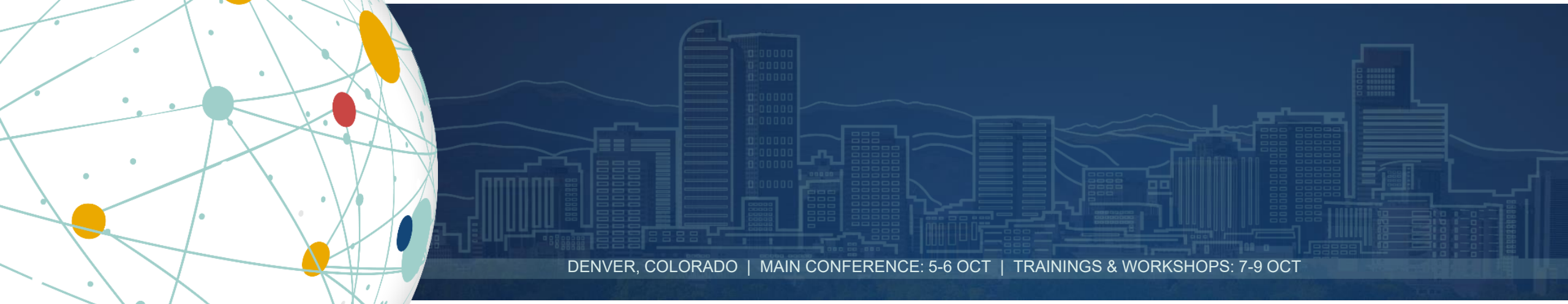


2026 US Interchange

THE FUTURE IS CONNECTED: STANDARDS AND AI POWERING DIGITAL TRANSFORMATION



RWD Lineage: A CDISC Standard for Reliable Real-World Data (RWD)

Anita Umesh, Data Standards Consultant
Tasha Nagamine, CTO, Droice Labs

Speakers



Tasha Nagamine
CTO
Droice Labs



Anita Umesh
Data Standards Consultant

Agenda

1. RWD Lineage: Background & Motivation
2. RWD Lineage in Define-XML
3. Implementation Examples
4. Project Status

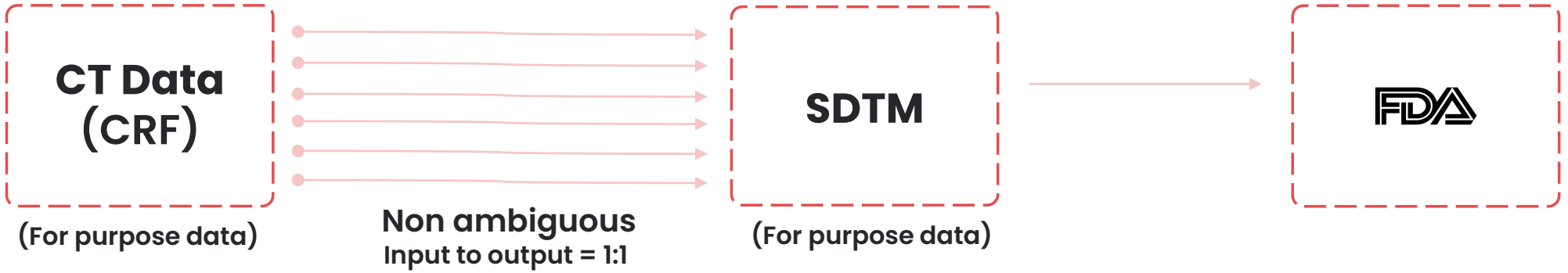
RWD Lineage: Background & Motivation

Introducing RWD Lineage

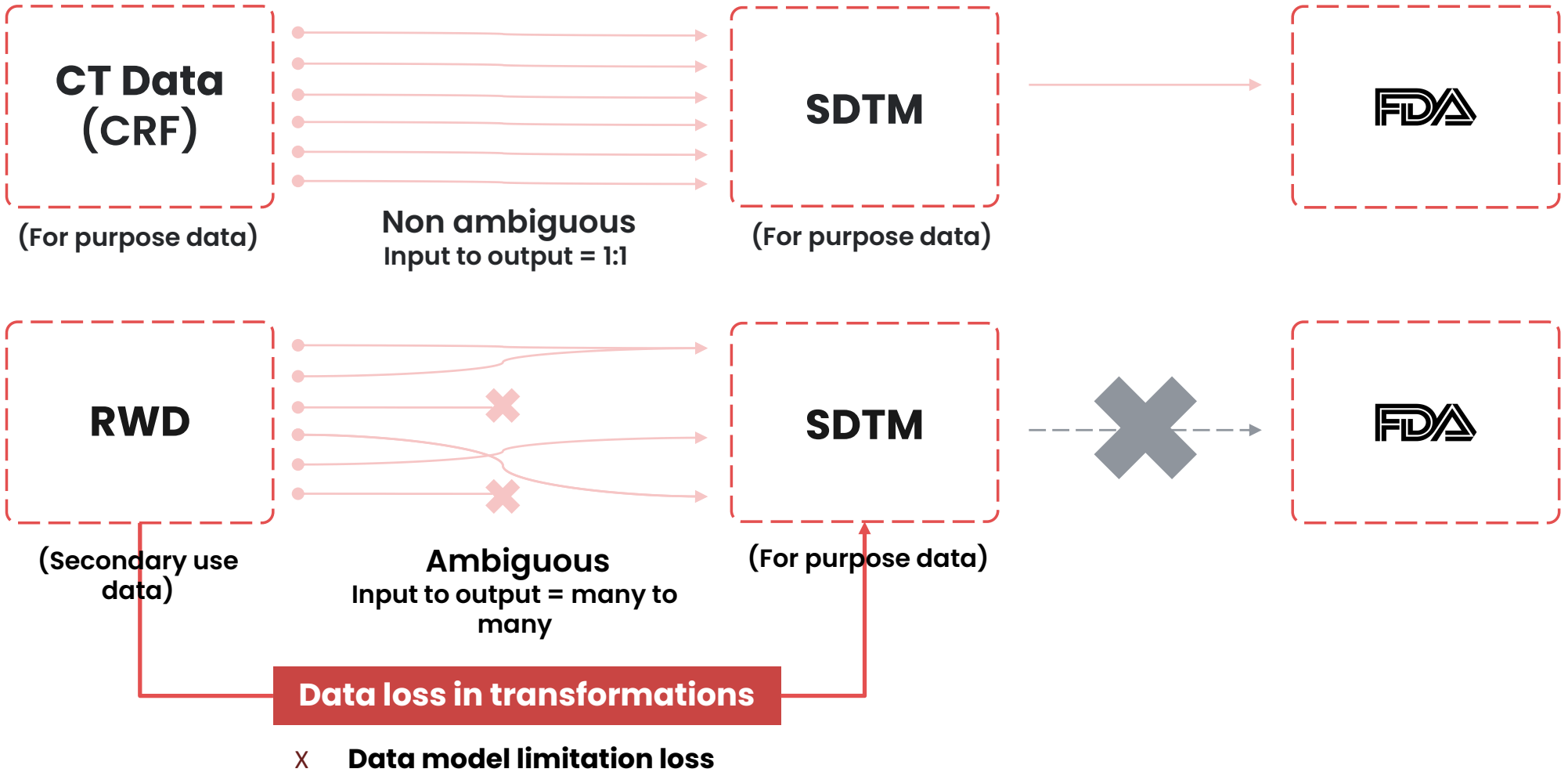
Project Goal

Create a CDISC data exchange standard for lineage metadata that is supplied along with RWD-derived SDTM, which provides the data reliability required by FDA to use RWE as primary evidence.

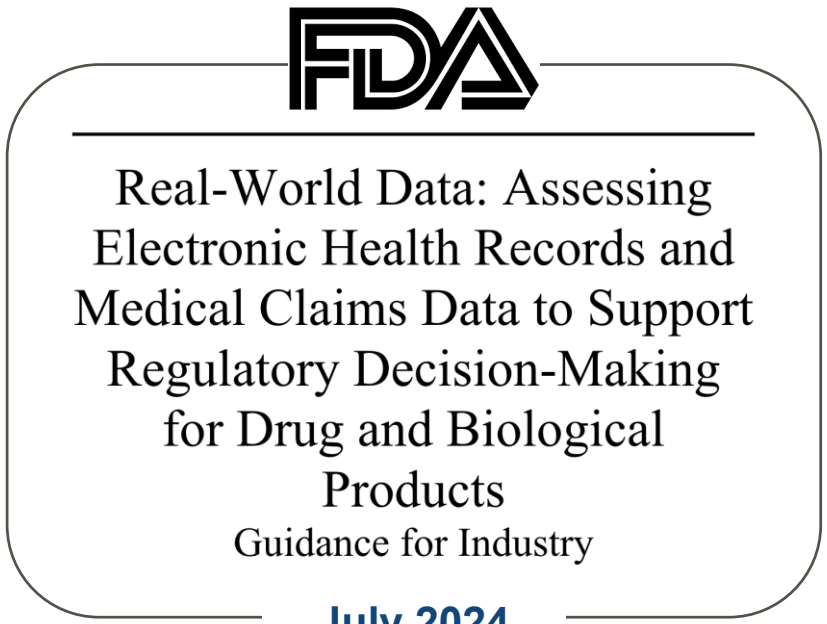
Reliability in RWD: Challenges



Reliability in RWD: Challenges



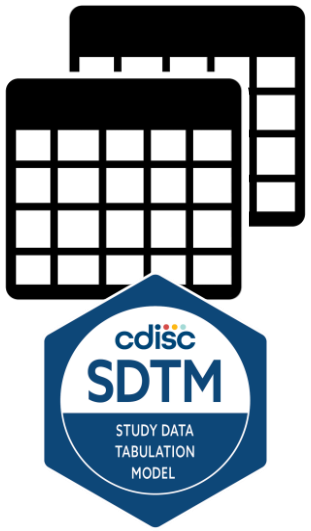
FDA: No Lowered Quality Standards for RWD/RWE



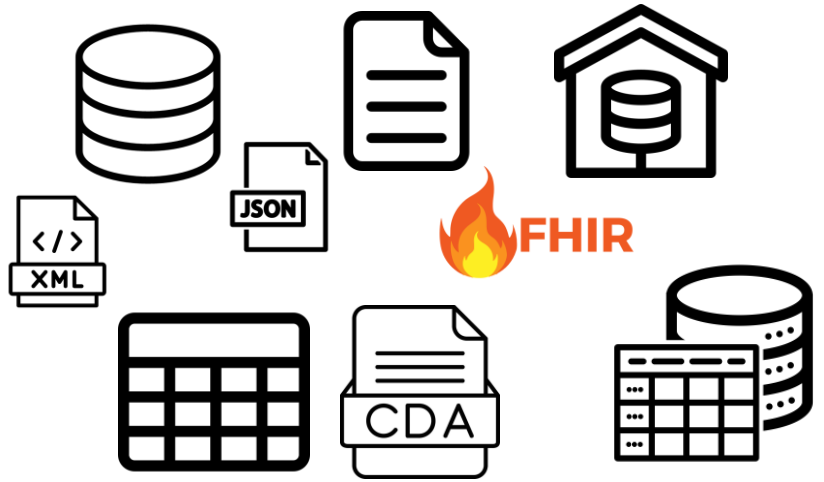
"For all study designs, it is important to ensure the **reliability**… of the data used to help support a regulatory decision. For the purposes of this guidance, the term **reliability includes accuracy, completeness, and traceability**."

RWD: Challenges for FDA Review

When reviewing and inspecting RWD submission, FDA needs to *interrogate source data* to determine *reliability* of study data



Validation & verification



Disparate RWD data sources

Reliability: Atomic Lineage for Individual Data Points

SDTM

MH						
…	USUBJID	MHSEQ	MHTERM	MHPRESP	MHOCCUR	…
…	001	1	HISTORY OF MYOCARDIAL INFARCTION	Y	Y	…
…	001	2	TYPE 2 DIABETES	Y	N	…
…	001	3	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	002	15	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	003	27	HYPERTENSION	Y	N	…
…	…	…	…	…	…	…

FDA

Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products

Guidance for Industry

July 2024

"...sponsors should evaluate the completeness, accuracy, and plausibility of the data, including **verifying data against its original source** (e.g., discharge notes, pathology reports, registry records)…"

"...subject matter experts' **review of medical records** (including structured and unstructured data) may be a preferred reference standard for validation of clinical events identified by diagnosis codes or automated algorithms…"

cdisc

https://www.fda.gov/regulatory-information/search-fda-guidance-documents/real-world-data-assessing-electronic-health-records-and-medical-claims-data-support-regulatory

Reliability: Atomic Lineage for Individual Data Points

SDTM

MH						
…	USUBJID	MHSEQ	MHTERM	MHPRESP	MHOCCUR	…
…	001	1	HISTORY OF MYOCARDIAL INFARCTION	Y	Y	…
…	001	2	TYPE 2 DIABETES	Y	N	…
…	001	3	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	002	15	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	003	27	HYPERTENSION	Y	N	…
…	…	…	…	…	…	…

Source RWD

PT_DX		
PT_ID	ICD10	TERM
21962	I10	Essential hypertension
21962	N18.3	Chronic kidney disease, stage 3
…	…	…

Reliability: Atomic Lineage for Individual Data Points

SDTM

MH						
…	USUBJID	MHSEQ	MHTERM	MHPRESP	MHOCCUR	…
…	001	1	HISTORY OF MYOCARDIAL INFARCTION	Y	Y	…
…	001	2	TYPE 2 DIABETES	Y	N	…
…	001	3	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	002	15	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	003	27	HYPERTENSION	Y	N	…
…	…	…	…	…	…	…

Source RWD

PT_DX		
PT_ID	ICD10	TERM
21962	I10	Essential hypertension
21962	N18.3	Chronic kidney disease, stage 3
…	…	…

VITALS		
Patno	Vital	Value
19251	BP	150/110
19251	BMI	28.3
…	…	…

NOTES	
PT_ID	TEXT
19251	Medical history: Type 2 DM on insulin, CHF, HTN, CKD3, …
19251	Discharge summary: …
…	…

Reliability: Atomic Lineage for Individual Data Points

SDTM

MH						
…	USUBJID	MHSEQ	MHTERM	MHPRESP	MHOCCUR	…
…	001	1	HISTORY OF MYOCARDIAL INFARCTION	Y	Y	…
…	001	2	TYPE 2 DIABETES	Y	N	…
…	001	3	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	002	15	HYPERTENSION	Y	Y	…
…	…	…	…	…	…	…
…	003	27	HYPERTENSION	Y	N	…
…	…	…	…	…	…	…

True positive

True positive

False negative

Interrogate source data to validate the accuracy of individual data points

Atomic lineage provides traceability of individual data points, including coordinates and values, from target SDTM to original source data

RWD Lineage in Define-XML

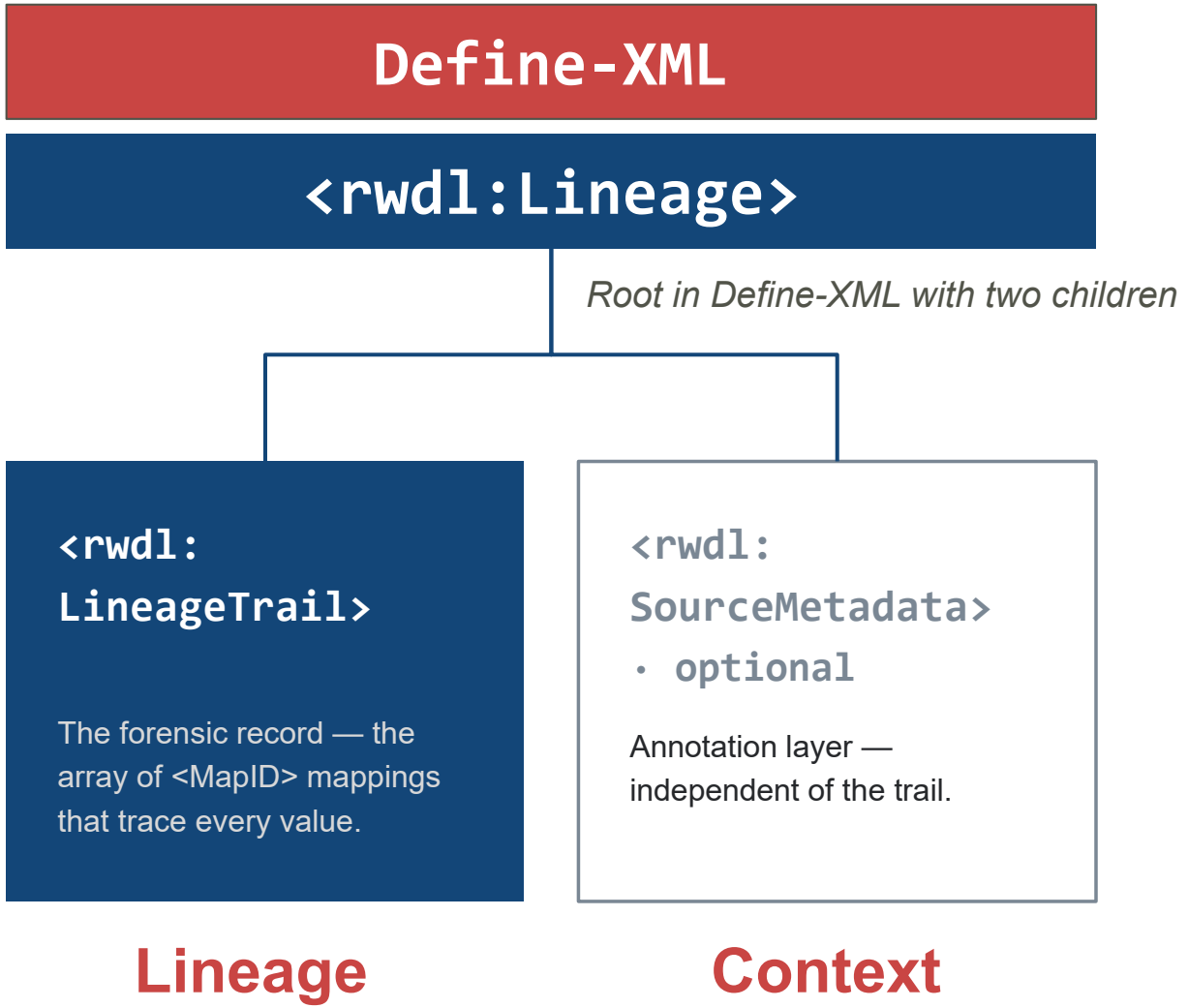
RWD Lineage Specification Overview

RWD Lineage

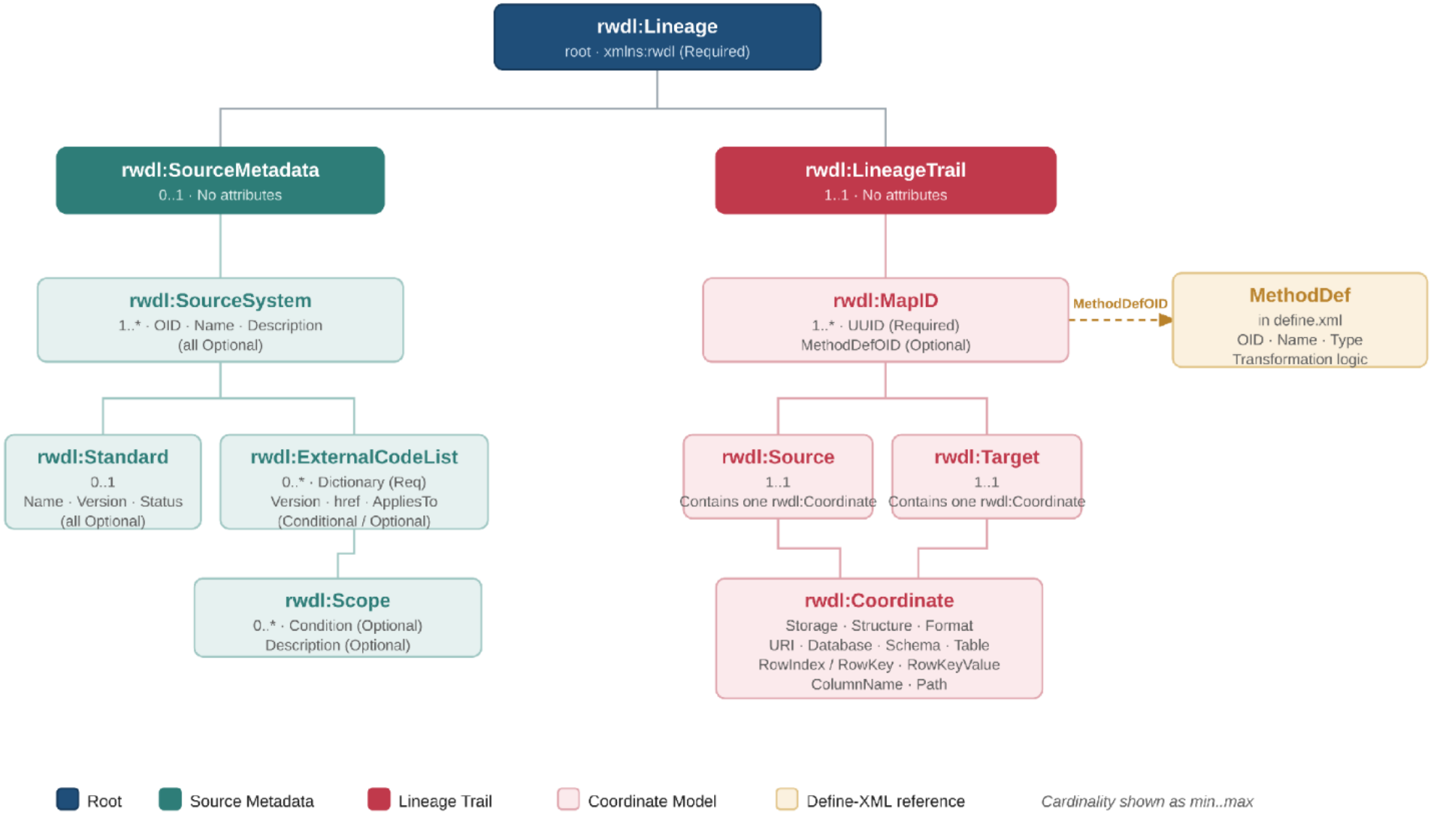
A machine-readable XML extension to Define-XML that captures **data point-level lineage** metadata — tracing every value in an RWD-derived SDTM dataset back to its exact source in the originating EHR, lab system, or clinical data store.

Why?

FDA and other regulators need data reliability evidence to accept RWE as primary evidence. RWD-Lineage provides that audit trail in a standardized, machine-readable format.



Inside the Lineage Trail



Inside the Lineage Trail

Implementation in Define-XML

Implemented as a namespace extension to Define-XML (rwdl:lineage), provided inside Define-XML or referenced as a standalone XML file alongside SDTM datasets.

- Source - source data point (e.g., in EHR/FHIR/…)
- Target - target data point (e.g., in SDTM)
- UUID - unique identifier for a source-target data point pair
- MethodDefOID - a reference to Define-XML MethodDef to describe the transformation between the source-target pair

Lineage Trail Unit: <MapID>

`UUID`	Deterministic UUIDv5 from Source + Target hash
`MethodDefOID`	OID ref to a Define-XML MethodDef (transform logic)
`rwdl:Source`	Coordinate of the origin RWD data point
`rwdl:Target`	Coordinate of the destination SDTM cell

`Coordinate = Storage x Structure`

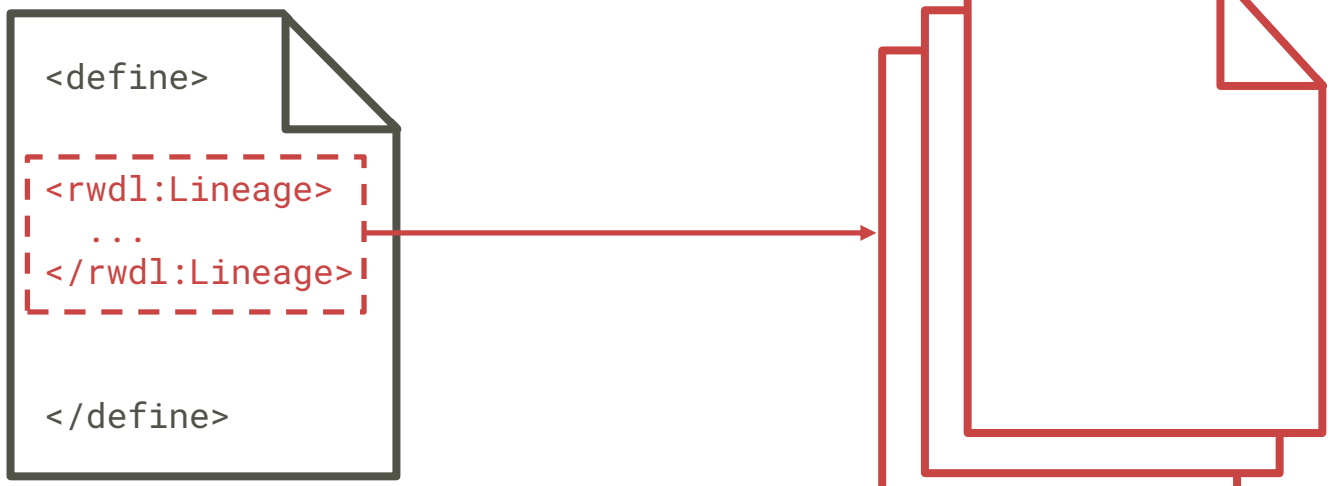
Database | Filesystem | API | Message x Tabular | Path | Object

Draft specification in GitHub: https://github.com/cdisc-org/DataExchange-RWD-Lineage/blob/main/documents/RWD-Lineage_Data_Standard_Specification.md

RWD Lineage in Define-XML

Self-contained <rwdl:lineage> XML tag referencing RWD Lineage .XML files

Detailed data point level lineage metadata stored in one or more XML files referenced in Define-XML



Define-XML

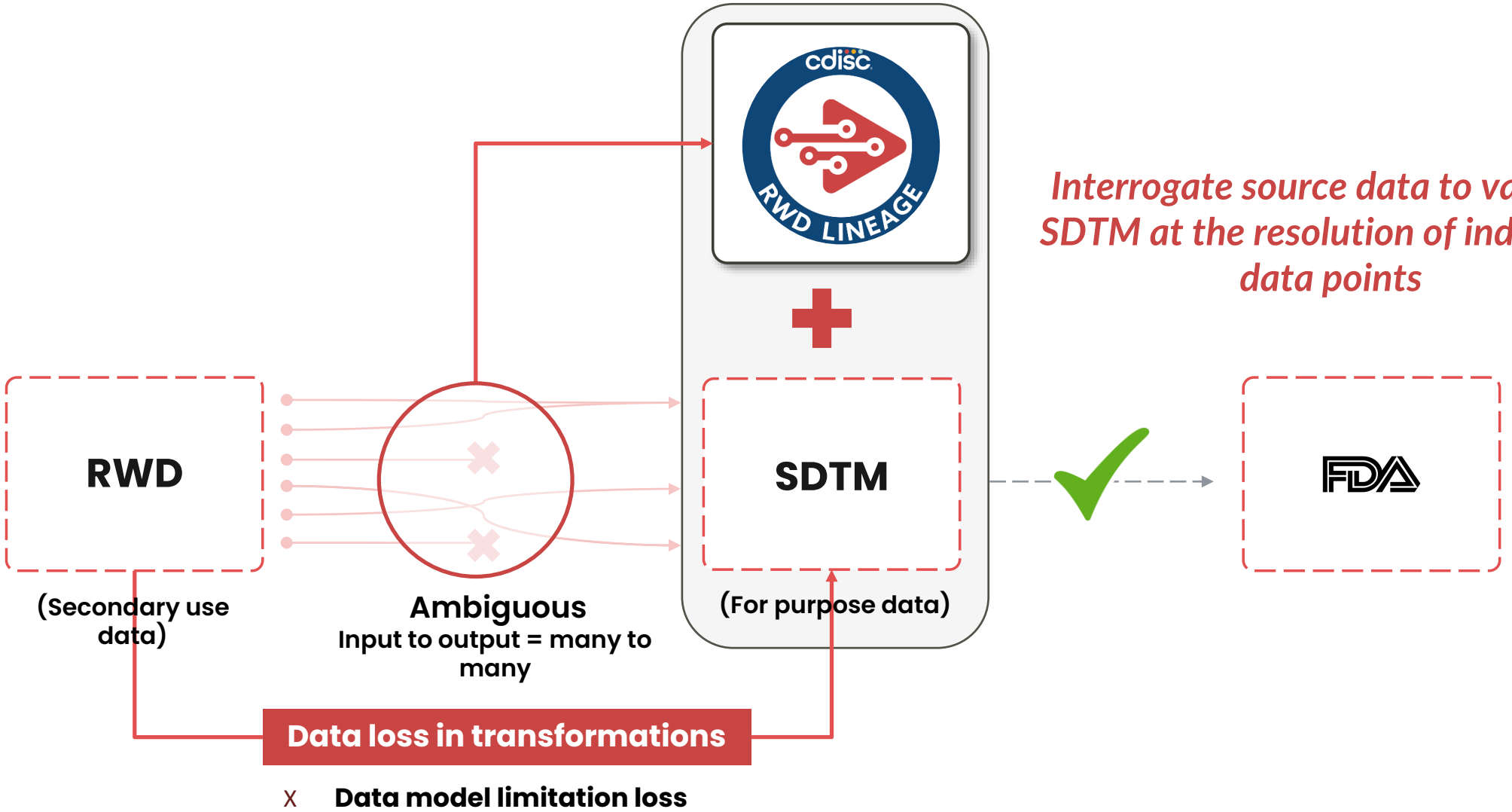
RWD Lineage

RWD Lineage in Define-XML

- Define-XML: already in the data standards catalog
- Minimal impact to existing workflows
  - Self-contained lineage metadata references in Define-XML
  - Lineage metadata can be stored in standalone files
- Easy and pain-free adoption



RWD LINEAGE for Reliable RWE



Interrogate source data to validate SDTM at the resolution of individual data points

Implementation Examples

Key requirements for examples to show

- Data-point level lineage
- Can “stack” multiple stages of transformations
  - Example: FHIR > CDM > SDTM
  - Example: Multiple aggregated EHR sources > SDTM
- Supports commonly used formats of RWD
  - Tabular data: e.g., EHR, claims, aggregated EHR in common data models
  - Tree-based formats: e.g., FHIR (JSON), CCDA (XML)
  - Objects: e.g., documents, PDF charts, images/signals, all at the file-level

Takeaway: Examples to show RWD Lineage standard will cover all the general use cases of RWD in regulatory decision-making

Considerations and Assumptions

- Considerations
  - Integrate into current submission workflow with minimal disruption to current process
    - Use of Define.xml as a backbone with lineage xml section
  - Data privacy issues
    - Lineage xml points to coordinates, not the data values
    - Data values to be accessible via coordinates with appropriate access controls
  - Data transformations
    - Expressed within Define.xml using existing MethodDef
    - Not standardized
  - Multiple data sources
    - Expressed within source metadata section of the lineage xml
  - Lineage xml size
    - Divide Lineage xmls into smaller files
  - Introduction of new metadata standards (e.g., Dataset-JSON)
    - Lineage xml concepts are transferable
  - Study cohort creation from screening cohort
    - Described in protocol

- Assumptions
  - Source data is immutable: the data is a “snapshot”, much like a database lock, with an associated lineage that is only valid for that snapshot

Worked Examples

Data-point level lineage across different clinical data transformation scenarios

Example 1

Clinical Events (CE)

EHR diagnoses, vitals & clinical notes → SDTM CE domain

Traceability scenario:

Multi-source evidence convergence
A single SDTM cell (CEOCCUR) traces back to source data points with ICD-10 diagnosis codes, blood pressure measurements, and NLP-extracted findings from clinical notes — all contributing to one clinical event determination.

3 source tables • 2 subjects • CE domain

Example 2

Labs & Adverse Events (LB, AE)

EHR LOINC lab results → SDTM LB + AE domains

Traceability scenario:

Cross-domain derivation chain
Raw lab result strings are parsed, converted to standard units, and evaluated against reference ranges in LB. Elevated values then derive an adverse event record in AE — full traceability from data points in source EHR through both domains.

1 source table • 2 subjects • LB + AE domains

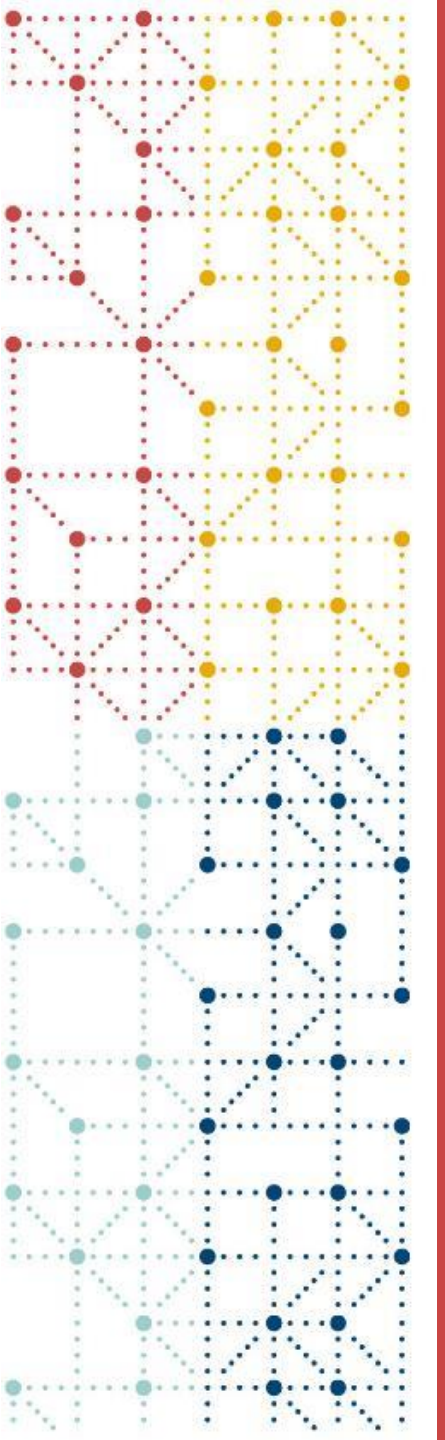
Example 3

Breast Cancer (MH)

Two RWD Sources → SDTM MH domain

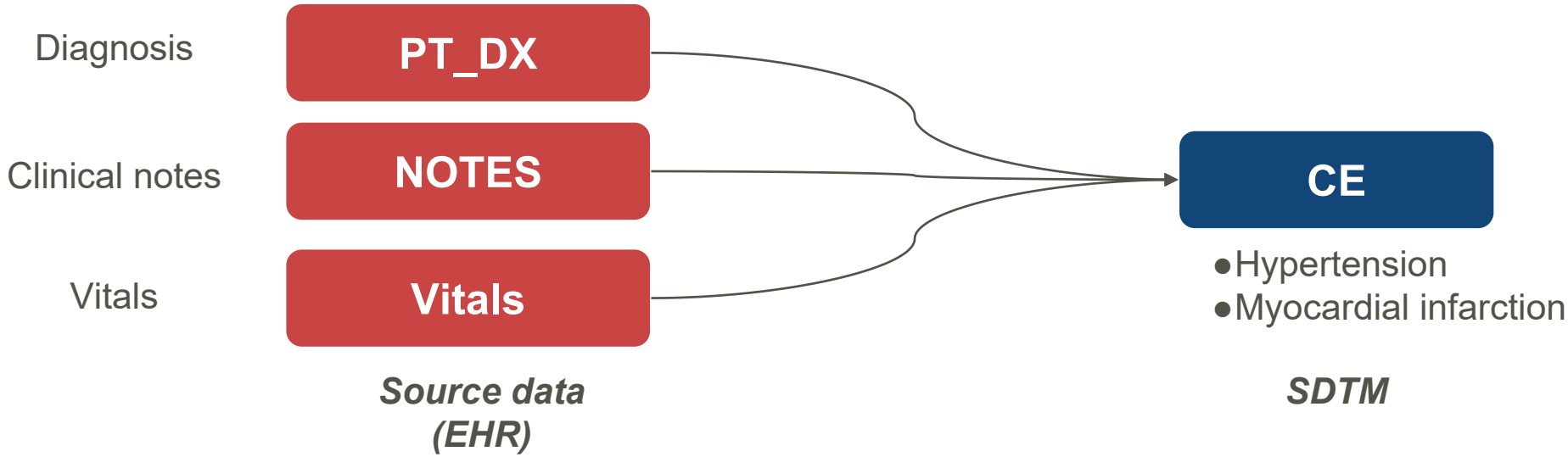
Traceability scenario:

Multisource RWD
EHR (MIMIC) and research database (the Cancer Genome Atlas (TCGA)) are combined to create SDTM MH domain. Traceability is preserved for multiple sources in a real research database setting.



Example 1: Hypertension and Myocardial Infarction Clinical Events (CE)

Example 1: SDTM CE Derivation



Two prespecified clinical events:

- Hypertension onset within the followup period
  - Patient is newly assigned ICDI0 diagnosis code I10–I16 (Hypertensive diseases)
  - Patient has at least two elevated blood pressure measurements (>=130/80 mmHg) over the course of a 3-month period within the followup period
  - Patient has newly diagnosed hypertension in clinical notes
- Myocardial infarction occurs within the followup period
  - Patient is assigned ICD10 diagnosis code I21 (Acute myocardial infarction) or I122 (Subsequent ST elevation (STEMI) and non-ST elevation (NSTEMI) myocardial infarction)
  - Patient has records of acute myocardial infarction in notes within the followup period

Example 1: Subject 001 Hypertension

SDTM

CE				
USUBJID	**CESEQ**	**CETERM**	**CEPRESP**	**CEOCCUR**
001	1	MYOCARDIAL INFARCTION	Y	Y
001	2	HYPERTENSION	Y	Y
002	1	MYOCARDIAL INFARCTION	Y	N
002	2	HYPERTENSION	Y	Y

Patient is newly assigned ICDI0 diagnosis code I10–I16 (Hypertensive diseases) within the followup period

Source RWD

PT_DX			
PT_ID	**DATE**	**ICD10**	**TERM**
21962	05/30/2025	I21.3	ST elevation (STEMI) myocardial infarction of unspecified site
21962	06/27/2025	N18.3	Chronic kidney disease, stage 3
21962	06/27/2025	I25.2	Old myocardial infarction
21962	06/27/2025	I10	Essential hypertension
19251	04/16/2024	E11.9	Type 2 diabetes mellitus without complications
19251	05/28/2024	E11.65	Type 2 diabetes mellitus with hyperglycemia
19251	07/25/2024	E11.65	Type 2 diabetes mellitus with hyperglycemia
19251	07/25/2024	K80.11	Calculus of gallbladder with chronic cholecystitis with obstruction

Inside the Lineage Trail

Implementation in Define-XML

Implemented as a namespace extension to Define-XML (rwdl:lineage), provided inside Define-XML or referenced as a standalone XML file alongside SDTM datasets.

- Source - source data point (e.g., in EHR/FHIR/…)
- Target - target data point (e.g., in SDTM)
- UUID - unique identifier for a source-target data point pair
- MethodDefOID - a reference to Define-XML MethodDef to describe the transformation between the source-target pair

Lineage Trail Unit: <MapID>

`UUID`	Deterministic UUIDv5 from Source + Target hash
`MethodDefOID`	OID ref to a Define-XML MethodDef (transform logic)
`rwdl:Source`	Coordinate of the origin RWD data point
`rwdl:Target`	Coordinate of the destination SDTM cell

`Coordinate = Storage x Structure`

Database | Filesystem | API | Message x Tabular | Path | Object

Draft specification in GitHub: https://github.com/cdisc-org/DataExchange-RWD-Lineage/blob/main/documents/RWD-Lineage_Data_Standard_Specification.md

Example 1: Subject 001 Hypertension

SDTM

CE				
USUBJID	**CESEQ**	**CETERM**	**CEPRESP**	**CEOCCUR**
001	1	MYOCARDIAL INFARCTION	Y	Y
001	2	HYPERTENSION	Y	Y
002	1	MYOCARDIAL INFARCTION	Y	N
002	2	HYPERTENSION	Y	Y

Patient is newly assigned ICDI0 diagnosis code I10–I16 (Hypertensive diseases) within the followup period

Source RWD

PT_DX			
PT_ID	**DATE**	**ICD10**	**TERM**
21962	05/30/2025	I21.3	ST elevation (STEMI) myocardial infarction of unspecified site
21962	06/27/2025	N18.3	Chronic kidney disease, stage 3
21962	06/27/2025	I25.2	Old myocardial infarction
21962	06/27/2025	I10	Essential hypertension
19251	04/16/2024	E11.9	Type 2 diabetes mellitus without complications
19251	05/28/2024	E11.65	Type 2 diabetes mellitus with hyperglycemia
19251	07/25/2024	E11.65	Type 2 diabetes mellitus with hyperglycemia
19251	07/25/2024	K80.11	Calculus of gallbladder with chronic cholecystitis with obstruction

Example 1: Subject 001 Hypertension

```
<?xml version="1.0" encoding="UTF-8"?>
<rwdl:Lineage xmlns:rwdl="http://www.cdisc.org/ns/rwdl/v1.0">

    <!-- LAYER 1: Source metadata - assertions about the source systems -->
    <rwdl:SourceMetadata>
        <rwdl:SourceSystem OID="SRC.CSV.1"
                    Name="Example 1 Clinical Source CSV Files"
                    Description="CSV exports from clinical source system; includes patient diagnoses (pt_dx.csv), vitals (vitals.csv), and clinical notes
            <rwdl:ExternalCodeList Dictionary="ICD-10-CM" Version="2024"
                                   AppliesTo="pt_dx.csv ICD10"/>
        </rwdl:SourceSystem>
    </rwdl:SourceMetadata>

    <!-- LAYER 2: Lineage trail - Source -> Target pairs -->
    <rwdl:LineageTrail>

        <!-- pt_dx.csv row 4 -> ce.csv row 2 -->
<rwdl:MapID UUID="35060134-fc2f-4cdf-9abe-491924739bd5">
    <rwdl:Source>
        <rwdl:Coordinate Storage="FILESYSTEM" Structure="TABULAR" Format="CSV">
            <rwdl:URI>https://github.com/cdisc-org/DataExchange-RWD-Lineage/blob/main/examples/example1/data/source/pt_dx.csv</rwdl:URI>
            <rwdl:RowIndex>4</rwdl:RowIndex>
            <rwdl:ColumnName>ICD10</rwdl:ColumnName>
        </rwdl:Coordinate>
    </rwdl:Source>
    <rwdl:Target>
        <rwdl:Coordinate Storage="FILESYSTEM" Structure="TABULAR" Format="CSV">
            <rwdl:URI>https://github.com/cdisc-org/DataExchange-RWD-Lineage/blob/main/examples/example1/data/sdtm/ce.csv</rwdl:URI>
            <rwdl:RowIndex>2</rwdl:RowIndex>
            <rwdl:ColumnName>CEOCCUR</rwdl:ColumnName>
        </rwdl:Coordinate>
    </rwdl:Target>
</rwdl:MapID>
```

MapID

Source ← 4th row of column *ICD10* in table PT_DX

maps to

Target ← 2nd row of column CEOCCUR in table CE

https://github.com/cdisc-org/DataExchange-RWD-Lineage/blob/main/examples/example1/data/define/rwd-lineage.xml

Example 1: Subject 001 Hypertension

SDTM

CE				
USUBJID	**CESEQ**	**CETERM**	**CEPRESP**	**CEOCCUR**
001	1	MYOCARDIAL INFARCTION	Y	Y
001	2	HYPERTENSION	Y	Y
002	1	MYOCARDIAL INFARCTION	Y	N
002	2	HYPERTENSION	Y	Y

Patient is newly assigned ICDI0 diagnosis code I10–I16 (Hypertensive diseases) within the followup period

Source RWD

PT_DX			
PT_ID	**DATE**	**ICD10**	**TERM**
21962	05/30/2025	I21.3	ST elevation (STEMI) myocardial infarction of unspecified site
21962	06/27/2025	N18.3	Chronic kidney disease, stage 3
21962	06/27/2025	I25.2	Old myocardial infarction
21962	06/27/2025	I10	Essential hypertension
19251	04/16/2024	E11.9	Type 2 diabetes mellitus without complications
19251	05/28/2024	E11.65	Type 2 diabetes mellitus with hyperglycemia
19251	07/25/2024	E11.65	Type 2 diabetes mellitus with hyperglycemia
19251	07/25/2024	K80.11	Calculus of gallbladder with chronic cholecystitis with obstruction

Example 1: Subject 001 Hypertension

```
31              </rwdl:Coordinate>
32          </rwdl:Target>
33      </rwdl:MapID>
34
35      <rwdl:MapID UUID="77d5b278-692e-4d9b-8232-179b34f8e6b3">
36          <rwdl:Source>
37              <rwdl:Coordinate Storage="FILESYSTEM" Structure="TABULAR" Format="CSV">
38                  <rwdl:URI>https://github.com/cdisc-org/DataExchange-RWD-Lineage/blob/main/examples/example1/data/source/pt_dx.csv</rwdl:URI>
39                  <rwdl:RowIndex>4</rwdl:RowIndex>
40                  <rwdl:ColumnName>TERM</rwdl:ColumnName>
41              </rwdl:Coordinate>
42          </rwdl:Source>
43          <rwdl:Target>
44              <rwdl:Coordinate Storage="FILESYSTEM" Structure="TABULAR" Format="CSV">
45                  <rwdl:URI>https://github.com/cdisc-org/DataExchange-RWD-Lineage/blob/main/examples/example1/data/sdtm/ce.csv</rwdl:URI>
46                  <rwdl:RowIndex>2</rwdl:RowIndex>
47                  <rwdl:ColumnName>CEOCCUR</rwdl:ColumnName>
48              </rwdl:Coordinate>
49          </rwdl:Target>
50      </rwdl:MapID>
51
52      <rwdl:MapID UUID="7e376beb-7dad-4f5c-a212-88283ac22eba" MethodDefOID="MT.AFTERIDXDATE">
53          <rwdl:Source>
54              <rwdl:Coordinate Storage="FILESYSTEM" Structure="TABULAR" Format="CSV">
55                  <rwdl:URI>https://github.com/cdisc-org/DataExchange-RWD-Lineage/blob/main/examples/example1/data/source/pt_dx.csv</rwdl:URI>
```

Source

← 4th row of column *TERM* in table PT_DX

maps to

Target

← 2nd row of column CEOCCUR in table CE

https://github.com/cdisc-org/DataExchange-RWD-Lineage/blob/main/examples/example1/data/define/rwd-lineage.xml

Example 1: Subject 001 Hypertension

SDTM

CE				
USUBJID	**CESEQ**	**CETERM**	**CEPRESP**	**CEOCCUR**
001	1	MYOCARDIAL INFARCTION	Y	Y
001	2	HYPERTENSION	Y	Y
002	1	MYOCARDIAL INFARCTION	Y	N
002	2	HYPERTENSION	Y	Y

Patient is newly assigned ICDI0 diagnosis code I10–I16 (Hypertensive diseases) within the followup period

Source RWD

PT_DX			
PT_ID	**DATE**	**ICD10**	**TERM**
21962	05/30/2025	I21.3	ST elevation (STEMI) myocardial infarction of unspecified site
21962	06/27/2025	N18.3	Chronic kidney disease, stage 3
21962	06/27/2025	I25.2	Old myocardial infarction
21962	06/27/2025	I10	Essential hypertension
19251	04/16/2024	E11.9	Type 2 diabetes mellitus without complications
19251	05/28/2024	E11.65	Type 2 diabetes mellitus with hyperglycemia
19251	07/25/2024	E11.65	Type 2 diabetes mellitus with hyperglycemia
19251	07/25/2024	K80.11	Calculus of gallbladder with chronic cholecystitis with obstruction

Example 1: Subject 001 Hypertension

```
47                <rwdl:ColumnName>CEOCCUR</rwdl:ColumnName>
48            </rwdl:Coordinate>
49        </rwdl:Target>
50    </rwdl:MapID>
51
52    <rwdl:MapID UUID="7e376beb-7dad-4f5c-a212-88283ac22eba" MethodDefOID="MT.AFTERIDXDATE">
53        <rwdl:Source>
54            <rwdl:Coordinate Storage="FILESYSTEM" Structure="TABULAR" Format="CSV">
55                <rwdl:URI>https://github.com/cdisc-org/DataExchange-RWD-Lineage/blob/main/examples/example1/data/source/pt_dx.csv</rwdl:URI>
56                <rwdl:RowIndex>4</rwdl:RowIndex>
57                <rwdl:ColumnName>DATE</rwdl:ColumnName>
58            </rwdl:Coordinate>
59        </rwdl:Source>
60        <rwdl:Target>
61            <rwdl:Coordinate Storage="FILESYSTEM" Structure="TABULAR" Format="CSV">
62                <rwdl:URI>https://github.com/cdisc-org/DataExchange-RWD-Lineage/blob/main/examples/example1/data/sdtm/ce.csv</rwdl:URI>
63                <rwdl:RowIndex>2</rwdl:RowIndex>
64                <rwdl:ColumnName>CEOCCUR</rwdl:ColumnName>
65            </rwdl:Coordinate>
66        </rwdl:Target>
67    </rwdl:MapID>
68
69    <!-- vitals.csv row 3 -> ce.csv row 4 -->
70    <rwdl:MapID UUID="880af520-a197-44c7-976c-3770bdc46575">
71        <rwdl:Source>
72            <rwdl:Coordinate Storage="FILESYSTEM" Structure="TABULAR" Format="CSV">
73                <rwdl:URI>https://github.com/cdisc-org/DataExchange-RWD-Lineage/blob/main/examples/example1/data/source/vitals.csv</rwdl:URI>
```

Source

← **Comparison to index date using MethodDefOID**

← **4th row of column *DATE* in table PT_DX**

maps to

Target

← **2nd row of column CEOCCUR in table CE**

https://github.com/cdisc-org/DataExchange-RWD-Lineage/blob/main/examples/example1/data/define/rwd-lineage.xml

Example 1: Subject 001 Hypertension

SDTM

CE				
USUBJID	**CESEQ**	**CETERM**	**CEPRESP**	**CEOCCUR**
001	1	MYOCARDIAL INFARCTION	Y	Y
001	2	HYPERTENSION	Y	Y
002	1	MYOCARDIAL INFARCTION	Y	N
002	2	HYPERTENSION	Y	Y

Note that these other points in the SDTM record would have their own MapID's defined

Source RWD

PT_DX			
PT_ID	**DATE**	**ICD10**	**TERM**
21962	05/30/2025	I21.3	ST elevation (STEMI) myocardial infarction of unspecified site
21962	06/27/2025	N18.3	Chronic kidney disease, stage 3
21962	06/27/2025	I25.2	Old myocardial infarction
21962	06/27/2025	I10	Essential hypertension
19251	04/16/2024	E11.9	Type 2 diabetes mellitus without complications
19251	05/28/2024	E11.65	Type 2 diabetes mellitus with hyperglycemia
19251	07/25/2024	E11.65	Type 2 diabetes mellitus with hyperglycemia
19251	07/25/2024	K80.11	Calculus of gallbladder with chronic cholecystitis with obstruction

Example 1: Subject 001 Myocardial Infarction

SDTM

CE				
USUBJID	**CESEQ**	**CETERM**	**CEPRESP**	**CEOCCUR**
001	1	MYOCARDIAL INFARCTION	Y	Y
001	2	HYPERTENSION	Y	Y
002	1	MYOCARDIAL INFARCTION	Y	N
002	2	HYPERTENSION	Y	Y

Patient is assigned ICD10 diagnosis code I21 (Acute myocardial infarction) or I122 (Subsequent ST elevation (STEMI) and non-ST elevation (NSTEMI) myocardial infarction) within the followup period

Source RWD

PT_DX			
PT_ID	**DATE**	**ICD10**	**TERM**
21962	05/30/2025	I21.3	ST elevation (STEMI) myocardial infarction of unspecified site
21962	06/27/2025	N18.3	Chronic kidney disease, stage 3
21962	06/27/2025	I25.2	Old myocardial infarction
21962	06/27/2025	I10	Essential hypertension
19251	04/16/2024	E11.9	Type 2 diabetes mellitus without complications
19251	05/28/2024	E11.65	Type 2 diabetes mellitus with hyperglycemia
19251	07/25/2024	E11.65	Type 2 diabetes mellitus with hyperglycemia
19251	07/25/2024	K80.11	Calculus of gallbladder with chronic cholecystitis with obstruction

Example 1: Subject 001 Myocardial infarction

SDTM

CE				
USUBJID	**CESEQ**	**CETERM**	**CEPRESP**	**CEOCCUR**
001	1	MYOCARDIAL INFARCTION	Y	Y
001	2	HYPERTENSION	Y	Y
002	1	MYOCARDIAL INFARCTION	Y	N
002	2	HYPERTENSION	Y	Y

Patient has records of acute myocardial infarction in notes within the followup period

Source RWD

NOTES		
PT_ID	**DATE**	**TEXT**
21962	06/27/2025	PAST MEDICAL HISTORY: 1. Stage III CKD with baseline creatinine in the 2 range. 2. Status **post MI on May 30, 2024**.
19251	04/16/2024	PAST MEDICAL HISTORY: Type 2 DM, hypothyroidism, hyperlipidemia, HTN, obstructive sleep apnea
19251	07/25/2024	ADMISSION DIAGNOSES: Hyperglycemia, cholelithiasis, obstructive sleep apnea, diabetes mellitus, and hypertension. DISCHARGE DIAGNOSES: Hyperglycemia, cholelithiasis, obstructive sleep apnea, diabetes mellitus, hypertension, and cholecystitis.

Example 1: Subject 002 Hypertension

SDTM

CE				
USUBJID	**CESEQ**	**CETERM**	**CEPRESP**	**CEOCCUR**
001	1	MYOCARDIAL INFARCTION	Y	Y
001	2	HYPERTENSION	Y	Y
002	1	MYOCARDIAL INFARCTION	Y	N
002	2	HYPERTENSION	Y	Y

Patient has newly diagnosed hypertension in clinical notes

Source RWD

NOTES		
PT_ID	**DATE**	**TEXT**
21962	06/27/2025	PAST MEDICAL HISTORY: 1. Stage III CKD with baseline creatinine in the 2 range. 2. Status **post MI on May 30, 2024**.
19251	04/16/2024	PAST MEDICAL HISTORY: Type 2 DM, hypothyroidism, hyperlipidemia, **HTN**, obstructive sleep apnea
19251	07/25/2024	ADMISSION DIAGNOSES: Hyperglycemia, cholelithiasis, obstructive sleep apnea, diabetes mellitus, and **hypertension**. DISCHARGE DIAGNOSES: Hyperglycemia, cholelithiasis, obstructive sleep apnea, diabetes mellitus, **hypertension**, and cholecystitis.

Example 1: Subject 002 Hypertension

SDTM

CE				
USUBJID	**CESEQ**	**CETERM**	**CEPRESP**	**CEOCCUR**
001	1	MYOCARDIAL INFARCTION	Y	Y
001	2	HYPERTENSION	Y	Y
002	1	MYOCARDIAL INFARCTION	Y	N
002	2	HYPERTENSION	Y	Y

Patient has at least two elevated blood pressure measurements (>=130/80 mmHg) over the course of a 3-month period within the followup period

Source RWD

Vitals			
Patno	**Date**	**Vital**	**Value**
21962	06/27/2025	BP	160/115
21962	06/27/2025	BMI	33.6
19251	04/16/2024	BP	142/86
19251	04/16/2024	BMI	28.5
19251	05/28/2024	BP	140/82
19251	05/28/2024	BMI	28.2
19251	07/25/2024	BP	142/85
19251	07/25/2024	BMI	28.3

Example 1: RWD Lineage XML links & tabular view of lineage

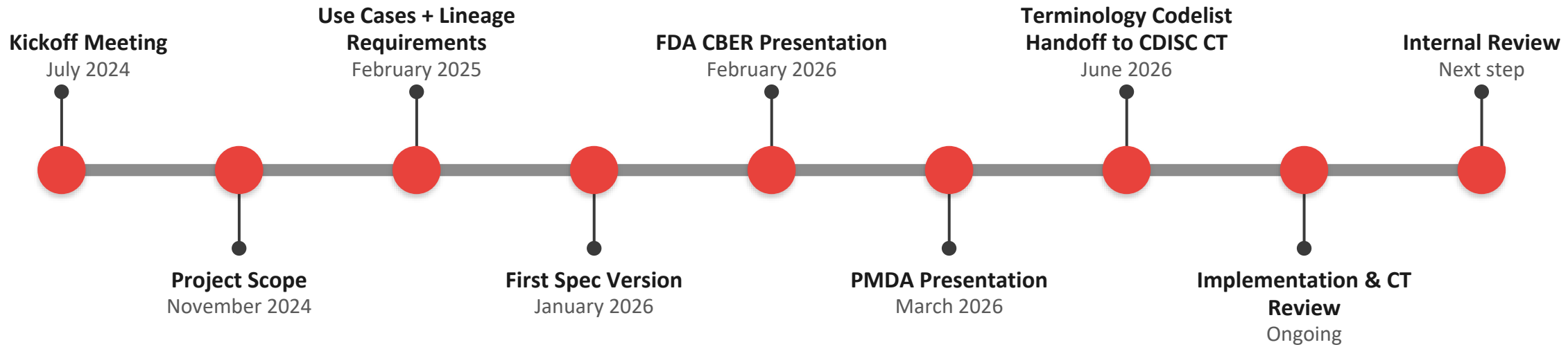
Define-XML: https://github.com/cdisc-org/DataExchange-RWD-Lineage/blob/main/examples/example1/data/define/define.xml
RWD-Lineage-XML: https://github.com/cdisc-org/DataExchange-RWD-Lineage/blob/main/examples/example1/data/define/rwd-lineage.xml

	TARGET					SOURCE							
uuid	storage	structure	URI	rowIndex	columnName	storage	structure	URI	rowIndex	columnName	spanBegin	spanEnd	value
35060134-fc2f-4c	Filesystem	Tabular	https://github.cor	2	CEOCCUR	Filesystem	Tabular	https://github.cor	4	ICD10			I10
77d5b278-692e-4	Filesystem	Tabular	https://github.cor	2	CEOCCUR	Filesystem	Tabular	https://github.cor	4	TERM			Essential hyperte
7e376beb-7dad-4	Filesystem	Tabular	https://github.cor	2	CEOCCUR	Filesystem	Tabular	https://github.cor	4	DATE			06/27/2025
880af520-a197-4	Filesystem	Tabular	https://github.cor	4	CEOCCUR	Filesystem	Tabular	https://github.cor	3	Vital			BP
99e07ea5-bdef-4	Filesystem	Tabular	https://github.cor	4	CEOCCUR	Filesystem	Tabular	https://github.cor	3	Value			142/86
9a5e2706-c919-4	Filesystem	Tabular	https://github.cor	4	CEOCCUR	Filesystem	Tabular	https://github.cor	3	Date			04/16/2024
897e5ede-3913-4	Filesystem	Tabular	https://github.cor	4	CEOCCUR	Filesystem	Tabular	https://github.cor	5	Vital			BP
0a44299a-87c5-4	Filesystem	Tabular	https://github.cor	4	CEOCCUR	Filesystem	Tabular	https://github.cor	5	Value			140/82
b8fb983e-6d14-4	Filesystem	Tabular	https://github.cor	4	CEOCCUR	Filesystem	Tabular	https://github.cor	5	Date			05/28/2024
b70adc19-9e75-4	Filesystem	Tabular	https://github.cor	4	CEOCCUR	Filesystem	Tabular	https://github.cor	7	Vital			BP
44eead9c-8781-4	Filesystem	Tabular	https://github.cor	4	CEOCCUR	Filesystem	Tabular	https://github.cor	7	Value			142/85
76237584-4c5c-4	Filesystem	Tabular	https://github.cor	4	CEOCCUR	Filesystem	Tabular	https://github.cor	7	Date			07/25/2024
0465945a-2b61-4	Filesystem	Tabular	https://github.cor	4	CEOCCUR	Filesystem	Tabular	https://github.cor	2	TEXT	707	710	HTN
7c8b1ce7-dd0d-4	Filesystem	Tabular	https://github.cor	4	CEOCCUR	Filesystem	Tabular	https://github.cor	3	TEXT	100	112	hypertension
39c0d747-67dc-4	Filesystem	Tabular	https://github.cor	4	CEOCCUR	Filesystem	Tabular	https://github.cor	3	TEXT	209	221	hypertension
0f185720-823f-4e	Filesystem	Tabular	https://github.cor	1	CEOCCUR	Filesystem	Tabular	https://github.cor	1	ICD10			I21.3
de098b66-7ac9-4	Filesystem	Tabular	https://github.cor	1	CEOCCUR	Filesystem	Tabular	https://github.cor	1	TERM			ST elevation (ST
3fefb802-1332-4f	Filesystem	Tabular	https://github.cor	1	CEOCCUR	Filesystem	Tabular	https://github.cor	1	DATE			05/30/2025
67e89898-8383-4	Filesystem	Tabular	https://github.cor	1	CEOCCUR	Filesystem	Tabular	https://github.cor	1	TEXT	278	305	acute myocardial
8f0f595a-9d6a-46	Filesystem	Tabular	https://github.cor	1	CEOCCUR	Filesystem	Tabular	https://github.cor	1	TEXT	1211	1234	post MI on May 3

Tabular view: https://github.com/cdisc-org/DataExchange-RWD-Lineage/blob/main/examples/example1/Example1.xlsx (RWDLineage-Table tab)

Project Status

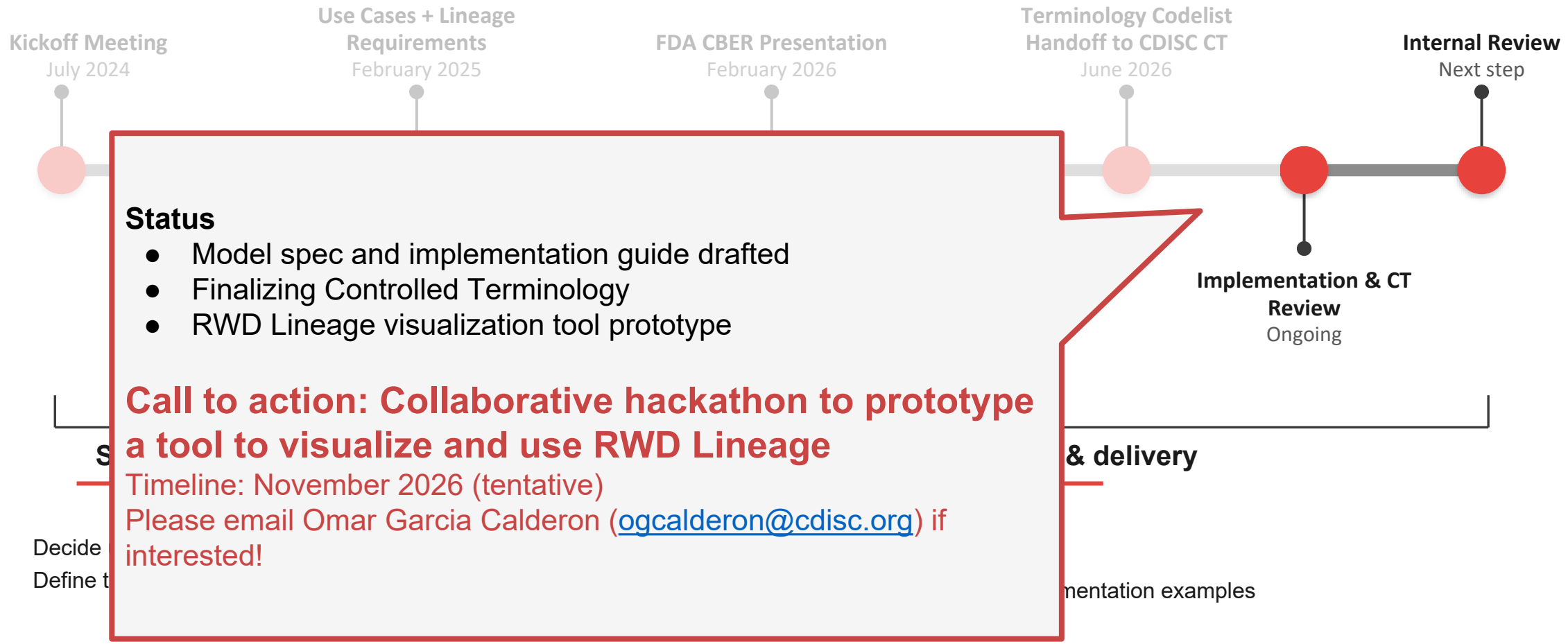
Project Timeline

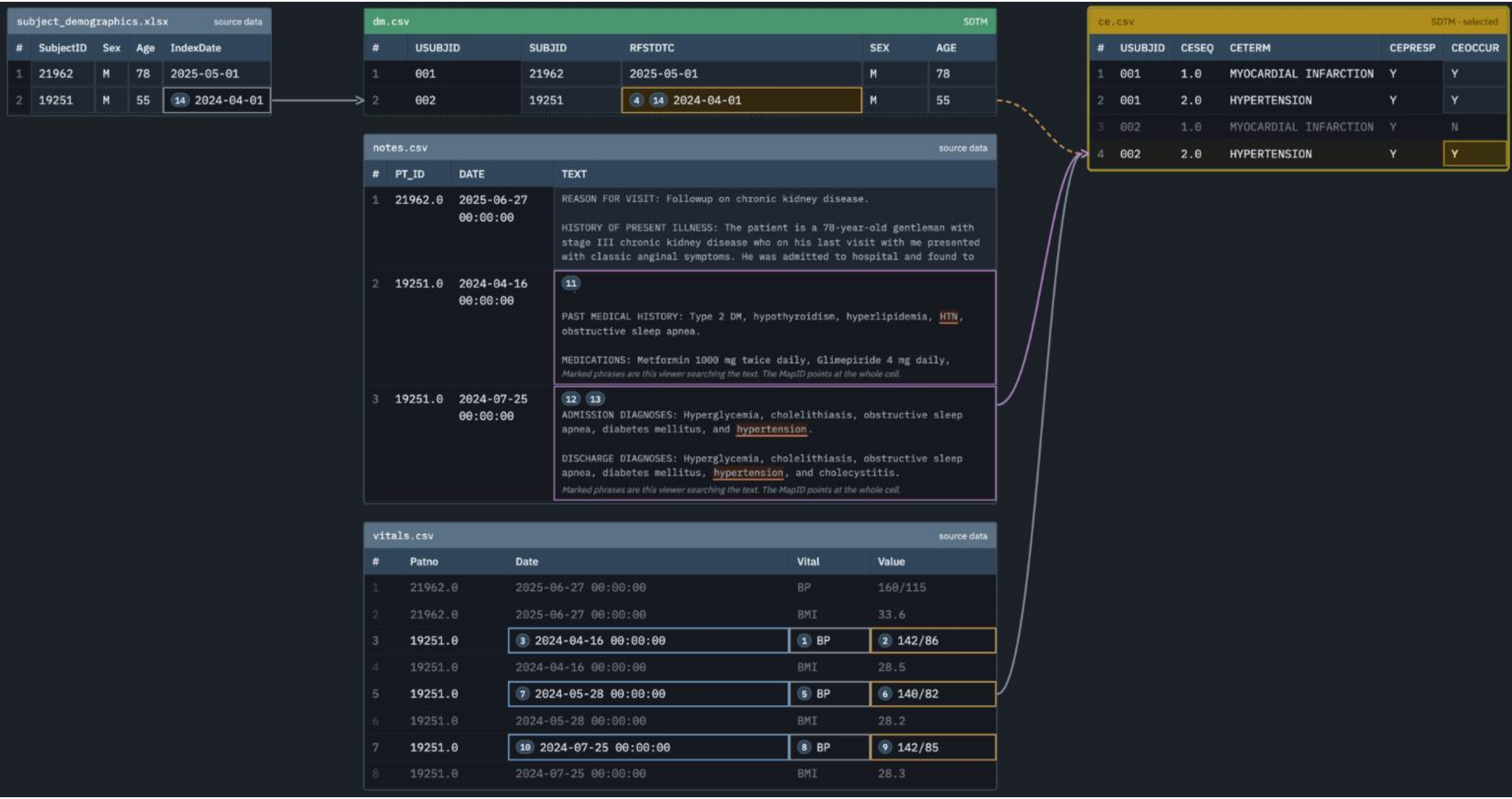


- Decide use cases
- Define timelines

- Implement use cases to inform the lineage
- Fine-tune model specification & build implementation examples
- Package lineage into a standard

Project Timeline




subject_demographics.xlsx
source data
SubjectID Sex Age IndexDate
1 21962 M 78 2025-05-01
2 19251 M 55 14 2024-04-01
dm.csv
SDTM
USUBJID SUBJID RFSTDTC SEX AGE
1 001 21962 2025-05-01 M 78
2 002 19251 4 14 2024-04-01 M 55
ce.csv
SDTM - selected
USUBJID CESEQ CETERM CEPRESP CEOCCUR
1 001 1.0 MYOCARDIAL INFARCTION Y Y
2 001 2.0 HYPERTENSION Y Y
3 002 1.0 MYOCARDIAL INFARCTION Y N
4 002 2.0 HYPERTENSION Y Y
notes.csv
source data
PT_ID DATE TEXT
1 21962.0 2025-06-27 00:00:00
REASON FOR VISIT: Followup on chronic kidney disease.
HISTORY OF PRESENT ILLNESS: The patient is a 78-year-old gentleman with stage III chronic kidney disease who on his last visit with me presented with classic anginal symptoms. He was admitted to hospital and found to
2 19251.0 2024-04-16 00:00:00
11
PAST MEDICAL HISTORY: Type 2 DM, hypothyroidism, hyperlipidemia, HTN, obstructive sleep apnea.
MEDICATIONS: Metformin 1000 mg twice daily, Glimepiride 4 mg daily,
Marked phrases are this viewer searching the text. The MapID points at the whole cell.
3 19251.0 2024-07-25 00:00:00
12 13
ADMISSION DIAGNOSES: Hyperglycemia, cholelithiasis, obstructive sleep apnea, diabetes mellitus, and hypertension.
DISCHARGE DIAGNOSES: Hyperglycemia, cholelithiasis, obstructive sleep apnea, diabetes mellitus, hypertension, and cholecystitis.
Marked phrases are this viewer searching the text. The MapID points at the whole cell.
vitals.csv
source data
Patno Date Vital Value
1 21962.0 2025-06-27 00:00:00 BP 160/115
2 21962.0 2025-06-27 00:00:00 BMI 33.6
3 19251.0 3 2024-04-16 00:00:00 1 BP 2 142/86
4 19251.0 2024-04-16 00:00:00 BMI 28.5
5 19251.0 7 2024-05-28 00:00:00 5 BP 6 140/82
6 19251.0 2024-05-28 00:00:00 BMI 28.2
7 19251.0 10 2024-07-25 00:00:00 8 BP 9 142/85
8 19251.0 2024-07-25 00:00:00 BMI 28.3

Feedback appreciated!

Working version of the RWD Lineage model specification & implementation examples: https://github.com/cdisc-org/DataExchange-RWD-Lineage/tree/main

Please circulate with your colleagues in the RWD space!



Thank You!



Join us as a volunteer!

cdisc