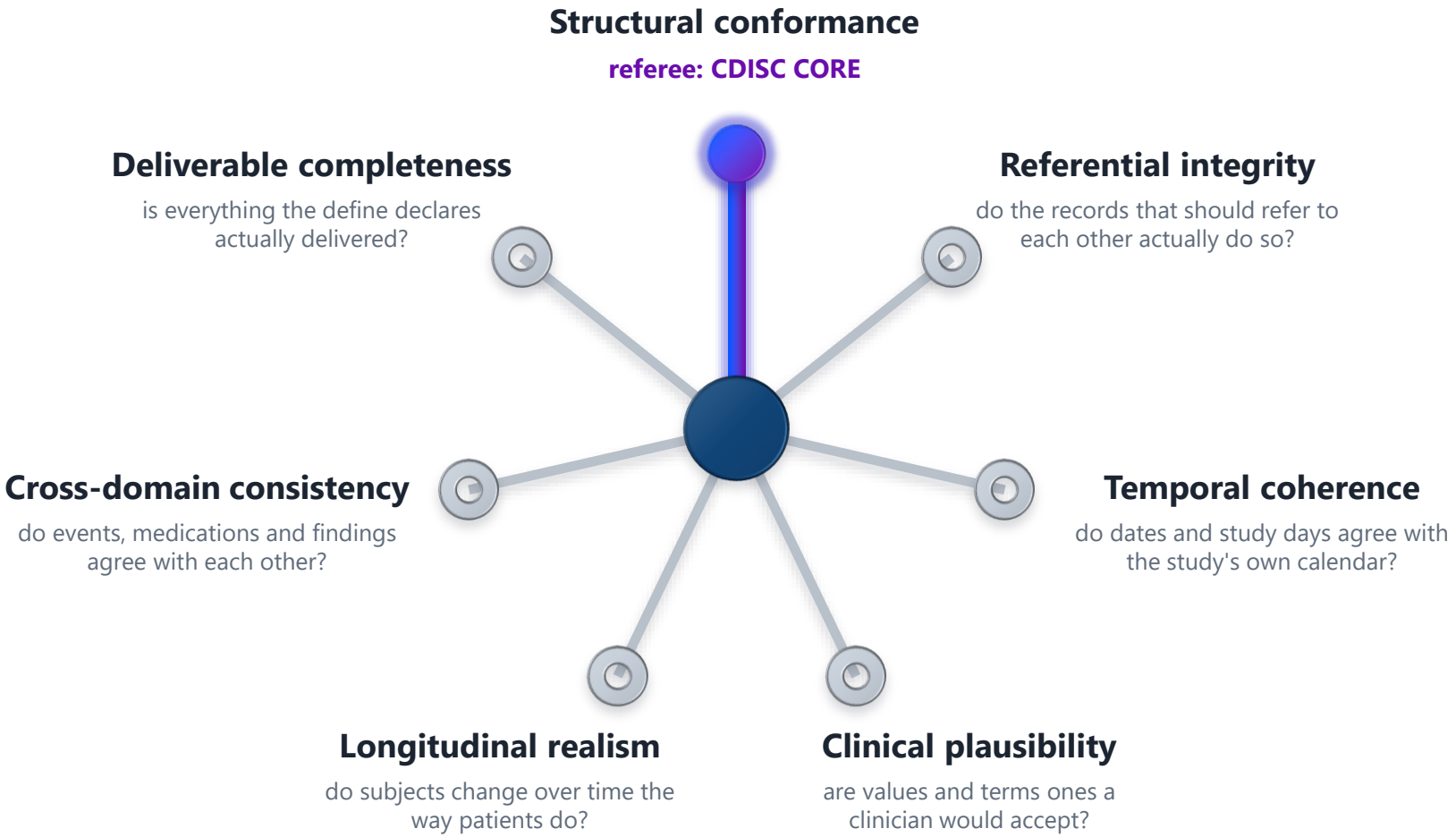
On a sponsor-format study the same loop satisfied the sponsor's own checks and 318 official rules

Study	Official CORE rules passing, of 430	Datasets delivered	The sponsor's own edit checks
VXH-401 · sponsor-format demo, run of record	**318**	18 of 18	33 of 52 solved outright; violations across the 52 fell from 52 to 2
VXH-301 · demo, end state	**320**	17	—
VXH-501 · demo, end state	**320**	17	—

VXH-401 was imported from a CRF authoring system in the sponsor's own format; the loop read the sponsor's edit checks as well as the standard's rules.

VXH-401 run of record 2026-09-13 on CORE 0.17.1, CORE report 55, bundle run 0c9c9886…; VXH-301 and VXH-501 end state 2026-09-13 (plan §22), CORE reports 55 and 57.

What defines synthetic data quality?
The same loop can learn from the other six assessment axes



- The loop needs a scorer for each axis the way it has CDISC CORE for the first.
- The methodology draft defines the other six and how each is scored.
- A scorer that reads a dataset and reports an axis can sit exactly where CORE sits in the ring.

Conformance is the first axis because it has an independent referee today. The loop is built to take six more.

Methodology v0.1 draft, section 4.

Why define synthetic data quality?

Synthetic data quality = what is verifiable before first patient-in

1	**Edit checks and rules**	Test every edit check and derived rule on a clean twin and a defective twin before the database exists.	needs a clean twin that is truly clean, and defects that are the only defects	**conformance · referential integrity**
2	**Analysis programs**	Write and dry-run the ADaM and table programs on data shaped like the real study.	needs real structure: visits, baselines, reference ranges, treatment arms	**temporal coherence · longitudinal realism**
3	**Mappings and the define**	Validate the SDTM transformation and the define on the executed bundle, not on the generator's tables.	needs breadth: a transformation checked on thin data passes trivially	**deliverable completeness**
4	**External data deliveries**	Rehearse vendor files on the vendor's cadence and layout, from central lab to ECG to imaging.	needs files that look and behave like the vendor's, joined to the CRF data	**cross-domain consistency**
5	**Blinding and disclosure**	Rehearse unblinding risk and masked exports before real subjects exist.	needs plausible treatment effects and a masking layer that survives inspection	**clinical plausibility**

Poor synthetic data fails quietly: the check passes, the program runs, the file loads, and none of it would have happened on real data. Each rehearsal depends on a different property, which is why quality is seven measures and not one score.

cdisc®

Use cases as shipped (docs/Feature & Options Catalog.md); the quality properties are the seven axes of slide 3.

Read the method, score your own data, and help define synthetic data quality

- The methodology draft is open for comment.
- The scorer runs in your browser on SDTM datasets, nothing is uploaded.
- Test the scorer, give your feedback, shape the next version.



clinventive.com

Questions?



Thank You!

