TAMING REAL WORLD DATA (RWD) WITH CDISC STANDARDS: A PRACTICAL, STANDARDS-BASED APPROACH FOR REGULATORY-READY EVIDENCE

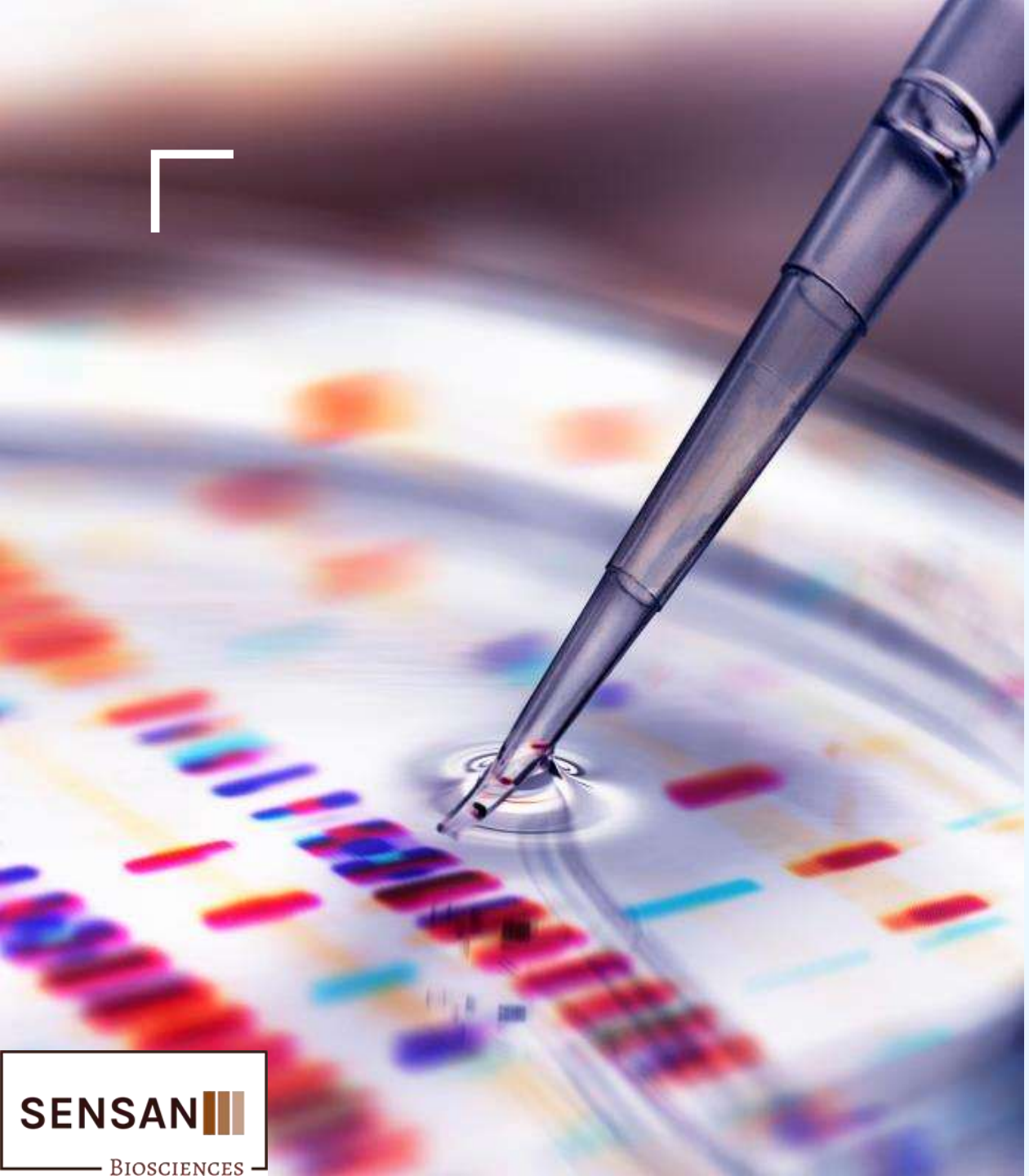
Presented by

Tamilselvi Narayanasamy

Sensan Biosciences

BENGALURU, INDIA
PharmaSUG SDE
APRIL 11, 2026

SENSAN BIOSCIENCES



AGENDA

- Define Real World Data (RWD)
- Sources of RWD
- Importance of RWD
- Key challenges of RWD
- Applying CDISC to Real World Data
- Challenges in Implementing CDISC for Real World Data
- Sentinel and Open EHR- Managing and utilizing Electronic Health Record (EHR) data
- Key Steps in Mapping Real-World clinical data to CDISC models - Sentinel Initiative and open EHR-based RWD examples
- Conclusion

"**Real-world data** are data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources."

SOURCES OF REAL WORLD DATA

Electronic Health Records (EHRs)

Health Claims

Patient Registries

Mobile and Wearable devices

Billing Records

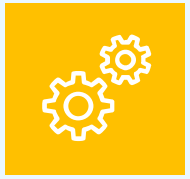
Patient Reported Outcomes

Why RWD is Important ?

Helps evaluate treatment effectiveness.

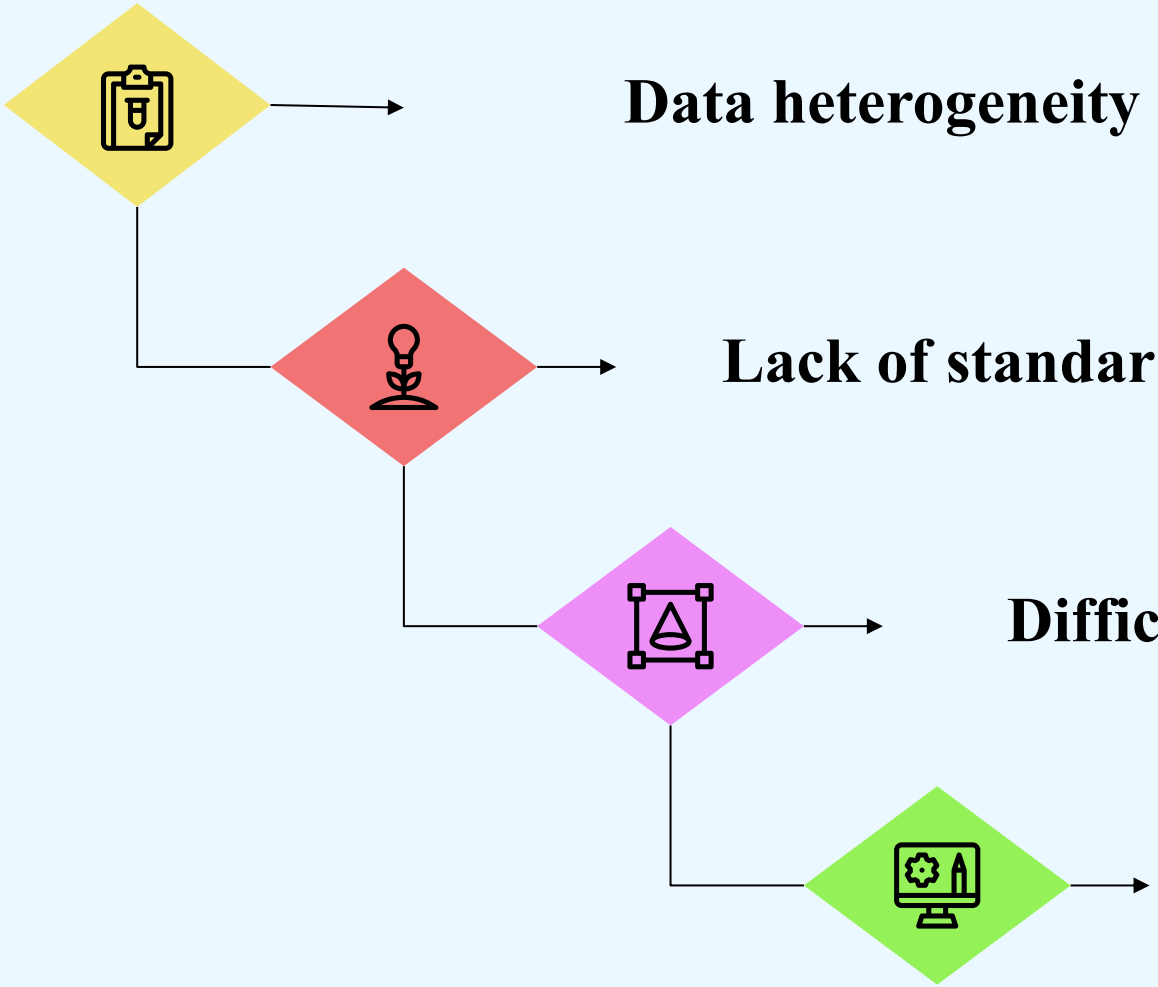
Provides insights from real clinical practice.

Supports regulatory decision making.



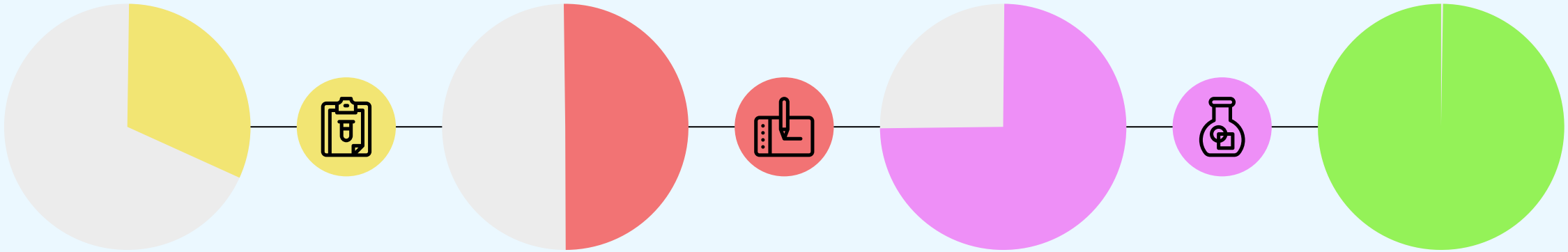
Real-world studies can be done more quickly and are more cost-effective.

Key Challenges of Real-World Data



- Data heterogeneity
- Lack of standard terminology
- Difficult integration across sources
- Missing or inconsistent values

Applying CDISC to Real World Data (Standards-Based Workflow)



Data Ingestion

Collect RWD from multiple sources (Hospitals, Claims databases, Registries)

Data Standardization

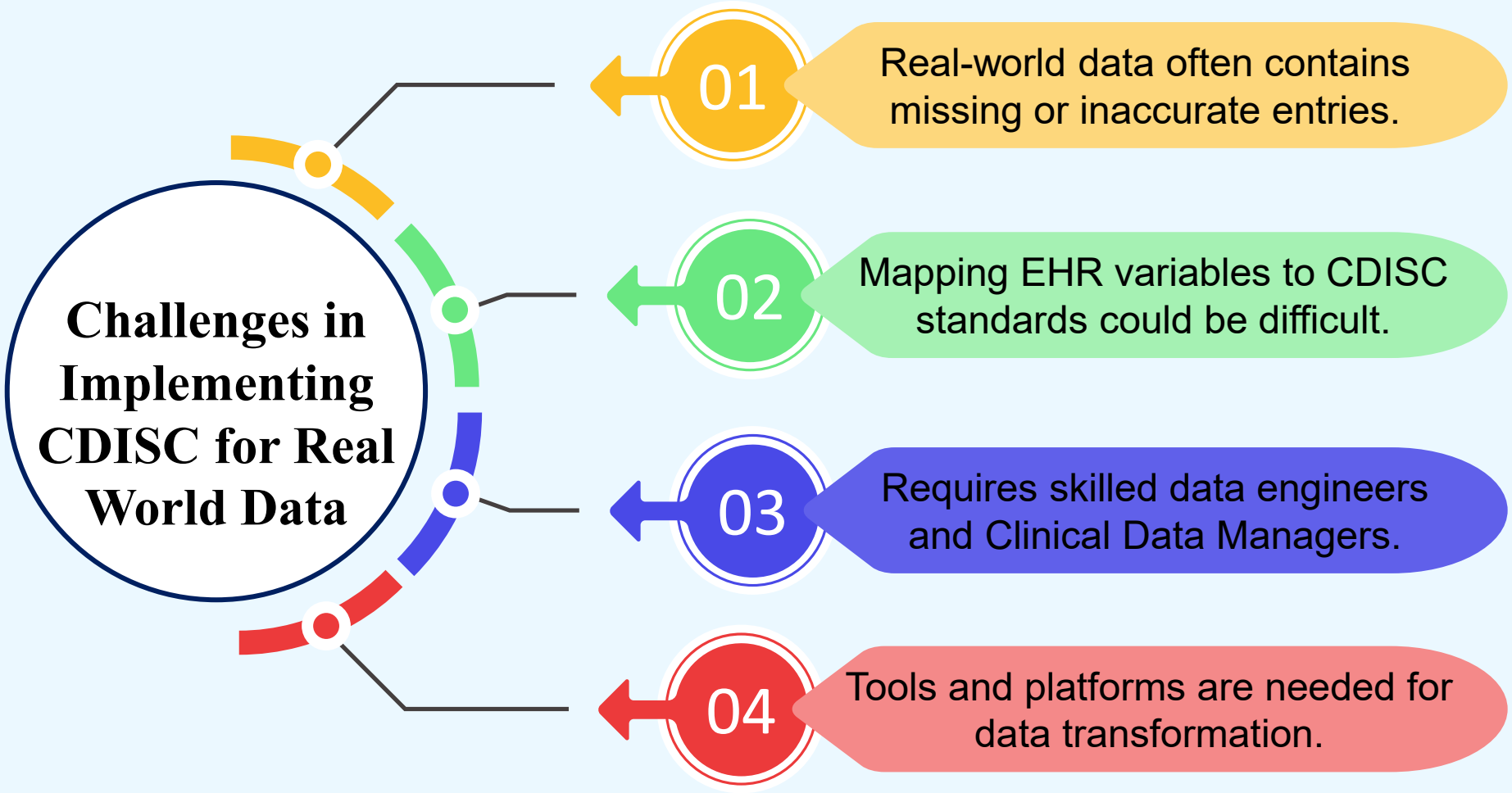
Apply CDISC terminology and variable mapping

Data Transformation

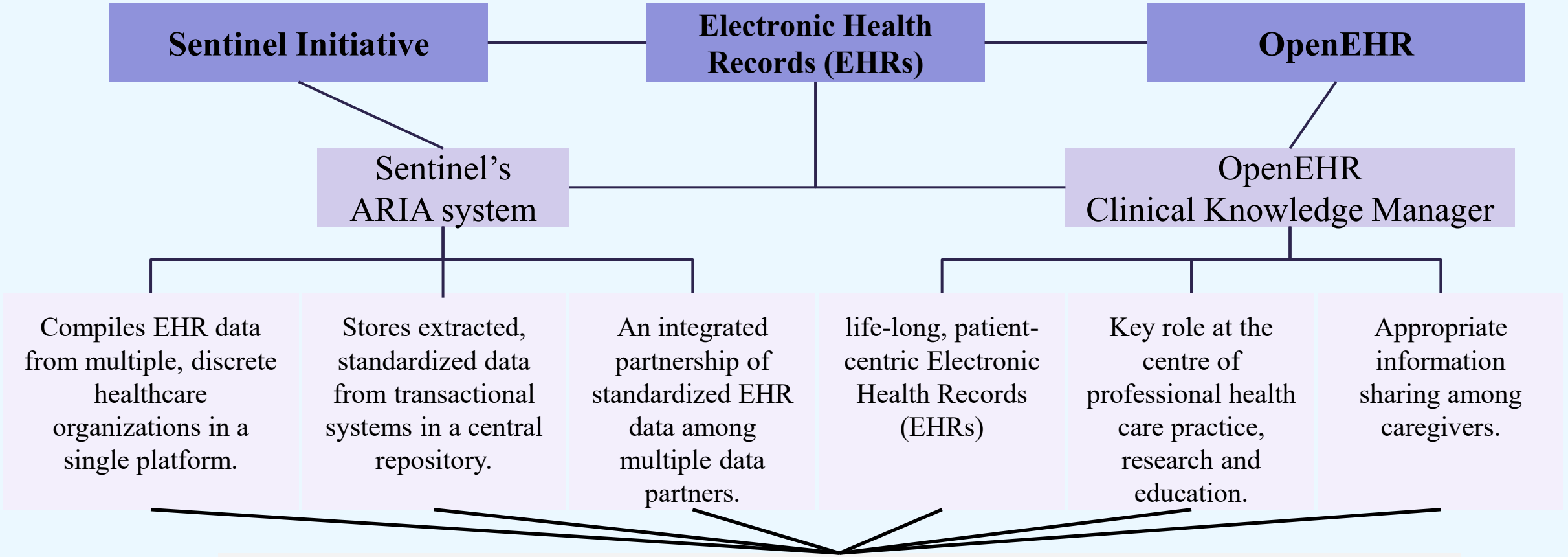
Convert into SDTM datasets(Identify class and domain, Determine list of variables to be included in the domain)

Analysis

Use ADaM datasets for statistical analysis


Challenges in Implementing CDISC for Real World Data
01
Real-world data often contains missing or inaccurate entries.
02
Mapping EHR variables to CDISC standards could be difficult.
03
Requires skilled data engineers and Clinical Data Managers.
04
Tools and platforms are needed for data transformation.

Sentinel and Open EHR - Managing and utilizing Electronic Health Record (EHR) data



openEHR focusing on standardized data structure and storage and **Sentinel focusing on distributed analysis for safety surveillance**.

Key Steps in Mapping Real-World clinical data to CDISC models - Sentinel Initiative and open EHR-based RWD examples

Data Acquisition and Understanding

Data Harmonization and Standard Terminologies

Validation and Quality Checks

Mapping to CDISC Models

Reproducible Data Transformation

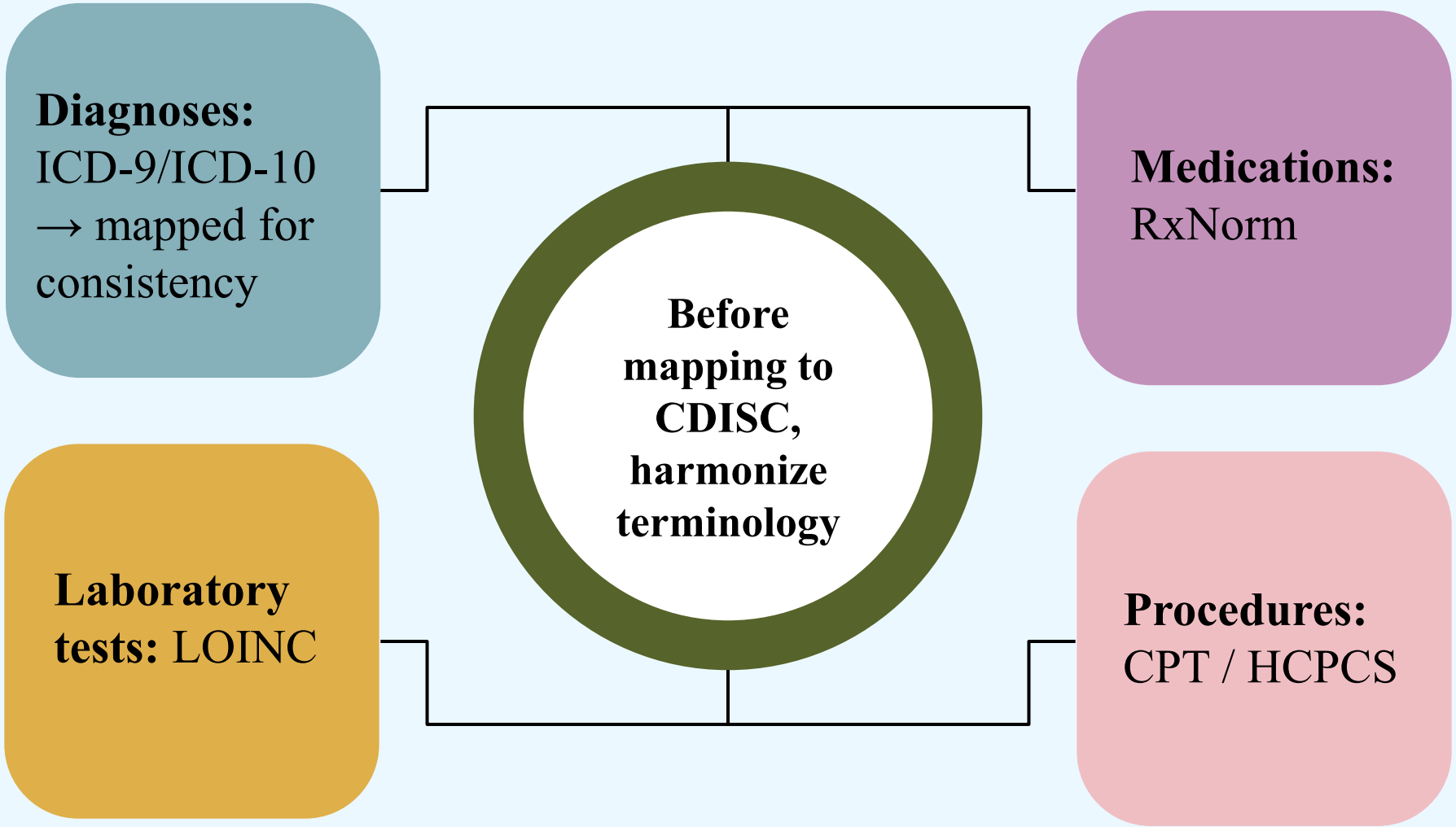
Maintaining Traceability

SENSAN BIOSCIENCES

Data Acquisition and Understanding

1. Obtain publicly available datasets (e.g., Sentinel distributed database (ARIA System), open EHR datasets (Example: Lab Datasets))
2. Review Data structure, Coding systems (ICD, CPT, LOINC, RxNorm, etc.), Metadata and data dictionaries.
3. Identify **source domains** such as Demographics, Diagnoses, Procedures, Medications, and Lab results.

Data Harmonization and Standard Terminologies



Mapping to CDISC Models

- Restructuring data into **domain-based datasets**
- Assigning **controlled terminology**
- Creating **unique subject identifiers**

Example Mapping:

Source (EHR/Sentinel) EHR Field	Example Value	CDISC SDTM Domain	CDISC Variable
Patient ID	12345	DM	USUBJID

EHR: Patient_ID = 12345

→ SDTM: USUBJID = "STUDY01-12345"

Example Mapping:

Source (EHR/Sentinel) EHR Field	Example Value	CDISC SDTM Domain	CDISC Variable
Test Name	Haemoglobin	LB	LBTEST
Result Value	13.5	LB	LBORRES
Unit	g/dL	LB	LBORRESU

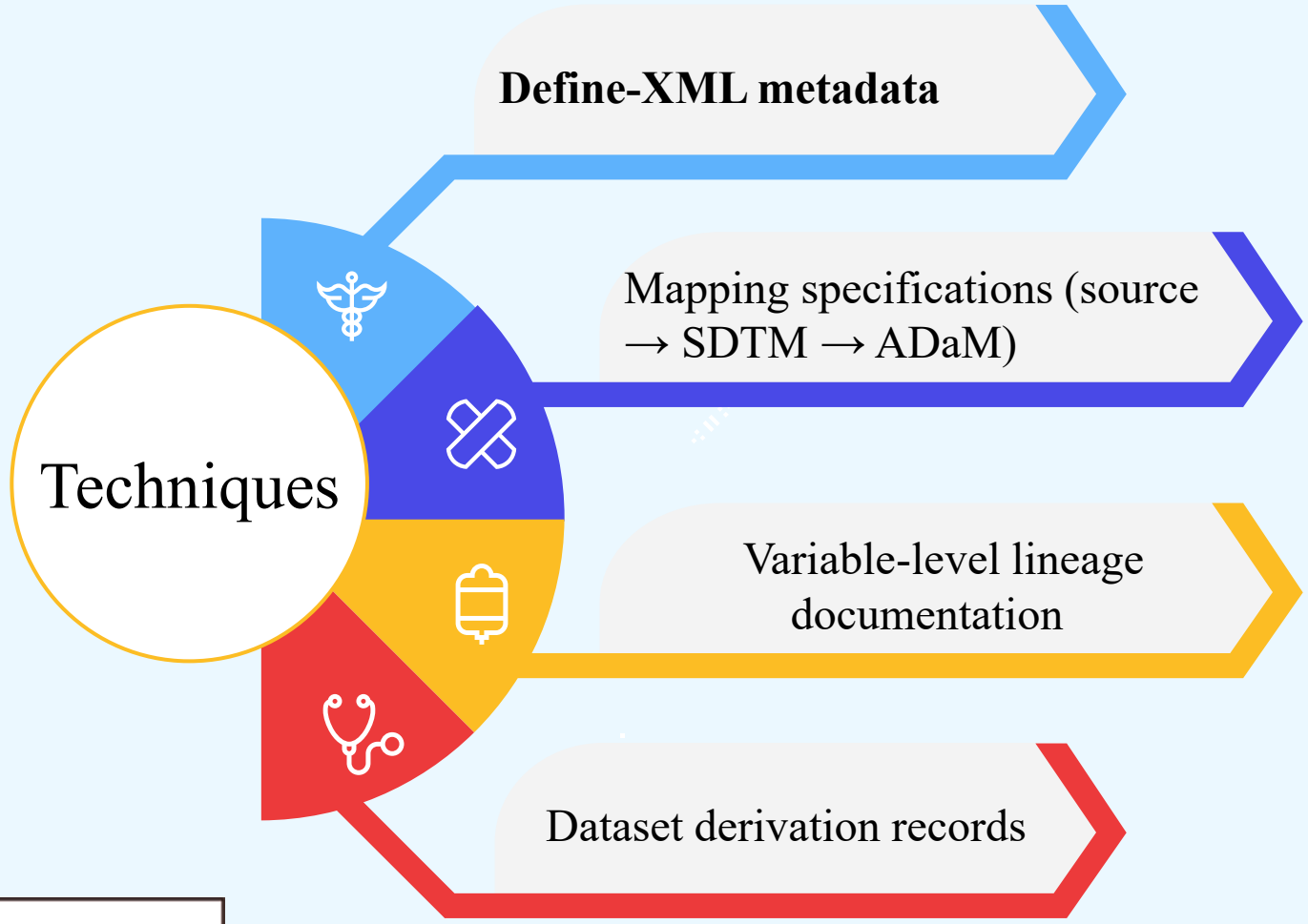
EHR: Hb = 13.5 g/dL
→ SDTM: LBTEST = "HEMOGLOBIN"
LBORRES = "13.5"
LBORRESU = "g/dL"

Source (EHR/Sentinel) EHR Field	Example Value	CDISC SDTM Domain	CDISC Variable
Diagnosis	Type 2 Diabetes	MH	MHTERM

EHR: Diagnosis = Type 2 Diabetes
→ SDTM: MHTERM = "TYPE 2 DIABETES MELLITUS"

Maintaining Traceability

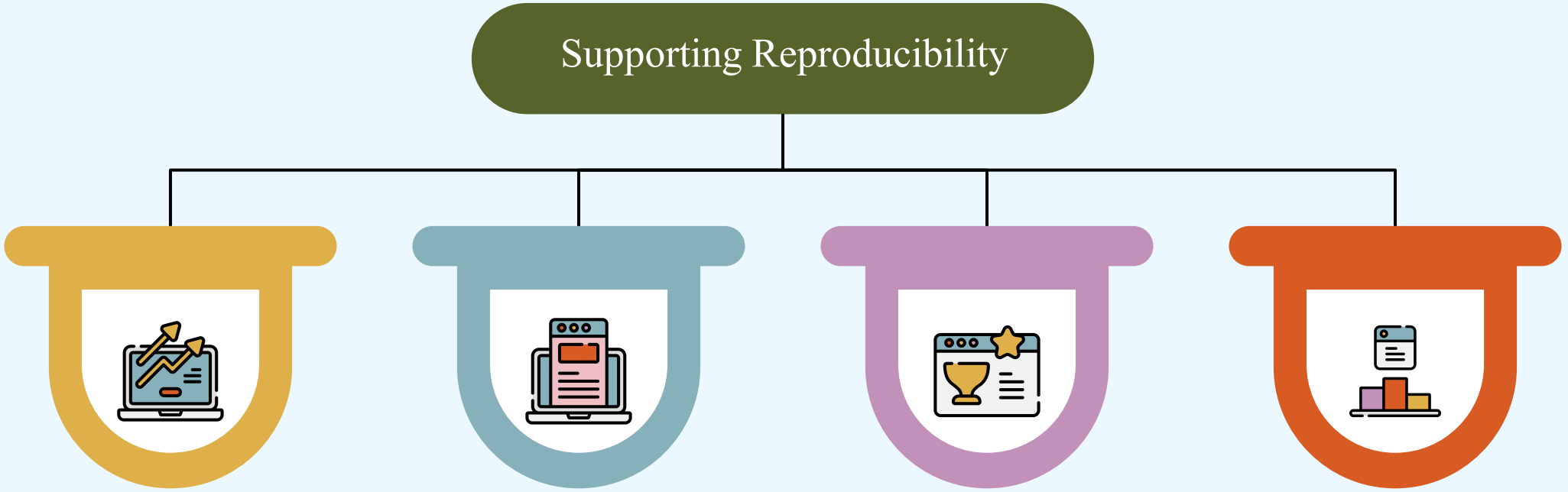
Traceability ensures every analysis result can be traced back to the original source data.



Example traceability chain:

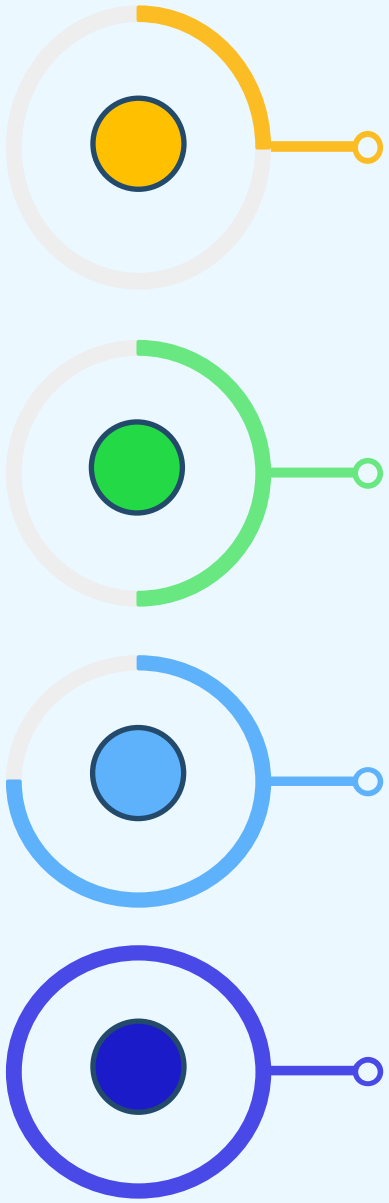
EHR encounter table
→ SDTM SV (Subject Visits)
→ ADaM ADSL (Subject-Level Analysis Dataset)
→ Statistical output (e.g., survival analysis)

Reproducible Data Transformation



Validation and Quality Checks

Quality control ensures compliance with CDISC and regulatory expectations.



- Domain structure validation
- Controlled terminology checks
- Cross-domain consistency
- Conformance to SDTM/ADaM rules

CONCLUSION

RWD provide a valuable and rich data source beyond the confines of traditional epidemiological studies, clinical trials, and lab-based experiments, with lower cost in data collection compared to the latter. Therefore, **combining RWD with CDISC standards is essential for modern clinical research and regulatory decision-making.**

BENGALURU, INDIA
PharmaSUG SDE
APRIL 11, 2026

ANY QUESTIONS?

SENSAN
BIOSCIENCES

Contact Information

Name: Tamilselvi Narayanasamy

Organization: SENSAN Biosciences Pvt. Ltd.

Address: Temple Steps, No.184-187, 3rd Floor, A-Block, Little Mount, Anna Salai.

City, State ZIP: Chennai - 600 015, Tamil Nadu

Work Phone: 7358316928

E-mail: tamilselvi.narayanasamy@sensanbio.com

Web: www.sensanbio.com