connect·share·advance

Beyond SDTM/SEND:

Metadata-Driven Data for Deterministic AI

PHUSE Shanghai Single Day Event
Pingping Gong, PointCross Life Sciences
6 Nov 2026

phuse.global

connect·share·advance

Table of Contents

1. The Core Problem

SDTM / SEND: Exchange-oriented standards, not repository foundations

2. Proposed Architecture

5-layer metadata-driven repository model (UDM / MDR / DSL / Computed Signals / AI-LLM)

3. Regulated-AI Guardrails

Reproducible | Auditable | Human-in-the-Loop (FDA TCG 6.2)

4. Real-World Validation

Non-clinical signal-detection evidence & SDTM/SEND data gaps

5. End-to-End Implementation

Unified workflow for Clinical (SDTM-ADaM) & Non-clinical (SEND)

6. Diagnostic Evaluation Criteria

Five key questions for SDTM/SEND-based repository assessment

7. Key Takeaways & Closing

SDTM/SEND: Built for Exchange, Not Repository Mining

EXCHANGE LAYER (SDTM/SEND)

→ Single-study normalization for FDA submission

→ Standard domains: DM, VS, LB, TU, MI

→ Point-in-time snapshot at submission lock

REPOSITORY LAYER (MISSING)

→ As-collected, not post-transformed

→ Protocol & design metadata as entities

→ Computed signals: TR, Concordance, Severity

Why it matters: Most companies today lack the translational data infrastructure to perform cross-study analysis without extensive reprogramming, due to a lack of an invariant, repository data model. With SEND as an example, every study's SEND is transformed independently on a study-by-study basis, so

SDTM/SEND should be the generated output of the data model — not the storage layer.

Metadata-Layered Repository Model Separates Data from Intelligence

connect·share·advance

L5 · AI / LLM LAYER
Assisted report drafting · Human-in-the-loop · Every output traces to signals

L4 · COMPUTED SIGNALS
Treatment Relatedness (TR) · Concordance flags (55 checks) · Dose-response

L3 · DATA SERVICES LAYER (DSL)
Fold-change normalization · Bio-statistics · Group summaries · Incidence counts

L2 · METADATA REGISTRY (MDR)
Protocol & study design · SOPs · Terminology · Thresholds & rules · Templates

L1 · UNIFIED DATA REPOSITORY (UDM)
As-collected: instruments, assays, tablet EDC, cage-side, pathology

→ **SEND / SDTM**
Generated output
Not the storage model

Principle: Raw data at bottom, metadata rules the middle, signals and AI at top.
Each layer independently queryable.

AI/LLM in Regulated Systems: deterministic

Validation extends to traceability (Refer FDA TCG 6.2, June 2026)

1. REPRODUCIBLE

Same data in, same computed result. Every time, no randomness, no drift.

Math-based:

fold-change
p-values
thresholds

2. AUDITABLE

Every AI/LLM output traces back to a computed signal or raw data point. No black-box conclusions.

Rules-based:

55 concordance checks,
TR flags

3. HUMAN-IN-LOOP

AI/LLM drafts text; the Study Director reviews, edits, and approves every published conclusion.

Human-in-loop at every decision point

Why it matters:

- Generative AI/LLM hallucinates — unacceptable in a 21 CFR Part 11 regulated environment.
- AI/LLM assists: deterministic computed signals carry the evidentiary weight.
- The Study Director, not the model, owns the final judgment.

Validated on 100's of Nonclinical Studies
— Signals That Group-Level Statistics Miss

Deterministic concordance checks produce auditable evidence — not probabilistic guesses.

01

Subject-Level Concordance

Single-animal findings surface automatically — even at n=1.

Example: one animal with ↓thymus weight + ↓WBC + lymphoid depletion is flagged by the concordance engine. Group-mean incidence gates previously suppressed this signal.

Xbiom miss rate drops after roll-up fix

02

Compartment-Aware Checks

CSF markers evaluated only within Interface 23 (Checks 50–52).

Not cross-compared against systemic hematology or serum chemistry — the blood-brain barrier makes cross-compartment concordance biologically invalid.

Prevents false signals from comparing incomparable compartments.

03

Triangulation Across Data Types

Peripheral blood ↔ bone marrow cytology ↔ histopathology.

Pharmacological-toxicological axis: efficacy signals bunch at low doses, adverse signals at high doses.

Directly relevant for vaccine studies where dose-response runs opposite to traditional toxicology.

External validation:
Nova In Silico (in silico clinical modeling) confirmed this approach aligns with regulatory expectations for auditable computational evidence.

AI Can Mine Intelligence When Findings Data Is Structured

connect·share·advance

WHAT SDTM/SEND CAPTURES

- **Observation records**
  Body weight values, lab results, organ weights
- **Group-level summaries**
  Mean, SD, incidence counts by dose group
- **Terminology coding**
  MedDRA / SNOMED coded findings

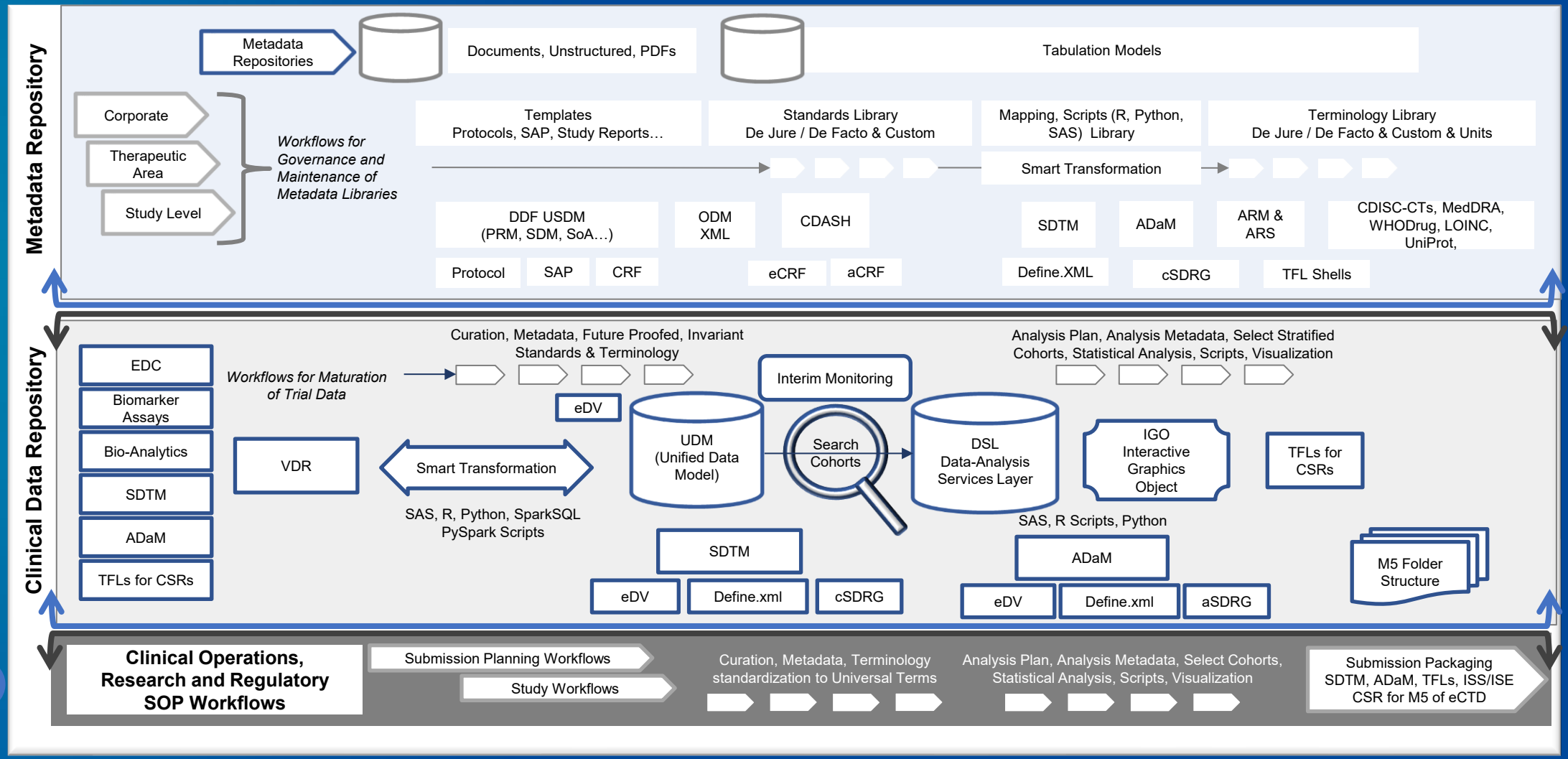
WHAT IS NOT CAPTURED (THE GAP)

- **Pathologist judgment**
  Treatment-related? Adverse? Biologically significant?
- **Concordance assessment**
  Do in-life, pathology, and clinical signals align?
- **Dose-response patterns**
  Is the finding dose-dependent across studies?

The opportunity:

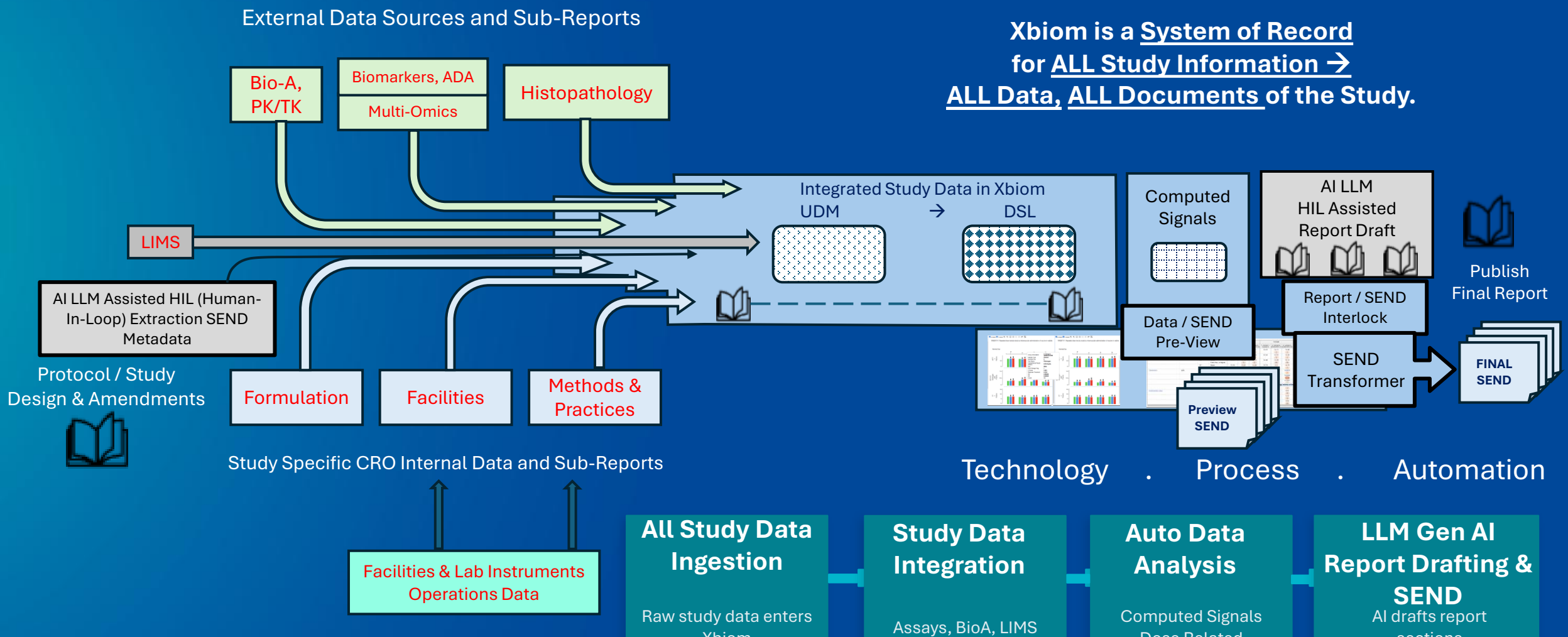
- When TR flags, concordance scores, and severity grades are stored as structured metadata， not buried in narratives
- AI can connect signals across hundreds of studies. That is where cross-study pattern recognition and predictive safety truly begin.

Xbiom – Platform & Repository Functional Components for Clinical Trial Data



AI-Driven Data Maturation, Analysis & Signal Generation

connect·share·advance

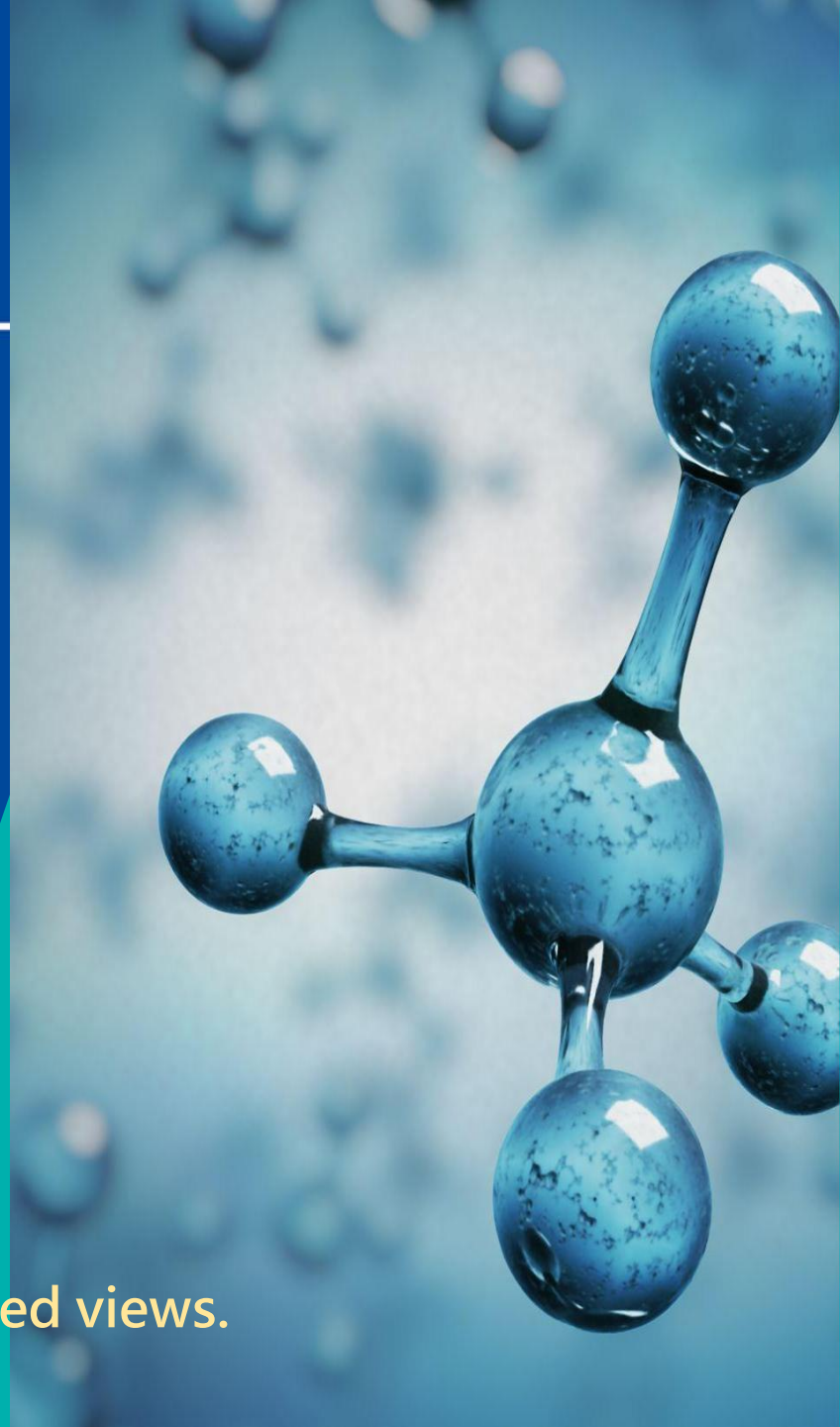


Questions to ask

SDTM or SEND based repository

1	What happens to loaded studies when IG/terminology updates?	Versions
2	Where to store data without standard domain?	Coverage
3	Can we recover original lab report after coding mapping?	Fidelity
4	Can we align multiple studies to one timeline?	Time
5	Can submission datasets be regenerated from the repository?	Direction

The repository holds source data, exchange standards are derived views.



The repository holds the data, and exchange standards are views of it

Source and assay data

Any vendor or lab format, batch or streamed

→

Universal data model

Harmonized, indexed, longitudinal and and extensible, with metadata that records every mapping and terminology decision

→

- **SDTM-ADaM and SEND for regulators**
- **Longitudinal and cohort analysis**
- **AI and signal detection**

Deterministic
the same data always give the same output, with every decision on record.

Universal data model concept: PointCross, PHUSE CSS poster jointly published with Roche.



Closing statements

- SDTM / SEND：单项研究监管审阅标准，是申报必备交付物。
- **以交**换为中心模式的局限： **静**态快照，不利于可追溯、可复现的跨项目 、**跨**临床-**非**临床AI
- UDM 价值：
  - **新增研究、数据**类型，无需重新映射历史数据
  - 统一受试者/动物时间线，打通临床与**非**临床检测数据，支持批量 + **流式数据接入**
  - **映射与**术语规则人工审批留痕，保证输入一致则结果确定性

核心结论：

SDTM-ADaM/SEND **是和**监管沟通的载体
不宜将其作为跨研究 **跨**阶段AI **挖掘的底**层持久数据底座；**而**UDM**可以**。

UDM universal data model
indexed, integrated, longitudinal patient data
in streamed data and batches

connect·share·advance

Questions and discussion welcome

Ping ping@pointcross.com

phuse

The Global Healthcare Data Science Community

Contact Channels
- office@phuse.global
- phuse.global

Social Media
- /company/phuse
- /phusetube
- /phuse_global
- /phusebook