



cdisc® 2026 Europe Interchange

THE FUTURE IS CONNECTED: STANDARDS AND AI POWERING DIGITAL TRANSFORMATION



MILAN, ITALY | MAIN CONFERENCE: 20-21 MAY | TRAININGS & WORKSHOPS: 18, 19, & 22 MAY

Climbing Out of the Silos: Connecting CDISC Standards and Open-Source Tools for End-to-End Automation

John McDade, Associate Director, Phastar

Speakers



John McDade
Associate Director
Phastar

- I have been in the industry programming for almost 20 years and been with Phastar for the last 10 years as part of the Operational Excellence team. Here I enjoy working on a variety of process and tool development initiatives, focused on open-source / AI / web app development.

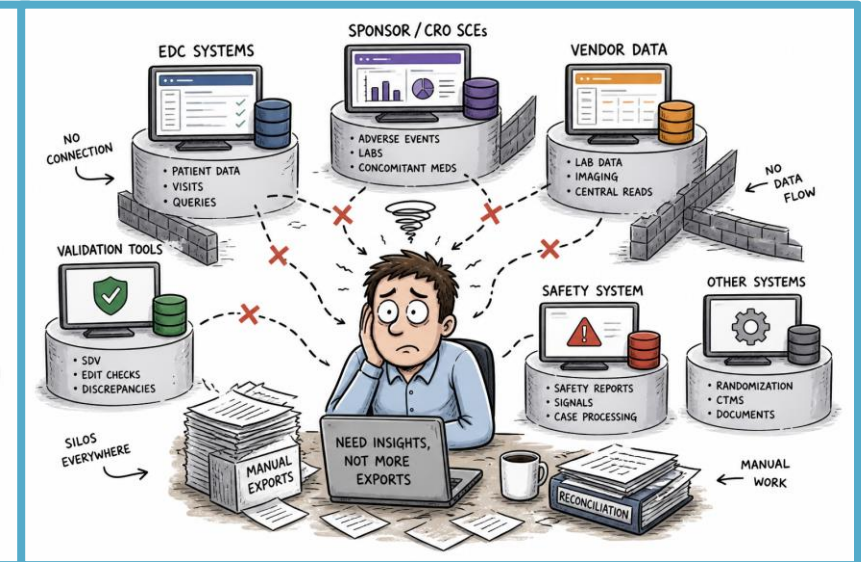
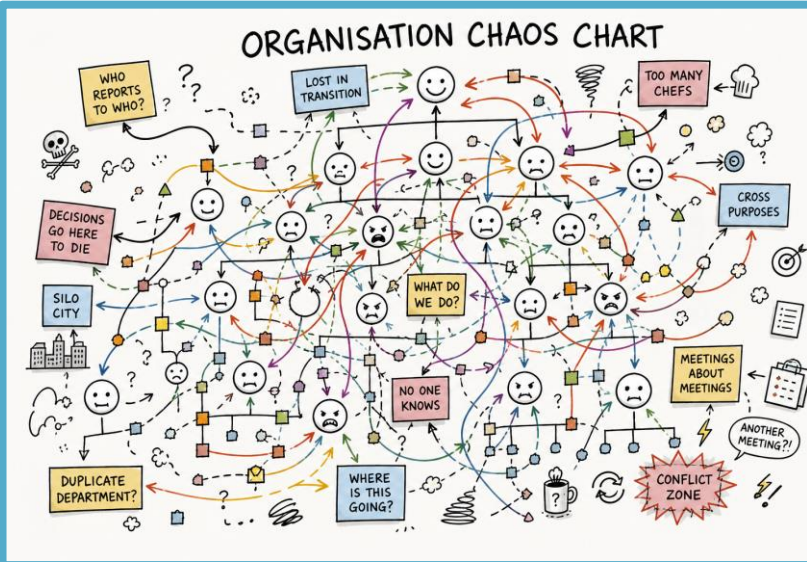
Agenda

1. Silos
2. Enablers
3. JSON Mapping Metadata
4. Automation
5. OAK Adoption
6. Conclusion



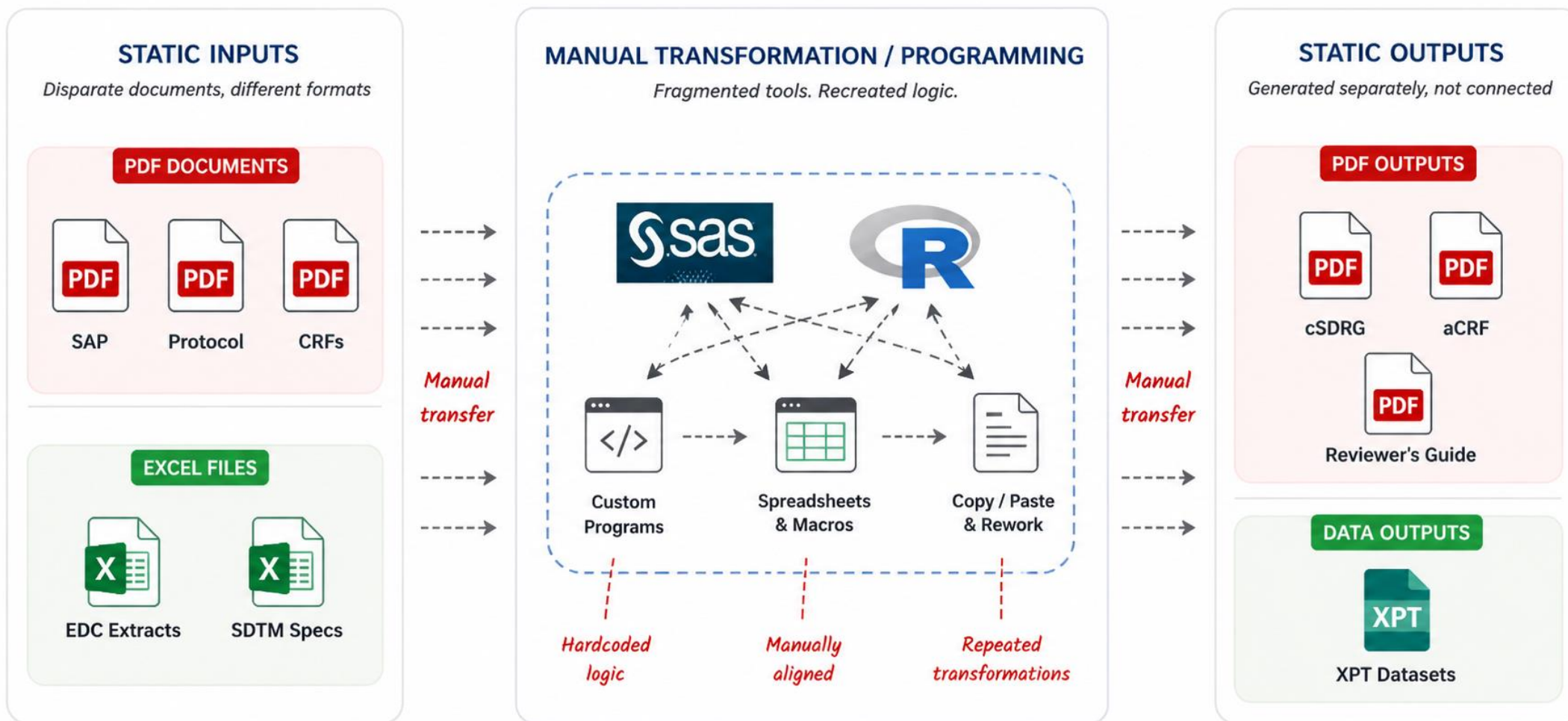
Clinical Trial Silos

The Obvious Silos

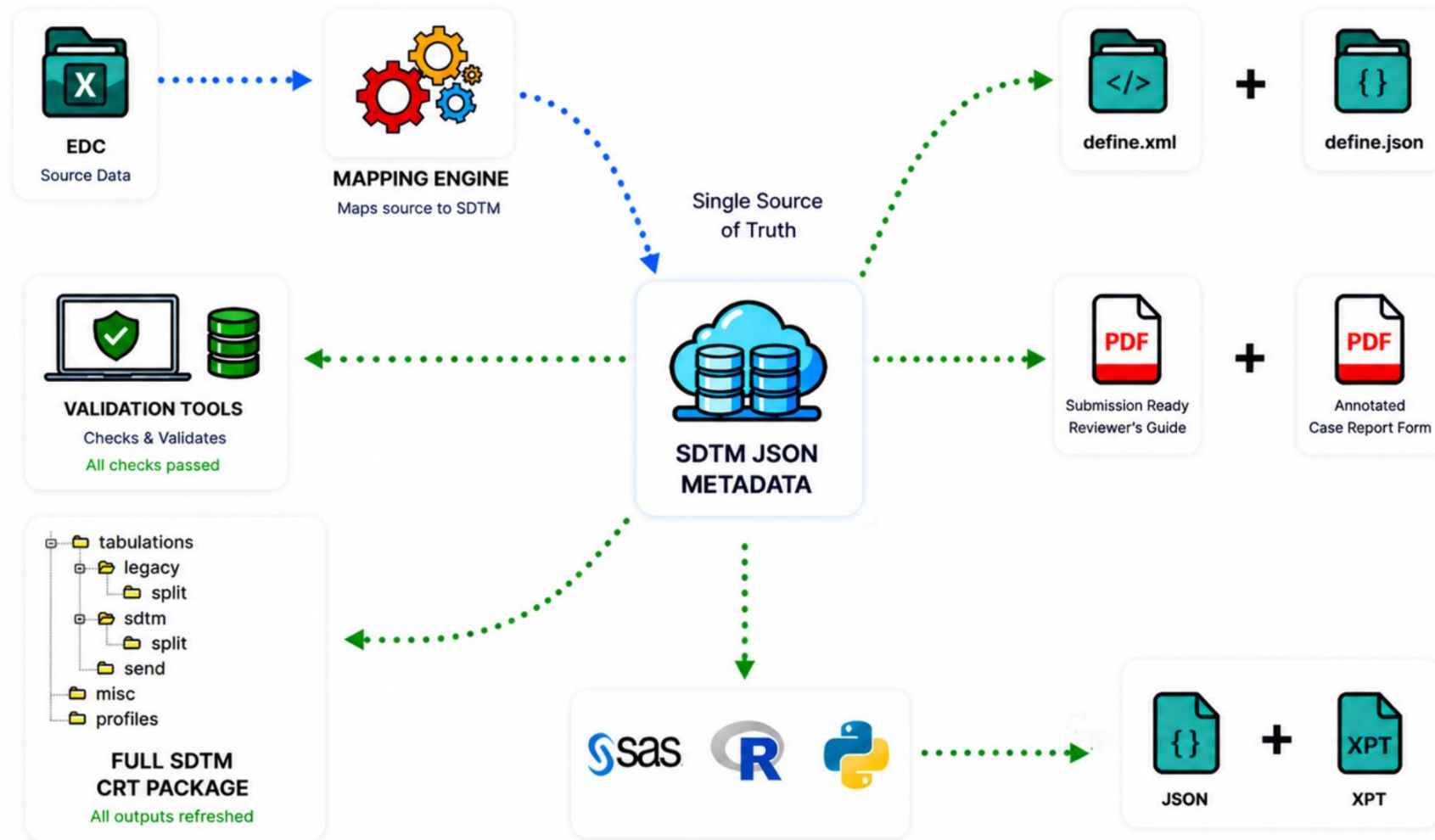


Database go-live / lock alignment across every stakeholder can feel more like a minor miracle than an operational masterstroke

Silos Within Silos – SDTM



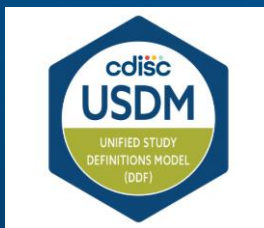
Connected SDTM Workflow





The Enablers

- Dataset JSON / API
- CDISC CORE
- USDM / Biomedical Concepts
- Open-source
- OAK



SDTM JSON Metadata — Codelist & Derivation

Codelist mapping — AE.AEACN

```
{
  "source": {
    "raw_table": "AE",
    "raw_column": "AEACN",
    "raw_label": "Action Taken, Investigational Product",
    "raw_codelist": "$AEACN_2"
  },
  "targets": [{
    "sdm_domain": "AE",
    "sdm_variable": "AEACN",
    "mapping_type": "coded",
    "value_mappings": [
      {
        "raw_value": "0", "raw_decode": "Dose not changed",
        "sdm_value": "DOSE NOT CHANGED" },
      {
        "raw_value": "2", "raw_decode": "Dose reduced",
        "sdm_value": "DOSE REDUCED" },
      {
        "raw_value": "3", "raw_decode": "Drug interrupted",
        "sdm_value": "DRUG INTERRUPTED" },
      {
        "raw_value": "4", "raw_decode": "Drug discontinued",
        "sdm_value": "DRUG WITHDRAWN" }
    ]
  }
]
```



SDTM JSON
METADATA

SAS + R derivations — VS.VSDAT → VSDTC

```
{
  "source": {
    "raw_table": "VS",
    "raw_column": "VSDAT",
    "raw_label": "Vital Signs Date"
  },
  "targets": [{
    "sdm_domain": "VS",
    "sdm_variable": "VSDTC",
    "mapping_type": "derivation",
    "derivation": {
      "sas_code": "catx('T', vmdat, substr(put(input(vstim, time5.), e8601tm.), 1, 5))",
      "r_code": "ifelse(is.na(VSTIM), as.character(VSDAT), paste0(VSDAT, 'T', format(strptime(VSTIM, '%H:%M'), '%H:%M')))"
    }
  }],
  "linked_variables": ["VSTIM"]
}
```

SDTM JSON Metadata — Variable Linking & SUPP

Variable linking (1/2) — VS.HEIGHT (6 SDTM targets, one findings group)

```
{
  "mapping_id": "VS_HEIGHT",
  "mapping_type": "findings_group",
  "source": { "raw_table": "VS", "raw_column": "HEIGHT", "raw_label": "Height" },
  "sdm_condition": "HEIGHT NE ",
  "condition_group_id": "VS_HEIGHT_GROUP",
  "targets": [
    { "sdm_domain": "VS", "sdm_variable": "VSCAT", "sdm_assigned_value": "\"BODY MEASUREMENT\"" },
    { "sdm_domain": "VS", "sdm_variable": "VSTEST", "sdm_assigned_value": "\"Height\"" },
    { "sdm_domain": "VS", "sdm_variable": "VSTESTCD", "sdm_assigned_value": "\"HEIGHT\"" },
    { "sdm_domain": "VS", "sdm_variable": "VSORRES", "sdm_assigned_value": "HEIGHT" },
    { "sdm_domain": "VS", "sdm_variable": "VSSTRESC", "sdm_assigned_value": "HEIGHT" },
    { "sdm_domain": "VS", "sdm_variable": "VSSTRESN", "sdm_assigned_value": "HEIGHT" }
  ]
}
```

Variable linking (2/2) — VS.HEIGHTU (units, joined via the same condition_group_id)

```
{
  "mapping_id": "VS_HEIGHTU",
  "source": { "raw_table": "VS", "raw_column": "HEIGHTU", "raw_codelist": "$VSRES3F" },
  "condition_group_id": "VS_HEIGHT_GROUP",
  "targets": [
    { "sdm_domain": "VS", "sdm_variable": "VSORRESU", "mapping_type": "coded",
      "value_mappings": [
        { "raw_value": "C49668", "raw_decode": "cm", "sdm_value": "cm" },
        { "raw_value": "C48500", "raw_decode": "inch", "sdm_value": "inch" }
      ]
    },
    { "sdm_variable": "VSSTRESU", "mapping_type": "coded",
      "value_mappings": [ /* same as VSORRESU */ ]
    }
  ]
}
```

SUPP variable — EG.INTPR → SUPPEG.EGSPFY

```
{
  "source": {
    "raw_table": "EG",
    "raw_column": "INTPR",
    "raw_label": "Reason, Abnormal  
Overall ECG Evaluation"
  },
  "sdm_condition": "INTP NE \"\"",
  "targets": [{
    "sdm_domain": "SUPPEG",
    "sdm_variable": "EGSPFY",
    "mapping_type": "supp_qualifier",
    "supp_metadata": {
      "QNAM": "EGSPFY",
      "IDVAR": "EGSEQ",
      "QORIG": "CRF",
      "QLABEL": "Specification  
of ECG Finding"
    },
    "sdm_assigned_value":
      "uppercase(INTPR)"
  }]
}
```

Metadata Foundation — the supporting cast

1

CDISC reference data

Controlled Terminology · SDTM IG metadata · client-specific overlays · Biomedical Concepts package

2

Dataset keys management

Sort order and key variables derived from validated IG metadata — feeds dataset sorting, define-XML keys, validation key

3

Trial Design + I/E criteria

TA · TE · TV · TI · TS — protocol-imported Inclusion / Exclusion criteria flow into the TI domain and the Reviewer's Guide



**SDTM JSON
METADATA**

4

CDISC validation findings

CDISC CORE + Pinnacle 21 results — comments managed in-tool, fed straight into the Reviewer's Guide

5

Global derivations (multi-language)

STUDYID, USUBJID, DM-side and cross-domain derivations — SAS / R / Python expressions held centrally and reused across studies

6

Custom code, filters & merge logic

Per-domain SAS macros, R functions, pre / derivation / post snippets · dataset-level WHERE filters · SET vs MERGE vs Custom-Skip strategy

SDTM Mapping JSON



SDTM JSON
METADATA

Vital

13 variables 11 auto-mapped Add Filter

Global Mapping Variables Form Mapping Variables Create Linked Group Search

SDTM Targets: VS (11)

VS.DBPSIT FINDINGS VS 9 variables mapped VSCAT, VSORRES, VSSTRES

Sitting DBP (mmHg)

Category & Test 3

VSCAT	"PULSE AND BLOOD PRESSURE"
VSTEST	"Diastolic Blood Pressure"
VSTESTCD	"DIABP"

Results 3

VSORRES	DBPSIT
VSSTRESC	DBPSIT
VSSTRESN	DBPSIT

Units 2

VSORRESU	"mmHg"
VSSTRESU	"mmHg"

Qualifiers 1

VSPOS	"SITTING"
-------	-----------

FINDINGS Group Mapping

Auto-mapped 10 variables

Overview

Changing domain will auto-update all variables (e.g., EG→VS:1)

Available FINDINGS Variables for VS

Values auto-fill: ORRES → STRESC, ORRESU → STRESU. Use S

✓ VSTESTCD Req Topic

Vital Signs Test Short Name

<> Derivation

"DIABP"

✓ VSTEST Req Synonym Qualifier

Vital Signs Test Name

<> Derivation

"Diastolic Blood Pressure"

✓ VSCAT Perm Grouping Qualifier

Category for Vital Signs

```
144 # Linked Group: cond_DBPSIT_f7b8240c
145 # Condition: DBPSIT NE .
146 vs_cond_dbpsit_f7b8240c <- vs_data %>%
147   filter(!is.na(DBPSIT) & nzchar(DBPSIT))) %>%
148   mutate(
149     VSTESTCD = 'DIABP',
150     VSCAT = 'PULSE AND BLOOD PRESSURE',
151     VSORRES = as.character(DBPSIT),
152     VSORRESU = 'mmHg',
153     VSPOS = 'SITTING',
154     VSSTRESC = as.character(DBPSIT),
155     VSSTRESN = DBPSIT.
```

```
# =====
# Combine all topic datasets (each topic = one measurement = one row per subject-visit)
# =====
```

```
vs_combined <- bind_rows(
  vs_vs_height_group,
  vs_vs_temp_group,
  vs_vs_weight_group,
  vs_cond_dbpsit_f7b8240c,
  vs_cond_pulsesit_ce124ff1,
  vs_cond_sbpsit_4aa5a4d7,
  vs_cond_vsuperf_2e0ba03d
)
```

Dataset Keys

2

Dataset keys management

Sort order and key variables derived from validated IG metadata

Study-Specific SDTM Metadata

28 datasets

Edit study-specific metadata for Define.xml generation. Changes are saved to this study only.

Q Search datasets, labels, or descriptions...

All Classes

Show All 47

Dataset	Label	Class	Description	Purpose	Keys	Actions
AE	Mapped Adverse Events	Events	One record per adverse event per subject	Tabulation	STUDYID, USUBJID, AESTDTC, AEENDTC, AEDECOD, AESPID	Edit
CE	Mapped Clinical Events	Events	One record per event per subject	Tabulation	STUDYID, USUBJID, CESTDTC, CECAT, CETERM	Edit
CM	Mapped Concomitant/Prior Medications	Interventions	One record per recorded intervention occu...	Tabulation	STUDYID, USUBJID, CMSTDTC, CMENDTC, CMCAT, CMSCAT, CMTRT, CMLOC, CMSPID	Edit

STUDYID, USUBJID, EGDTC, EGCAT, EGTESTCD, VISITNUM, EGSPID, QNAM MODULE

Edit

```
# NOTE: Sort includes SUPP variable(s): MODULE
# oak: derive_seq(tgt_var='EGSEQ', rec_vars=c('STUDYID', 'USUBJID', 'EGDTC', 'EGCAT', 'EGTESTCD', 'VISITNUM', 'EGSPID', 'MODULE'))
# SEQ increments within each subject (STUDYID + USUBJID)
eg_seq <- eg_data %>%
  arrange(STUDYID, USUBJID, !is.na(EGDTC), EGDTC, !is.na(EGCAT), EGCAT, !is.na(EGTESTCD), EGTESTCD, !is.na(VISITNUM), VISITNUM, !is.na(EGSPID), EGSPID, !is.na(MODULE), MODULE) %>%
  group_by(STUDYID, USUBJID) %>%
  mutate(EGSEQ = row_number()) %>%
  ungroup()
```

```
*****;
** Code to create EGSEQ variable. **;
*****;

proc sort data=eg_addvars1;
  by studyid usubjid egdte egcat egtestcd visitnum egspid module;
run;

data eg_seq;
  set eg_addvars1 (drop=egseq);
  by studyid usubjid egdte egcat egtestcd visitnum egspid module;
```

EG [SDTMIG 3.4]	ECG Test Results	FINDINGS	One record per one record per ecg observation per replicate per time point or one record per ecg observation per beat per visit per subject	Tabulation	STUDYID, USUBJID, EGDTC, EGCAT, EGTESTCD, VISITNUM, EGSPID, QNAM.MODULE	eg.xpt
---------------------------------	------------------	----------	---	------------	---	------------------------

SDTM Mapping JSON in Action

The screenshot displays the SDTM Mapper application interface. On the left is a sidebar with navigation links: Dashboard, Clients, SDTM Standards, Administration, User Management, Client Management, Project Management, and Role Management. The main area is titled 'Adverse Events (+1 more)' and shows a progress bar at 100%. Below the progress bar, there are filters for 'All', 'Incomplete', 'Complete', and 'Excluded'. A search bar for datasets is present. The 'AE' dataset is highlighted with a 100% mapping progress indicator (197/197 mapped). Other datasets shown include AESI (62% mapped, 18/29 mapped), BMRESO (100% mapped, 14/14 mapped), and CAPRX. The main table lists SDTM Targets with columns for the target name, mapping status, and a link to view coded values. The table includes rows for AE.AD, AE.AEACN, AE.AEACN1, AE.AEACN2, AE.AEACN3, AE.AEACN4, AE.AEACN5, AE.AEACN6, AE.AEACN7, and AE.AEACN8. Each row shows the target name, a status indicator (e.g., SUPPAE), a link to view coded values, and a link to view 5 coded values. The 'auto-mapped' status is indicated for each row.

SDTM Mapper

Dashboard
Clients
SDTM Standards
Administration
User Management
Client Management
Project Management
Role Management

Adverse Events (+1 more) 100%

All Incomplete Complete Excluded

Search projects, standards... Search datasets...

197 variables 177 auto-mapped AETN="C49488" and AETERM NE ""

Global Mapping Variables Form Mapping Variables Create Linked Group Search mapping Show Excluded Restore Excluded (13) Status

Export

SDTM Targets: AE (31) FA (155)

Target	Mapping Status	Link	Link
AE.AD → Additional Drug	SUPPAE	SUPP Qualifier QNAM: AD	1 vars auto-mapped
AE.AEACN → Action Taken, Investigational Product	AE.AEACN ✓	Coded (5)	Click to view 5 coded values auto-mapped
AE.AEACN1 → Action Taken, Investigational Product 1	SUPPAE	SUPP Qualifier	1 vars Coded (4) Click to view 4 coded values auto-mapped
AE.AEACN2 → Action Taken, Investigational Product 2	SUPPAE	SUPP Qualifier	1 vars Coded (5) Click to view 5 coded values auto-mapped
AE.AEACN3 → Action Taken, Investigational Product 3	SUPPAE	SUPP Qualifier	1 vars Coded (5) Click to view 5 coded values auto-mapped
AE.AEACN4 → Action Taken, Investigational Product 4	SUPPAE	SUPP Qualifier	1 vars Coded (5) Click to view 5 coded values auto-mapped
AE.AEACN5 → Action Taken, Investigational Product 5	SUPPAE	SUPP Qualifier	1 vars Coded (4) Click to view 4 coded values auto-mapped
AE.AEACN6 → Action Taken, Investigational Product 6	SUPPAE	SUPP Qualifier	1 vars Coded (4) Click to view 4 coded values auto-mapped
AE.AEACN7 → Action Taken, Investigational Product 7	SUPPAE	SUPP Qualifier	1 vars Coded (4) Click to view 4 coded values auto-mapped
AE.AEACN8 →	SUPPAE	SUPP Qualifier	1 vars Coded (4) Click to view 4 coded values auto-mapped

I/E criteria

3

Trial Design + I/E criteria

TA · TE · TV · TI · TS — protocol-imported Inclusion / Exclusion criteria flow into the TI domain and the Reviewer's Guide

Inclusion/Exclusion Criteria (TI)

Define eligibility criteria. Criteria marked for Reviewer's Guide will be exported for the I/E criteria section.

Export for RG

Add Criterion

All (49)

Inclusion (14)

Exclusion (35)

I/E Criterion Short Name	I/E Criterion (Full)	SDTM Version (200 char)	I/E Criterion Category	I/E Criterion Subcategory	Protocol Criteria Version
--------------------------	----------------------	-------------------------	------------------------	---------------------------	---------------------------

Clients > Training Area > DEMO STUDY 0002 > Trial Design

Overview

Trial Arms 3

Elements 5

Visits 14

I/E Criteria 49

Trial Summary 27

Trial Design Overview

Study: DEMO STUDY 0002

Epochs: Screening Treatment Follow-up

Arms: C B A

Anchor visit defined (Day 0 reference)

Clear All Data

Reset to Template

Import USDM

Export SDTM Datasets

Amendment Version			
1.0	INCLUSION	INCL01	Participants must be ≥ 18 years of age (or the legal age of consent in the jurisdiction in which the study is taking place) at the time of signing the ICF.
1.0	INCLUSION	INCL02	Adult females, pre/peri-menopausal and/or post-menopausal, and adult males (a) Pre/peri-menopausal women (ie, those who do not meet the criteria for post-menopausal defined below) can be enrolled if

CDISC Validation

4

CDISC validation findings

CDISC CORE — comments managed in-tool, fed straight into the Reviewer's Guide

SDTM Domains

AE750

DM46

EX11

Select All

Clear All

Validation Rules

CORE-00019132

CORE-0001968

CORE-0002072

CORE-0002661

CORE-0005169

CORE-0007081

CORE-0007111

CORE-0007141

CORE-00072612

CORE-0007665

CORE-000767735

Select All

Clear All

Conformance

Datasets

Issue Summary

Issue Details

Rules Report

Issue Summary

Summary of validation issues by dataset and rule

Dataset	CORE-ID	Message	Issues	Explanation
AE.json	CORE-000207	AESTDTC is present in a Findings general observation class	1	THIS IS A
AE.json	CORE-000266	If AESER = "N" then none of the seriousness criteria (AESCAN, AESCONG, AE...	1	TEST
AE.json	CORE-000196	AECAT is equal to AEDECOD.	8	WORKING IN
AE.json	CORE-000766	Related record is present in the parent domain dataset and "Other, specify" v...	5	EXCEL FILES
AE.json	CORE-000767	Parent record exists and AEDECOD is not null, but FAOBJ is not equal to AEDE...	735	IS A SILO!!!

4.2 Issues Summary

Validation Report: report.json

Validation Date: 2026-01-16T12:16:37.793516

CDISC CORE Version: sdtmig 3-4

The following validation issues were identified (807 total):

Dataset	CORE-ID	Diagnostic Message	Count	Explanation
AE	CORE-000207	AESTDTC is present in a Findings general observation class	1	THIS IS A
AE	CORE-000266	If AESER = "N" then none of the seriousness criteria (AESCAN, AESCONG, AESDISAB, AESDTH, AESHOSP, AESLIFE, AESOD, AESMIE) could be equal to "Y".	1	TEST
AE	CORE-000196	AECAT is equal to AEDECOD.	8	WORKING IN
AE	CORE-000766	Related record is present in the parent domain dataset and "Other, specify" value was coded.	5	EXCEL FILES
AE	CORE-000767	Parent record exists and AEDECOD is not null, but FAOBJ is not equal to AEDECOD.	735	IS A SILO!!!
DM	CORE-000714	RFXSTDTC is greater than RFXENDTC	1	



Search projects, standards...



Clients > Training Area > DEMO STUDY 0002 > CDISC Validation

CDISC Validation

Validate SDTM datasets for DEMO STUDY 0002

CDISC Core

Pinnacle 21

Validation Configuration

Select the CDISC standard, version, and controlled terminology package

Standard

SDTM ▾

Configuration

SDTM-IG 3.4 (FDA) ▾

SDTM CT

2025-03-28 ▾

CDISC CORE Validation History

View and manage previous validation runs



AE, DM, EX +1



May 12, 12:55 PM

805 Issues



4 runs



Source Data

Upload SDTM datasets in XPT or JSON format

RAW Define.xml

Define.xml Generation

Generate Define 2.0 XML documentation using protocol information, dictionaries, documents, and mapped SDTM metadata

Protocol Info
Configured

Dictionaries
Database-backed

Documents
Database-backed

> Domain Inclusion 33 of 33 Included

Generate Define 2.0 XMLView Define 2.019/03/2026, 16:38:41

Generate Define 2.1 XMLView Define 2.119/03/2026, 14:04:09

Generate Raw Spec DefineView Raw Spec18/03/2026, 12:00:42

CMTRT (TRANSF.BLDPROD → SDTM)

Permitted Value (Code)	Display Value (Decode)
1 (PRBCs)	PRBCs
C12434 (Blood)	BLOOD
C12520 (Platelet)	PLATELET
C13356 (Plasma)	PLASMA
5 (Cryoprecipitate)	CRYOPRECIPITATE
6 (Exchange transfusion)	EXCHANGE TRANSFUSION

CMTRT [VLM](#)

▶ QNAM = "CXRATEC"	Radiotherapy Technique	text		200	CXRATEC (Raw→SDTM) [8 Terms]	CRF Annotated Case Report Form [723] [731] [771] [771] CAPRXR.CXRATEC
▶ QNAM = "CXRATECO"	Radiotherapy Technique, Other	text		200		CRF Annotated Case Report Form [723] [731] [771] [771] CAPRXR.CXRATECO
▶ QNAM = "CXRDREGO"	Radiotherapy Site/Region Location Other	text		200		CRF Annotated Case Report Form [723] [731] [771] [771] CAPRXR.CXRDREGO
▶ QNAM = "CXTRTST"	Treatment Status	text		200	CXTRTST (Raw→SDTM) [8 Terms]	CRF Annotated Case Report Form [711] [717] [717] CAPRX.TREATSTS
▶ QNAM = "FAILREAS"	Reason for Therapy Failure	text		200	FAILREAS (Raw→SDTM) • "1" = "Toxicity" • "2" = "Progression on therapy" • "3" = "Progression following completion of therapy" • "98" = "Not applicable"	CRF Annotated Case Report Form [711] [717] [717] CAPRX.FAILREAS

Blood Product 1 = PRBCs
Type

C12434 = Blood
C12520 = Platelet
C13356 = Plasma
5 = Cryoprecipitate
6 = Exchange transfusion

SDTM OAK — Lessons Learned



What we could adopt

Patterns

- Topic-first findings assembly with `bind_rows()` — the canonical `findings_domain` pattern
- Programming workflow: `read raw` → `oak_id_vars` → `CT` → `topic` → `qualifiers` → `bind_rows` → `derived vars` → `labels`
- Reusable algorithms concept — 22 algorithms cover 16,000+ Roche mappings, language-agnostic

Functions used directly

- `derive_seq()` · `derive_study_day()` · `derive_bfll()` — invoked as-is
- `library(sdtm.oak)` loaded by every generated R program

Semantic alignment retained

- # oak: `assign_no_ct(...)` / `hardcode_no_ct(...)` / `derive_seq(...)` annotations preserve OAK intent in every block
- Reference URL cited in every program header (pharmaverse.github.io/sdtm.oak)

What we had to adapt

Specification Dependent CT

- `ct_spec` as CSV / Excel — no CDISC Library hook, no CT-version awareness
- No SDTM IG version awareness; variable length / label / format = “Yet to be developed”

Code Automation

- Code can be verbose chaining many functions depending on e.g. number of `xxTESTCD` — needs automation layer

SUPP metadata support

- `generate_sdtm_supp()` is a splitter — user hand-builds `supp_qual_info` (QNAM / QLABEL / QORIG)

Out of scope

- Trial Design · SE · SV · RELREC · Associated Person · EPOCH — all out of scope for v0.2

Define.xml / Reviewer's Guide Integration

- These items need to be in sync with SDTM mapping for ‘single source of truth’



Conclusion

Holding SDTM mapping as a JSON structure with supporting CDISC standards and study level metadata, we can vastly improve heavily siloed workflows. APIs connect the metadata sources, ensuring updates are only ever in one place.

Outputs in the form of SDTM code, define.xml, reviewer's guides become rendered artifacts of the existing metadata rather than considerable manual effort, driven from a single source of truth.

Workflows are deterministic, easily defensible, reproducible and sustainable.



Thank You!





Questions?

