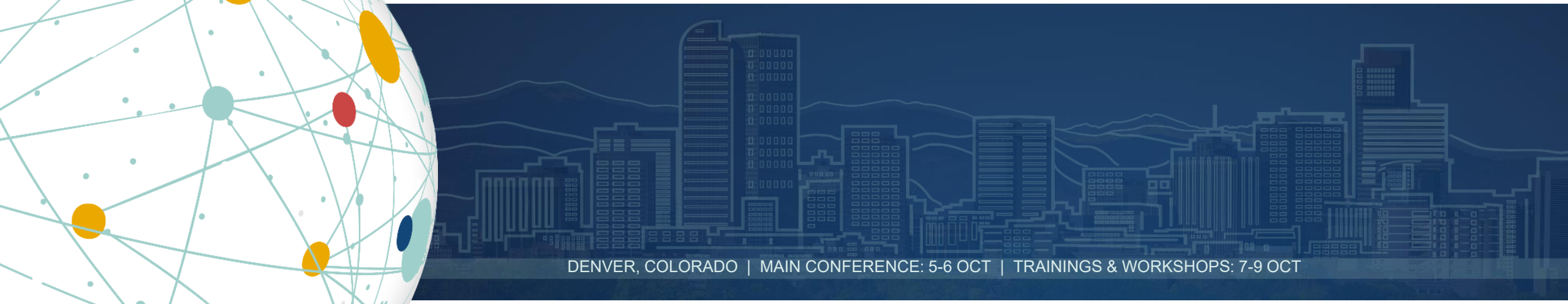


cdisc®

2026 US Interchange

THE FUTURE IS CONNECTED: STANDARDS AND AI POWERING DIGITAL TRANSFORMATION



Unlocking eSource Potential

A Call for Ecosystem Alignment on AI/NLP Standards

Jo-Zahn Baxter

Speakers



Jo-Zahn is a seasoned leader in clinical research and a pioneer in clinical innovation with nearly two decades of experience in clinical operations. She is adept at supporting R&D IT teams in AI-driven clinical trial technology selection, implementation, and global scale-up. Jo-Zahn focuses on technologies that accelerate clinical trial workflows, improve data quality, and reduce site burden.

Jo-Zahn Baxter

Director, Digital Transformation and R&D

Axtria

From isolated innovation to trusted infrastructure



1 What we learned

Applying AI/NLP to unstructured EHR text in a live setting



2 Where there’s room to grow

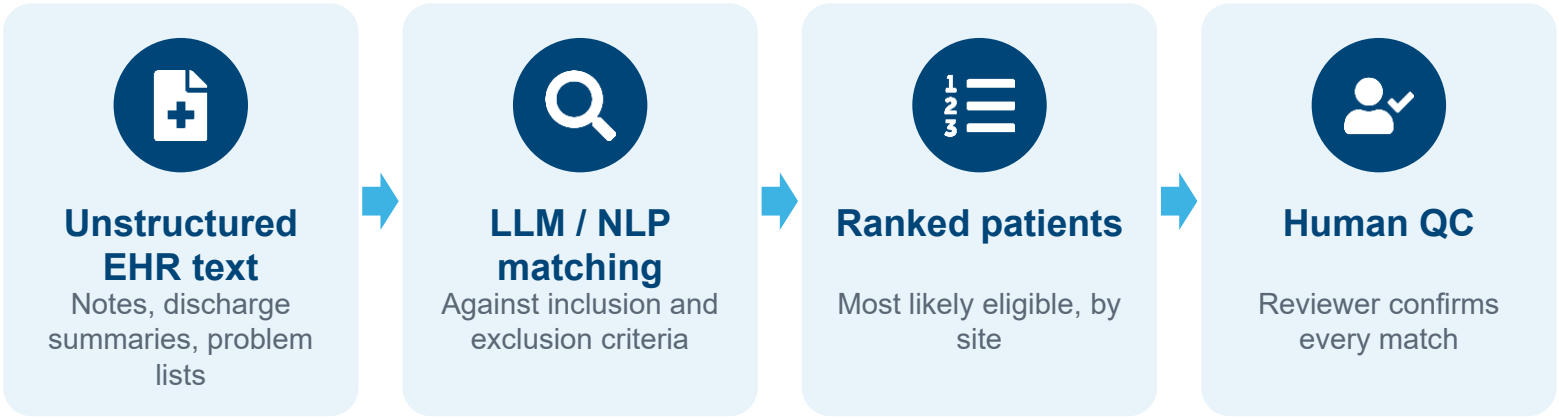
Regulatory guidance and data standards today



3 What a common standard could include

And how we might build it together

Using AI/NLP for eScreening



Part of a live clinical research workflow.

Notes carry the ***why***: the patient journey and context that structured fields often don't capture.

~20 min

to QC a ranked shortlist, versus hours of manual chart review

Faster and more consistent, surfacing candidates for the study team to consider.

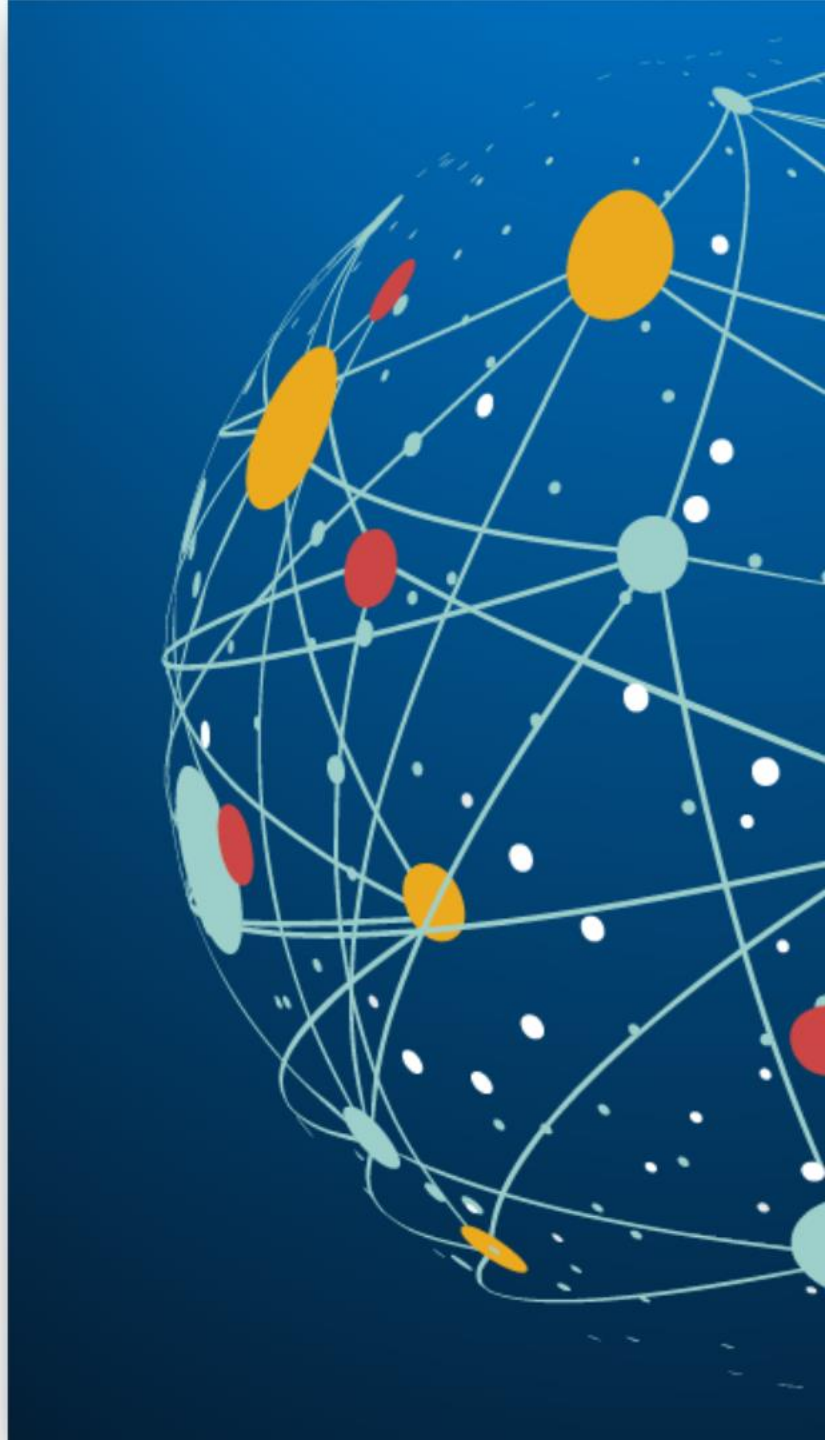
Where trial-relevant data lives

~80% of clinical information is unstructured: notes, reports, narratives

How sites capture source data today

- **Paper** — scanned or not
- **Paper + non-EHR DDC**
- **Paper + EHR**
- **DDC via EHR**
- **Standalone DDC** — discrete fields
- **Ambient voice** — emerging

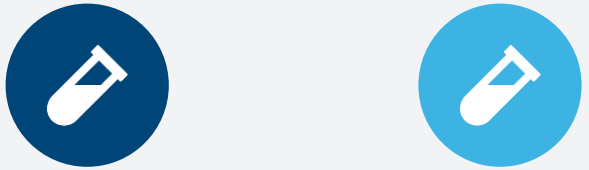
Source: Applied Clinical Trials (Sept 2025), citing Tufts CSDD. DDC = direct data capture.



EHRs were built for care and billing, not trials

IN THE EHR	WHAT THE TRIAL NEEDS
ICD-10	**MedDRA**
Codes exist because payers require them	
RxNorm / NDC	**WHODrug**
Codes exist because pharmacies need them	
"Hypertension"	**Onset? Severity? Attribution?**
Problem lists carry no trial context	

Same test, different codes



Hospital A

Hemoglobin: LOINC code A

Hospital B

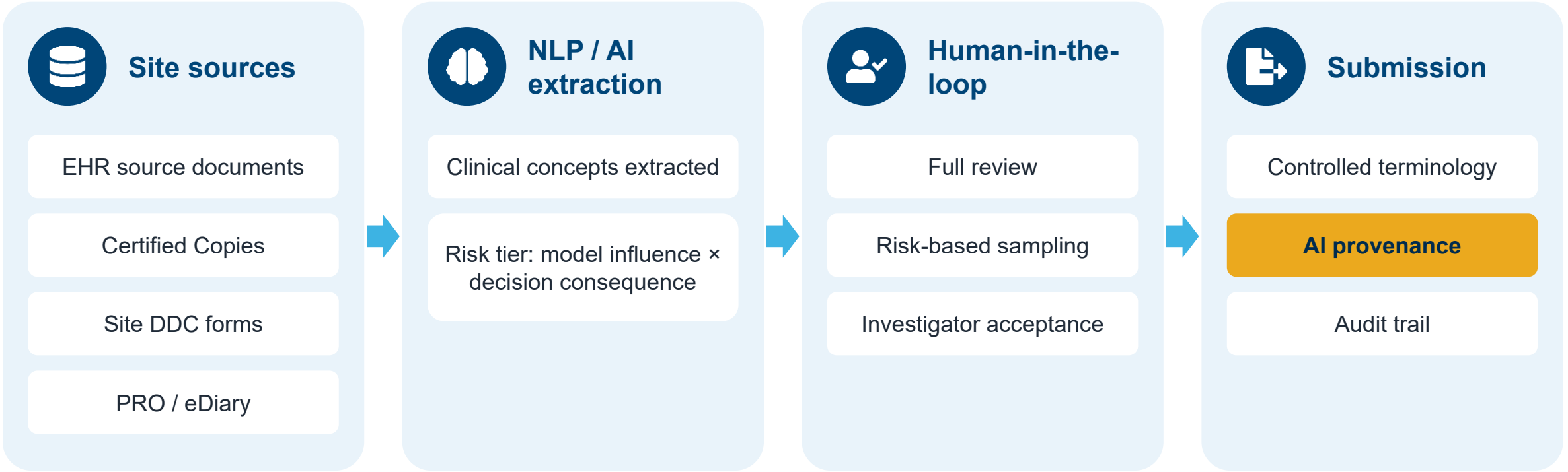
Hemoglobin: LOINC code B

Both codes are correct. Query one, and you miss the other hospital entirely.

EHR data does what it was designed to do. Moving from EHR to SDTM requires guidelines

The data flow is in place; provenance needs alignment

Structured FHIR path: 16.9% of data points, mostly labs and vitals. Unstructured notes are still largely out of reach.



The opportunity: align on what “AI provenance” contains, so teams can build on common ground.

Regulators have given us a vocabulary

Sep 2024

EMA reflection paper

Bias, training data, GxP traceability

Jan 2025

FDA draft guidance

Context of use, model risk, credibility

Jan 2026

Joint FDA–EMA principles

Ten principles of good AI practice

What's still evolving

- How AI/NLP systems should extract trial data and demonstrate accuracy
- How to encode AI-extracted data in a submission dataset
- A home for unstructured text in current standards

A shared approach will serve all of us better.

Today, audit trails vary by submission, and reviewers evaluate each approach on its own terms.

A starting point: seven AI provenance fields

1

Model ID & version

Unique and immutable

2

Training data lineage

Domain, time period, demographics, languages

3

Confidence score

Per element; recall, precision, error rates

4

HITL review flag

Not reviewed · accepted · modified · rejected

5

Reviewer ID

Pseudonymized, for every HITL record

6

Source-text pointer

eSource, certified copy, human-entered, NLP-extracted

7

Extraction timestamp

Per regulatory requirements

A practical starting point.

cdisc®

What shared infrastructure makes possible

Working independently

– Each sponsor develops its own validation approach

– Compliance costs are carried organization by organization

– First submissions carry more uncertainty

Working together

Reference validation datasets · annotation guidelines · performance benchmarks

- **Compliance cost spread across the industry**
- **First-submission risk becomes a first-mover advantage**

Precedent: SDTM moved from a voluntary option to an FDA-required standard through coordinated industry work.

cdisc®

An independent operational expertise to bring everyone together

Chairs working groups

and manages timelines

One industry voice

for regulatory engagement

Maintains shared artifacts

datasets, guidelines, benchmarks

Turns standards into SOPs

at site, sponsor, and vendor level

Sites are already collecting data in ways our submission standard wasn't designed to handle.

Binding standards are being developed. Taking part now helps shape a standard that works for everyone.

cdisc®

Questions?

