2026 US Interchange
THE FUTURE IS CONNECTED: STANDARDS AND AI POWERING DIGITAL TRANSFORMATION
DENVER, COLORADO | MAIN CONFERENCE: 5-6 OCT | TRAININGS & WORKSHOPS: 7-9 OCT

From Repository to Intelligent Platform: Connected Standards and AI for RWE eTMF

Chris Huang, Director Product Management, Oracle Health & Life Sciences
Tom Stonehouse, Senior Consulting Practice Manager, Oracle Life Sciences

2026 CDISC US Interchange Program



Speakers

Chris Huang

Director Product Management
Oracle Health & Life Sciences



Tom Stonehouse

Senior Consulting Practice Manager
Oracle Life Sciences



Disclaimer and disclosures

- **Views**

  The views and opinions expressed in this presentation are those of the authors and do not necessarily reflect the official policy or position of CDISC.

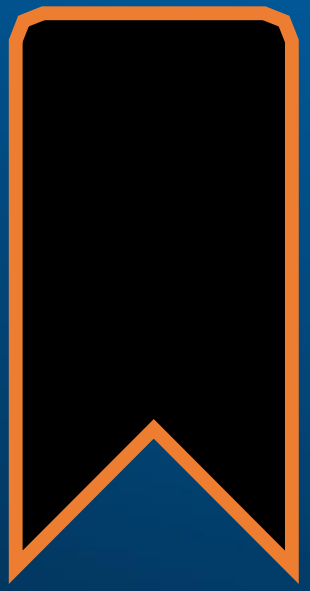
- **Disclosures**

  The authors have no real or apparent conflicts of interest to report.



Agenda

1	2	3	4
RWE Context	Data Supply Chain	AI & Oversight	Demo



RWE Context: Broader than a study type

RWE is not synonymous with non-interventional research

Observational RWE	Routine-care data, registries, claims, EHRs, labs, devices, PROs, or public health sources are analyzed without assigning treatment.
Pragmatic or point-of-care trials	Randomization may occur, but the trial is embedded in routine care and may use real-world data for outcomes or follow-up.
External control arms	A single-arm or interventional study may use RWD from registries, natural history, EHR, claims, prior trials, or expanded access as external context.
Hybrid evidence package	Investigational-product, safety, protocol, vendor, data-source, and analysis-lineage documentation can coexist in the same RWE evidence story.



Who uses RWE and what they want

Clinical Development	Natural history, Disease progression, Cohorts, Endpoints, Trial feasibility, Site/patient identification, External controls	Can we design a better development program?
Medical Affairs	Scientific exchange, Publications, Treatment Patterns, Patient journeys, Evidence gaps, KOL engagement	What is happening in real clinical practice?
HEOR	Burden of illness, Utilization, cost, Comparative effectiveness, Persistence, adherence, Quality of life	What is the clinical and economic burden?
Market Access	Value dossiers, Payer evidence, HTA support, Coverage decisions, Outcomes-based agreements	What evidence will payers trust?
Commercial	Patient segments, Treatment pathways, Adoption barriers, Regional variation, Competitive dynamics	Where is unmet need and adoption friction?
Regulatory / Safety	Label expansion, Postmarketing commitments, Safety context, Dosing changes, External controls, Signal evaluation	Can RWE support or contextualize a regulatory decision?

External control arms make traceability visible

The use case

Use external RWD to contextualize outcomes for a treated cohort when randomized concurrent controls are difficult, unethical, or infeasible.

The data sources

Natural-history studies, registries, EHRs, claims, prior trials, expanded access, or curated specialty datasets.

The risk

Differences in eligibility, index date, follow-up, outcome capture, standard of care, missing data, and confounding influences can weaken credibility.

The TMF need

Document source provenance, cohort comparability, endpoint definitions, data cuts, transformations, sensitivity analyses, and decisions.

The platform need

Connect the data source, study design, metadata, analysis package, review history, and inspection response narrative.

The takeaway

An ECA is not just a dataset. It is a defensible chain of evidence that must be inspectable and reproducible.

RWE data supply chain

From clinical event to RWE data product

1 Care or administration

- ❑ Clinical care
- ❑ Reimbursement
- ❑ Registry submission
- ❑ Lab testing
- ❑ Pharmacy dispensing
- ❑ Device use
- ❑ Public health reporting

2 Source aggregation

- ❑ Health system
- ❑ Payer
- ❑ Registry
- ❑ Lab
- ❑ Pharmacy
- ❑ Device platform aggregates routine data.

3 Vendor curation

- ❑ Data is de-identified
- ❑ Normalized
- ❑ Mapped
- ❑ Linked
- ❑ Curated
- ❑ Quality checked
- ❑ Packaged.

4 Sponsor access

- ❑ Sponsor receives an extract
- ❑ Secure workspace
- ❑ API
- ❑ Dashboard
- ❑ Analysis-ready dataset
- ❑ Aggregate output.

5 Evidence use

- ❑ The data inform cohorts
- ❑ Endpoints
- ❑ ECAs
- ❑ Disease progression
- ❑ Safety
- ❑ Market access
- ❑ Publications

How vendors receive RWE data

Secure file transfer

SFTP, managed file transfer, encrypted portals, or recurring file drops for EHR, registry, claims, lab, or pharmacy feeds.

Cloud landing zones

Encrypted object storage or controlled exchange areas for large recurring data cuts and high-volume feeds.

Database feeds

Secure views, extracts, replicated tables, or data warehouse interfaces for ongoing source-to-vendor ingestion.

API / FHIR exchange

Vendor-specific APIs or FHIR-aligned APIs, including population-level bulk export where supported.

Secure workspaces

Data remain in an enclave or clean room; approved users query or analyze under controlled access.

Aggregated outputs

For restricted data, the vendor may provide counts, tables, figures, dashboards, or study reports instead of row-level data.

Traceability: what should follow each data cut

Standard package
Data dictionary, metadata guide, source description, data quality report, version/refresh notes, and permitted-use documentation.

Study evidence package
Cohort logic, endpoint definitions, transformation specs, code lists, lineage, and data-quality checks tied to analysis datasets.

Audit support package
Vendor-held transfer logs, ETL logs, source-level evidence, inspection response support, retention commitments, and right-to-audit terms.

Contracting implication
Sponsors should not assume full traceability logs are included by default. Requirements belong in the DUA, SOW, quality agreement, or regulatory-support terms.

TMF implication
Track what was received, what is vendor-held, what is available on request, and what source-verification limits apply.

Inspection test
Can the team trace a final evidence output back to a specific data source, data cut, transformation, and review decision?

RWE value is unlocked only after the data work

Speed depends on whether data are already standardized, curated, traceable and analysis-ready.

The bottleneck is often not access to data — it is turning heterogeneous healthcare data into defensible evidence.

6–8 wks

simple RWE studies

EMA notes simple RWE studies may be completed in weeks; complex studies can take several months.

24%

failed validity checks

In one standardized safety-RWE program, one quarter of analyses failed objective diagnostics and were stopped.

11.5% → 77.3%

traceability improved

A JAMA Network Open study found traceability improved materially when advanced methods and multiple sources were applied.

Implication for eTMF:

Capture the evidence trail from source and data cut through quality checks, transformations and decisions — not just the final report.

The Cures Act changed the operating environment

Regulatory RWE priority

The Cures Act required FDA to evaluate how RWE can support new indications for approved drugs and postapproval requirements.

Interoperability push

ONC rules advanced electronic health information access, exchange, standardized APIs, and information-blocking expectations.

API / FHIR momentum

FHIR-based exchange and bulk access models make scalable real-world data movement more plausible.

Not a quality shortcut

Better data access does not make data automatically fit for purpose, complete, reproducible, or inspectable.

Documentation impact

Sponsors need provenance, permissions, cohort logic, endpoint definitions, quality checks, and transformation lineage.

eTMF implication

The platform must show not just that data were available, but that evidence was relevant, reliable, traceable, and permitted.

AI and unstructured RWE

Unstructured data is a major RWE challenge

Before modern AI

- ❑ Manual chart abstraction
- ❑ Case report forms / registry abstraction
- ❑ Medical coding
- ❑ Rule-based text search
- ❑ Dictionary / ontology-based NLP
- ❑ Clinical adjudication
- ❑ QC sampling and double review

With AI assistance

- ❑ Classify notes and source documents
- ❑ Extract clinical concepts, dates, outcomes, and metadata
- ❑ Suggest mappings and endpoint variables
- ❑ Detect missingness or conflicts
- ❑ Summarize provenance and lineage
- ❑ Route low-confidence items to review

How AI can organize RWE data upstream

Source classification

Identify EHR, claims, registry, lab, pharmacy, device, PRO, mortality, or public-health source type.

Metadata extraction

Extract study, source, data cut, coverage period, vendor, permitted use, region, version, and refresh date.

Terminology mapping

Suggest mappings across ICD, CPT, RxNorm, LOINC, SNOMED, MedDRA, OMOP, FHIR, or vendor models.

Record linkage support

Assist duplicate detection, longitudinal patient journeys, and cross-source linkage where permitted.

Quality monitoring

Flag missingness, outliers, unexpected source shifts, coding drift, inconsistent units, and metadata conflicts.

Lineage generation

Summarize how source data became curated variables, datasets, outputs, and inspection-ready evidence.

Classifier workflow: AI-assisted review

1 Synchronize context

- ❑ Select study and retrieve artifact metadata
- ❑ Capture file, source TMF, status, and study context
- ❑ No AI decision yet

2 Prepare & analyze

- ❑ Download and convert file to text
- ❑ Chunk content and create embeddings
- ❑ GenAI analyzes useful chunks and metadata conflicts

3 Recommend % decide

- ❑ Return TMF candidate, confidence, rationale, and alternatives
- ❑ Low confidence requires manual review
- ❑ Accept, override, or manually classify

Traceability thread: capture the evidence behind every recommendation and decision

- Source artifact + data cut
- Model/rule + reference version
- Chunk evidence + rationale
- Human decision + reason

Pre-filing and post-filing solve different quality problems

Pre-filing assist

- ❑ Trigger at upload
- ❑ Extract text and filename cues
- ❑ Suggest folder and metadata
- ❑ Apply naming and required-field checks
- ❑ Show confidence, rationale, and examples
- ❑ User accepts or overrides before commit

Post-filing sweep

- ❑ Run against existing filings
- ❑ Compare content, metadata, and expected content
- ❑ Flag anomalies and low confidence items
- ❑ Prioritize review queues
- ❑ Track remediation and aging
- ❑ Support continuous inspection readiness

Responsible AI controls for eTMF

- **Representative data**
  Train/test across document types, phases, vendors, regions, languages, templates, and sparse classes.
- **SME labels and thresholds**
  Use SME-reviewed labels, adjudication, gold-standard tests, and high-precision thresholds.
- **Explainable human review**
  Show confidence, rationale, evidence cues, alternatives, and review flags; capture reason-coded decisions.
- **Drift and performance monitoring**
  Monitor precision, acceptance, low-confidence trends, overrides, metadata drift, and performance slices.
- **Production controls**
  Validate automation and file conversion; maintain versioning, audit trail, change control, and revalidation.



eTMF Smart Classifier Demo

Blueprint: intelligent, interoperable RWE eTMF

The inspection-ready story should be assembled continuously

What data were used?	Sources, source type, data providers, repositories, data cuts, coverage period, refreshes, and ownership.
Was use permitted?	DUAs, consent/legal basis, privacy review, retention terms, redisclosure limits, and publication/submission permissions.
How did raw data become evidence?	Extraction, de-identification, mapping, linkage, curation, endpoint derivation, analysis datasets, and quality checks.
Which version supported each output?	Trace tables, code lists, transformation versions, data cut IDs, TLFs, publications, and submission evidence packages.
Where did AI assist?	Model/rule version, inputs, evidence snippets, confidence, human review, override rationale, and monitoring results.



Modernization roadmap

Start with traceability, prove value, then scale automation

0-90 days: assess

Inventory RWE artifact pain points, vendor data flows, current metadata, refresh handling, and traceability gaps.

90-180 days: standardize

Define minimum RWE metadata, data-cut evidence package, source provenance model, and expected content rules.

Pilot: configure

Pick a high-value RWE study type such as ECA, registry-based research, or longitudinal disease progression.

Pilot: assist

Use AI for classification, metadata recommendation, anomaly detection, and review queue prioritization with human oversight.

Scale: integrate

Connect APIs, FHIR-aligned data context, vendor deliverable tracking, analysis outputs, and inspection response workflows.

Sustain: govern

Monitor quality, drift, overrides, refresh impact, standards updates, and adoption by study, region, partner, and artifact class.

Measure what matters

- **Filing quality**

  Misfile rate, metadata completeness, low-confidence queue size, and correct classification rate.

- **Workflow efficiency**

  Time-to-correction, manual QC workload, review aging, and remediation hours avoided.

- **Inspection readiness**

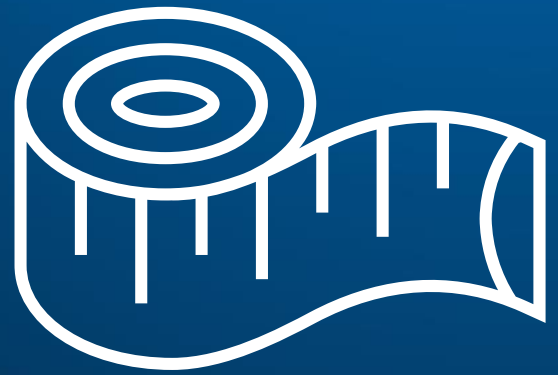
  Trace artifact, data cut, transformation, quality check, decision, and output without late reconstruction.

- **AI performance and operational insight**

  Pending reviews, accepted decisions, average confidence, override reasons, drift incidents, and performance slices.

- **Business impact**

  CAPA/audit trends, inspection-readiness lead time, cleanup effort, cost per 1,000 docs, and review productivity.



Key takeaways

1. RWE is an enterprise evidence asset

Clinical development, Medical Affairs, HEOR, Market Access, Commercial, Regulatory, and Safety use RWE for different decisions.

2. Vendor data creates chain-of-custody obligations

Sponsors need to know what was received, what changed, what is vendor-held, and what is available for audit support.

3. Refreshes create versioned evidence

Every new data cut should be tied to its quality checks, transformations, analysis outputs, and inspection narrative.

4. AI can organize, not absolve

AI can classify, extract, map, detect anomalies, and generate lineage, but human review, validation, and controlled improvement remain essential.

5. eTMF must become connected

The future TMF connects standards, metadata, APIs, vendor deliverables, analysis lineage, and continuous quality monitoring.

6. The goal is trust

Modernization reduces operational burden only when it also improves traceability, reproducibility, decision capture, and inspection readiness.

Thank You!

Questions?

Chris Huang | Tom Stonehouse
2026 CDISC US Interchange Program



Selected references and source material

CDISC Trial Master File resources / TMF Reference Model - https://www.cdisc.org/tmf
FDA Real-World Evidence resources - https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence
FDA Advancing Real-World Evidence Program - https://www.fda.gov/drugs/development-resources/advancing-real-world-evidence-program
FDA Considerations for RWD/RWE to Support Regulatory Decision-Making for Drugs - https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-real-world-data-and-real-world-evidence-support-regulatory-decision-making-drug
FDA EHR and Claims RWD Guidance - https://www.fda.gov/regulatory-information/search-fda-guidance-documents/real-world-data-assessing-electronic-health-records-and-medical-claims-data-support-regulatory
FDA Data Standards for Drug and Biological Product Submissions Containing RWD - https://www.fda.gov/regulatory-information/search-fda-guidance-documents/data-standards-drug-and-biological-product-submissions-containing-real-world-data
ONC Cures Act Final Rule - https://www.healthit.gov/regulations/cures-act-final-rule/
HL7 FHIR Bulk Data Access - https://www.hl7.org/fhir/uv/bulkdata/STU1/export/index.html
Oracle SharePoint eTMF Smart Classifier FDS, pitch deck, AI suitability document, and internal demo link provided by Tom Stonehouse; demo video not directly inspected due connector size limit.

