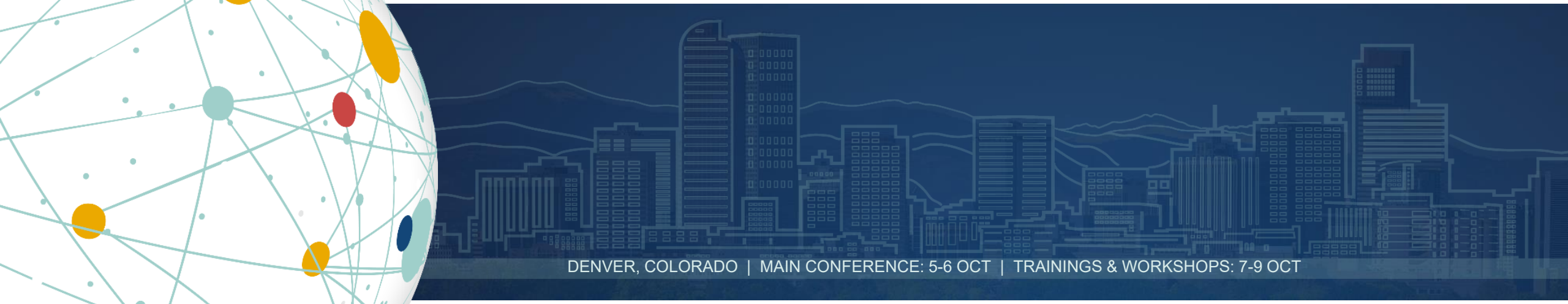


cdisc®

2026 US Interchange

THE FUTURE IS CONNECTED: STANDARDS AND AI POWERING DIGITAL TRANSFORMATION



Validating the Site ↔ Vendor ↔ Sponsor Data Workflow

Challenges and Opportunities in Clinical Data Validation

Heather Marino

Speakers



Heather is a senior leader in clinical strategy, research operations and trial execution bringing over 17 years of experience across clinical trials from early to late phase investigations. She brings strong experience in study concept design, protocol development, ICF design, data-informed site selection, AI-driven patient-to-trial matching, and data management. Heather has unique insights and experience in clinical innovation, having supported the development and led deployment of GxP and GDPR compliant tools leveraging Natural Language Processing (NLP), Machine Learning (ML) and GenerativeAI to streamline and automate clinical data workflows.

Heather Marino, MA, MLA

Senior Director, Digital Transformation and R&D

Axtria, Inc.

Following the data upstream

01 **Upstream data & the vendor market**
How sites capture source, the formats it arrives in, and who sells the tools

02 **Sponsor validation**
Sorting site tools into a menu and connecting them to one integrated stack

03 **Validation challenges**
Six places eSource adoption slips, and how to get ahead of each

04 **The business value**
Where eSource removes cost from a trial, and how to calculate it

Clinical research sites are increasingly adopting their own technology…

…creating opportunity and friction for data integration to sponsor systems

Patient-generated

eCOA · diaries · wearables

Created by the patient, often at home, on a device the site didn't build.

Direct data capture

Point-of-care entry

The research record is created at the moment of care, with no paper step.

eSource / non-CRF

Labs · imaging · genomics

Specialty systems built for clinical work, not for research.

EHR as source

Standard-of-care record

The site's medical record, used as the source for research data.

All four are source.
Same regulatory obligation: ALCOA+

A	L	C	O	A	+
Attributable	Legible	Contemporaneous	Original	Accurate	

Categories align to TransCelerate's eSource classifications. EHR-to-EDC is a subset of eSource, not a synonym for it.

cdisc®

Same fact, different shapes

Follow one blood test result as it leaves four different sites

ONE FACT

13.2

g/dL

A patient's hemoglobin result

A message from the lab system
Labeled fields a computer can read right away
Ready to use

A box typed into a form
Someone entered it at the point of care
Mostly ready

A scanned PDF of the lab report
A picture of text that has to be read out first
Must be extracted

A line in the doctor's note
“Hgb stable at 13.2”, buried in a paragraph
Hard to find

ONE STANDARD RECORD

Test	**Hemoglobin**
Result	**13.2**
Unit	**g/dL**
Date	**Visit date**
Patient	**Study ID**

Same fact, four shapes. It has to land as one standard record, every time.

cdisc®

The site → vendor → sponsor data workflow is broken

THE VENDOR MARKET what's for sale

EHR platforms & HIEs
Hold the data; not built for research workflows

Middleware integrators
Move EHR & eSource data to sponsor EDC; mixed structured / unstructured access

EDC platforms
Pull data into their own stack

Direct data capture eSource
Phase 1 CRUs, independent sites & networks; sponsor connectivity still maturing

≠

THE SITE LANDSCAPE what sites actually run

1 **Traditional paper** No structured capture at the source

2A **Hybrid: paper + non-EHR eSource** A mix of structured and unstructured access

2B **Hybrid: paper + EHR** EHR holds some, not all, source data

3A **eSource via EHR** Sponsor connectivity still maturing

3B **Standalone eSource** Site-owned system, not yet networked

No single vendor spans geography, data structure, and site reach, all at once.

cdisc®

How do sponsors validate different tools and datasets?

1 · Site intelligence

Sponsor and site walk through the technology the site has and how it is actually used, via survey.

None of this exists in a public database.

2 · Curate

A decision tree places each site in one of five curated profiles: the bring-your-own-technology (BYOT) menu.

1 2A 2B 3A 3B

3 · Connect

Each path is routed into the sponsor's ecosystem, mapped to what the sponsor already runs.

The site doesn't change its tools.

Understand what each site already has → curate it into the BYOT menu → route each path to the sponsor

The goal: one integrated stack, operating across sites & trials

WHATEVER THE SITE RUNS

- 1 **Paper**
- 2A **Paper + eSource**
- 2B **Paper + EHR**
- 3A **eSource via EHR**
- 3B **Standalone eSource**

▼ *Each site uses one or more of these pathways* ▼

Paper → Source Data Extraction
Sites upload PDFs; AI parses them into structured data

eSource / EHR → Sponsor
One standardized landing pattern; standalone eSource skips EDC

EHR → EDC
Structured + unstructured data populate eCRFs

Common ingestion layer
One landing pattern, mapped to standards, with end-to-end traceability

Clinical data lake

EDC

The site picks from a menu; the sponsor takes on the integration complexity.

cdisc®

Expecting the unexpected

Six places eSource adoption typically slips, and how we get ahead of each one

1 Define & redefine success metrics

Metrics set once and never revisited go stale.

→ **Define KPIs early; adjust as the program evolves.**

2 Align clinical & technical stakeholders early

Stakeholders brought in late have no runway to prepare.

→ **Bring them in early to build adoption collateral.**

3 Expect a non-linear site adoption curve

Sites start, pause, and stop. That's normal.

→ **Stay close to site-facing staff to catch drop-off.**

4 Rebuild, don't reuse, the RACI

Roles shift once EDC transcription decreases.

→ **Define a new RACI specific to eSource processes.**

5 Design studies for ingestion, not just entry

Forms built for manual entry don't always ingest cleanly.

→ **Refine eCRF design for the future state.**

6 What validation is enough for a site's source system?

The US has an EHR standard (ONC); Europe does not.

→ **Push for a shared industry standard.**

cdisc®

The clinical operations value of eSource

Where eSource removes cost from the trial, and how we propose to calculate value

	Data entry	Source data verification	Queries & rework	Investment
TODAY	Site enters data in the EHR, then re-keys it into EDC	Every data point checked by hand, by site and monitor	Typing errors trigger queries and rework	**Offset, not a saving** **CapEx** Build + validate **OpEx** Run + onboard sites
	▼ **Coordinator time saved**	▼ **Verification time saved**	▼ **Queries avoided**	−
WITH eSOURCE	**Entered once; a validated extract moves it**	**Extracted points need far less manual checking**	**Fewer errors at the source, fewer queries**	
HOW WE CALCULATE IT	*Points/patient × % extracted × time to enter × hourly rate*	*Points/patient × % extracted × time to verify × hourly rate*	*Points/patient × % extracted × error rate × cost per query*	*CapEx + OpEx*

Net value = (entry + SDV + query savings) − (CapEx + OpEx) per study and at portfolio scale

Sensitivity: every saving scales on % EHR data extracted. Pressure-test that input first.

cdisc®

The tools will keep multiplying.
The obligation won't change.

Questions?

