2026 Korea CDISC Day

29 October 2026 · Seoul, Korea

From Protocol to Data

: Clinical Trial Digitalization with CDISC USDM and AI

Eunhye Lee, Director C&R Research

29 October 2026



Agenda

1. Reusing Information Written Once
2. From Documents to Data — USDM · MDR · SDR
3. What We Actually Built
4. How Far We Have Come
5. Closing

Session title: The Future of CDISC: DDF, USDM, and Next-Generation Data Standards

Speaker



Eunhye Lee

Director
C&R Research

Eunhye Lee is a Director at C&R Research, where she leads the Clinical Data Intelligence (CDI) Team. She has over 18 years of experience in biostatistics and more than 15 years of expertise in CDISC standards, including SDTM and ADaM.

Her work focuses on clinical biostatistics, CDISC data standards, and the application of digital technologies and AI to enhance statistical analysis workflows, data quality, and efficiency in clinical trials.

She holds a B.S. in Mathematics and Statistics from Sungkyunkwan University and an M.S. and Ph.D. in Biostatistics from Yonsei University.

Generating Statistical Documents with USDM

From Protocol to SAP · TFL Shell · SDTM Spec · ADaM Spec

SECTION 01

Reusing Information Written Once

Today: the same value goes into four documents

- A real case: one sentence in four documents
- Copying forward vs. reading forward

One Sentence Goes into Four Documents

Document	Where it appears	Actual content	Author
Protocol	9.2.3.2 Secondary efficacy endpoint	MSSBP decrease of ≥20 mmHg and MSDBP decrease of ≥10 mmHg from baseline	Clinical
SAP	10.3 Secondary efficacy analysis	Same definition + analysis method (chi-square / Fisher)	Statistics
TFL Shell	14.2.2.2 Footnote [1]	responder = MSSBP ≥20 mmHg and MSDBP ≥10 mmHg reduction	Statistics
ADaM Spec	ADEFF.CRIT1 / CRIT1FL	Criterion sentence + missing after LOCF is a non-responder	Programming

Starting point The same value goes into four documents. What would change if it were written once and all four documents shared it?

The Current Flow and the Flow We Want

AS-IS

From document to document

- A person reads the Protocol and re-types it into the SAP
- A person reads the SAP and re-types it into the TFL Shell
- A person again re-types it into the SDTM / ADaM Specs
- The origin of each value lives only in people’s memory
- When one place changes, people hunt down the rest

TO-BE

Write once, read many times

- Structure the Protocol once
- SAP · TFL · SDTM · ADaM read that structure
- Each value keeps a record of where in the protocol it came from
- Fix the structure in one place and the readers follow
- Items that could not be filled show up as blanks

SECTION 02

From Documents to Data

USDM · MDR · SDR

- Writing for people vs. writing for machines
- Why standards matter, and the three layers
- The full pipeline on one page

Writing as a Document vs. Writing as Data

Document

Written to be read by people

- Written as sentences; wording differs from person to person
- Where things are is remembered by people
- To find something, you open the document and read it by eye
- When something changes, people propagate it to other documents

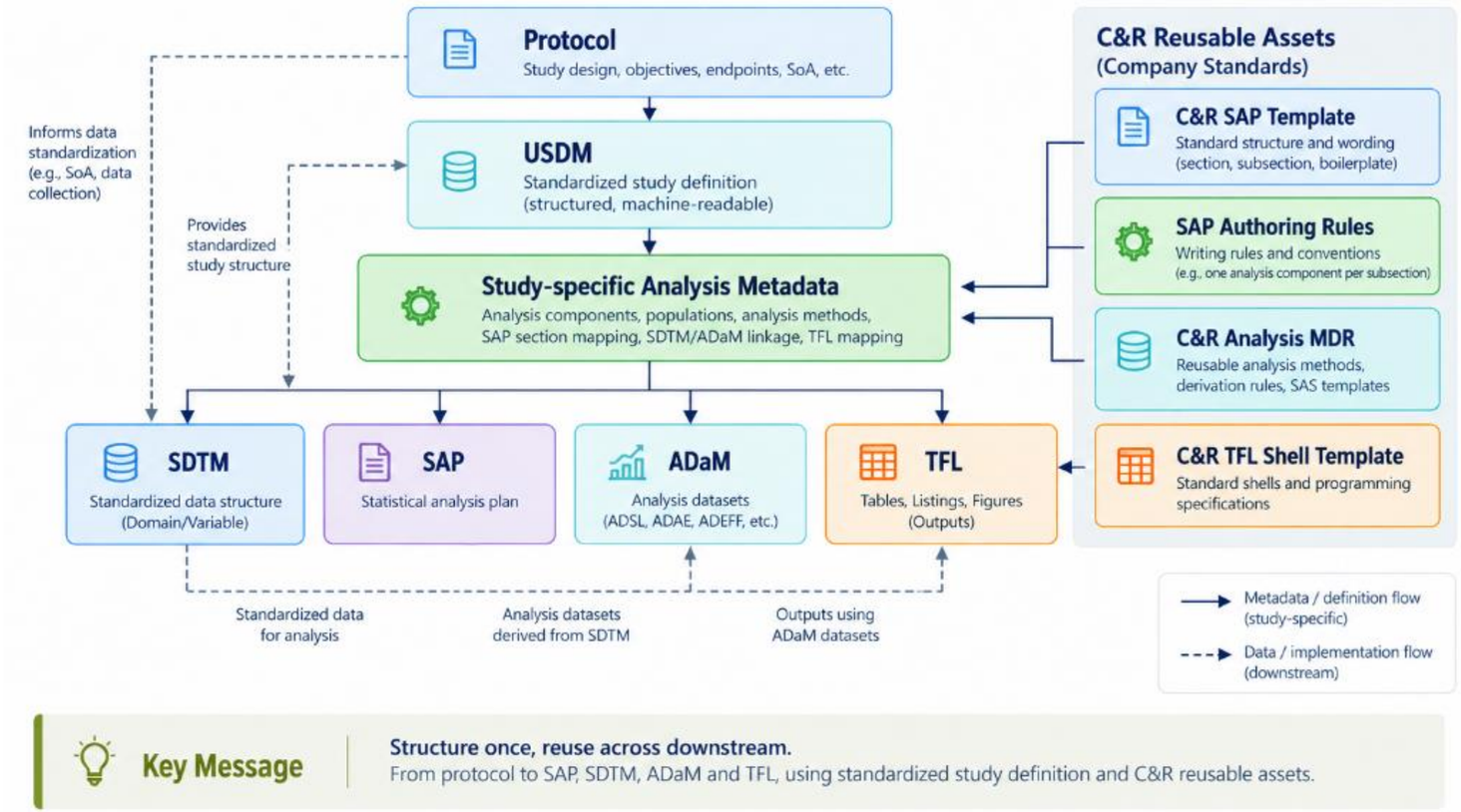
Data

Written so machines can find and use it

- Written as items; names and positions are fixed
- Rules define what is located where
- Look it up by name and the value comes back
- When something changes, the readers follow

Study Definition to Downstream Outputs

Study-specific flow with Reusable assets



What USDM Does

What it is	A common set of rules for describing the design of a single study as structured items. Objectives, endpoints, treatment arms, visits and analysis populations are recorded under the same names and in the same places, no matter who writes them.
Why it is needed	Without rules, every company and every system writes things differently. With rules, different systems can read the same file.
Who created it	Developed by CDISC and TransCelerate under the Digital Data Flow initiative. Recent versions are being aligned with the ICH M11 electronic protocol standard.
What it is not	It does not replace the document. It restates what the document says once more, in a form machines can use.

Scope for today We do not cover the detailed structure of USDM. We focus on how downstream documents change once this file exists.

What Is Standard and What Is This Study

Item	MDR — what belongs in the standard	SDR — what belongs to this study
Responder determination	Variable structure called CRITy / CRITyFL	20 mmHg · 10 mmHg · AND combination
Laboratory tests	A dictionary stating that uric acid is written as URATE	The 43 tests this study collects
Analysis populations	Standard terms FAS / PPS / SS	FAS = at least 1 dose + at least 1 post-baseline measurement
Change from baseline	Derivation rule CHG = AVAL − BASE	Which parameter, at which time point
Tables	Table skeleton · statistics blocks · footnote patterns	5 tables · Week 4 / 8 time points · treatment group names

In one sentence Will this value stay the same in the next study? If so, it is MDR. If you have to read the protocol to know it, it is SDR.

How One Protocol Becomes Deliverables

What actually happens at each step

Step	What happens	Output of this step
1 **Read the protocol**	Split the protocol into sections and record where in the document each sentence came from.	Section list **130 sections · 102 with body text**
2 **Convert to a standard format**	Restate the content as items under the common USDM rules: objectives, endpoints, analysis populations, etc.	USDM file **3 endpoints · 3 populations · 2 blanks**
3 **Define the analyses**	Fix what to analyze, when, in whom and how for this study. The responder criteria are also set here.	Analysis definition for this study **4 analyses · responder criteria 20 / 10 mmHg**
4 **Attach company standards**	Connect standards we already have instead of rebuilding: terminology, domain structure, table skeletons, derivation rules.	Standard assets **43 tests · 4 domains · 5 table skeletons**
5 **Generate deliverables**	Combine the study definition (step 3) with the standards (step 4) to generate documents and specifications. Each value’s source is recorded.	SAP · 5 TFL tables **SDTM 4 domains · ADaM 3 datasets**

Glossary

Terms used in “How One Protocol Becomes Deliverables”

USDM

Unified Study Definitions Model

A common CDISC rule set for describing study design as items. It holds “what this study is.”

MDR

Metadata Repository

The repository for standard metadata. It holds what is common to all studies, such as terminology dictionaries, domain structures and table skeletons.

Analysis Definition

Study Analysis Metadata

What to analyze, when, and how in this study. It is not a standard; it is created anew for each study.

From Study Definition to Connected Clinical Development

A connected workflow — structure the study once, then reuse, trace, and manage change.

1 UNDERSTAND THE STUDY

Protocol
↓
Protocol Section Registry
↓
USDM

2 DEFINE THE ANALYSIS

SAP Requirement Map
↓
PASS / GAP Check
↓
Study-specific Analysis Metadata
↓
\+ Reusable MDR

3 CREATE DOWNSTREAM OUTPUTS

SAP
↓
SDTM Specification
↓
ADaM Specification
↓
C&R TFL Shell

4 CONNECT & CONTROL

End-to-End Traceability
↓
Amendment
↓
Impact Detection
↓
Affected Outputs Review

Structure → Define → Reuse → Connect → Trace → Manage Change

SECTION 03

What We Actually Built

Results of generating four deliverables from one protocol

- The moment a document becomes data, and traceability
- Standard terminology lookup · SDTM / ADaM specifications
- One sentence running through four documents

The Moment a Document Becomes Data

Screen

registry.json → usdm_P3.json

Section decomposition

The protocol was split into 130 sections; 102 (78%) have body text. Each section stores its position in the document and a content hash.

Structured result

2 Objectives · 3 Endpoints · 3 AnalysisPopulations. The secondary efficacy endpoint definition comes in whole, including the response-rate sentence.

What could not be confirmed

2 gaps. The primary and secondary objective sections were not found. They are left blank rather than filled by guesswork.

Nature of this tool It does not fill in what it cannot find with something plausible. The very fact that it could not fill it in is recorded in the deliverable.

One Endpoint Splits into Several Analyses

USDM endpoints and actual statistical analyses are not 1 : 1

USDM endpoint		Analysis definition for this study · study_map.yaml
Endpoint_1 Primary Efficacy	→	**ANA_PRIMARY** TFL 14.2.1.1 Week 8 MSSBP change · ANCOVA · FAS
Endpoint_2 Secondary Efficacy	→	**ANA_SEC_MSSBP_W4** TFL 14.2.1.1 Week 4 MSSBP change · ANCOVA · FAS
	→	**ANA_SEC_MSDBP** TFL 14.2.2.1 Week 4 · 8 MSDBP change · ANCOVA · FAS
	→	**ANA_SEC_RESP** TFL 14.2.2.2 Week 4 · 8 response rate · chi-square / Fisher · sensitivity Logistic · FAS
Endpoint_3 Safety	→	**Safety analysis** TFL 14.3.1.1 Adverse event summary · TEAE definition rule not yet fixed

Why this is needed Endpoint_2 contains three endpoints. USDM only goes as far as “what this study is”; how to split it into analyses is the job of this study’s analysis definition. Safety has no rule yet, so it remains blank.

Lookup in a Standard Dictionary, Not Translation

Protocol term	SDTM standard code	Standard name	NCI code	Status
Uric acid	URATE	Urate	C64814	**Matched**
Creatinine	CREAT	Creatinine	C64547	**Matched**
γGT	GGT	Gamma Glutamyl Transferase	C64847	**Matched**
CPK	CK	Creatine Kinase	C64489	**Matched**
HbA1c	HBA1C	Hemoglobin A1C	C64849	**Matched**
eGFR(CKD-EPI)	Unconfirmed	Unconfirmed	—	**To confirm**

Note Abbreviations and standard codes often differ in practice. This conversion is the result of a CDISC terminology lookup, not an AI translation.

Points to Note in the SDTM · ADaM Specifications

Deliverable	Scope	Points to note
SDTM Spec	DM · VS · LB · AE (EX out of scope)	The mean of three BP readings is calculated by the EDC and recorded in the CRF. Neither SDTM nor ADaM recalculates it; both use that value as is. The role of the standard is not to calculate but to record where the value came from.
SDTM Spec	43 tests	Bilirubin and Glucose have the same test code in blood and urine. LBCAT and LBSPEC distinguish them.
ADaM Spec	ADSL · ADEFF · ADAE	The variables were not decided first and the tables fitted afterwards. We first wrote down which cell of which table requires which variable, then derived the variables backwards.
ADaM Spec	ADEFF: 24 variables	Each variable shows which TFLs use it. CRIT1FL is used only in 14.2.2.2.

Why work backwards Variables the tables do not require are not created. The specification always has a reason for why it looks the way it does.

Final Traceability — How Far One Item Travelled (1/2)

Layer	Identifier	What it holds
1 **Protocol source**	**PRT 9.2.3.2**	Secondary efficacy endpoint — “Blood pressure response rate at Weeks 4 and 8 (proportion of subjects with an MSSBP decrease of ≥20 mmHg and an MSDBP decrease of ≥10 mmHg)”
2 **SDR · USDM**	**USDM Endpoint_2**	EP-SECONDARY-EFFICACY · level C85827 (Secondary Objective) · holds the original text without truncation
3 **SDR · Mapping**	**ANA_SEC_RESP**	study_map.yaml — PARAMCD RESP · time points Week 4 / Week 8 · population FAS · test chi-square or Fisher
4 **Deliverable · SAP**	**SAP 10.3**	Secondary efficacy endpoint analysis — traceability.csv records “came from PRT 9.2.3.2”

Final Traceability — How Far One Item Travelled (2/2)

Layer	Identifier	What it holds
5 MDR	CRITy / CRITyFL	adam_structures.yaml — defines only the structure of the determination flag; it holds no threshold values such as 20 mmHg
6 SDTM	VS	VSTESTCD SYSBP / DIABP · VSPOS SITTING · VSSTRESN — collects the EDC-derived mean as is
7 ADaM	ADEFF.CRIT1 / CRIT1FL	CRIT1 = criterion sentence, CRIT1FL = Y/N. Determined across two parameters; missing after LOCF is set to N
8 Final deliverable	C&R TFL 14.2.2.2	Response rate table — footnote [1] is identical to the ADEFF.CRIT1 text. The source chain extends even to the numbers in the table

Conclusion for today From a single definition in the Protocol to a number in the final Table, where it came from can be traced at every step.

SECTION 04

How Far We Have Come

Implementation scope and remaining work

- USDM implementation scope and non-standard extensions
- What is needed to fill the SDR further
- When an amendment occurs

Filling the SDR Further

Priority	Item to enhance	What moves over	Effect
1	**Visits · schedule**	visits · analysis_visits	The basis for ADaM analysis visits (AVISIT) becomes standard
2	**Treatment arms**	arms	Secures the basis for SDTM DM.ARM. The design structure becomes standardized
3	**Estimand**	criteria · missing_data · statistics	Aligns with ICH E9(R1). All the ingredients already exist; they only need to be bundled
4	**Biomedical Concept**	Part of our own MDR	Restores the standard path from USDM to SDTM
5	**Standard extension path**	4 non-standard fields	Move them to ExtensionClass to secure conformance

A measurable indicator 7 of the 13 blocks in study_map.yaml properly belong to the SDR. The size of this file is exactly how much standardization is still missing.

Amendment — Adding a Week 12 Assessment

Amendment Add a Week 12 efficacy assessment visit, and additionally assess MSSBP · MSDBP and the blood pressure response rate at Week 12. The statistical methods stay the same.

Impact assessment by analysis unit

Analysis unit	Before	After	Verdict
ANA_PRIMARY · Week 8 MSSBP	Week 8	Week 8	No impact
ANA_SEC_MSSBP_W4 · MSSBP	Week 4	Week 4, 12	**Review needed**
ANA_SEC_MSDBP · MSDBP	Week 4, 8	Week 4, 8, 12	**Review needed**
ANA_SEC_RESP · response rate	Week 4, 8	Week 4, 8, 12	**Review needed**
Safety analysis · adverse events	Entire period	Entire period	No impact

What Stays Unchanged

MDR — all derivation · statistical rules
CHG · ANCOVA · CRITy / CRITyFL · chi-square / Fisher

TFL Shell — table skeleton
The time-point block repeats, so Week 12 follows the same pattern

SDTM — DM · AE · LB

ADaM — ADSL · ADAE

Amendment — Downstream Outputs to Review

Downstream outputs that need review

SDTM · VS

Baseline · Week 4 · Week 8
→ Add Week 12 records

ADaM · ADEFF

AVISIT 3 → 4
Derivation rules apply as is

SAP 10.3

3 secondary efficacy analysis items
Flagged REVIEW REQUIRED

TFL 14.2.1.1 / 2.1 / 2.2

Add a Week 12 time-point block
Reuse the shell pattern

Message The amendment is not reflected automatically. The changed information is structured once, and the linked deliverables that need review are pointed out.

Maturity stages

Impact Detection | **Affected Draft Update** | **Clinical / Statistical Review** | **Approved Update**

Today’s demo scope: up to change propagation, before the first stage

Closing

- Less time spent re-typing
- Every value keeps a record of where it came from
- When something changes, you know where to look

“

A method, not a tool

Thank You

Eunhye Lee · Director, C&R Research

2026 Korea CDISC Day · 29 October 2026

