

cdisc®

2026 US Interchange

THE FUTURE IS CONNECTED: STANDARDS AND AI POWERING DIGITAL TRANSFORMATION



SUPP to NSV: Transforming Data Representation for Improved Reviewer Utility

Soumya Rajesh, Sr. Standards Engineer, Global Data Standards, IQVIA

Speaker



Soumya Rajesh,
Sr. Standards Engineer,
Global Data Standards,
IQVIA

Soumya has over 21 years of experience in CDISC standards, programming, and regulatory operations across diverse therapeutic areas and study phases. A published author and frequent speaker since 2018, she has presented on SDTM & ADaM best practices, Clinical Classifications, Findings About and submission standards. She is also an authorized CDISC instructor (SDTM and CDASH) and has held key roles at standards development teams including Past Lead of the CDISC SDS Leadership Team and Co-Lead of the SDS NSV sub-team. Soumya is an active contributor to CDISC, PHUSE, and PharmaSUG initiatives. Today, she brings this wealth of expertise to share insights on the evolving data standards landscape.

Introduction:

Supplemental Qualifier (SUPPQUAL) datasets have served as the long-standing mechanism for representing non-standard data in SDTM. Using QNAM (variable name), QLABEL (label), and QVAL (value), SUPP datasets store non-standard variables in a vertical structure that requires transposition or pivoting to merge with parent domains. Although this design initially offered flexibility, limitations have accumulated as data complexity and regulatory expectations have evolved.

So, the shift from SUPPQUAL to Non-Standard Variables (NSV) in SDTM v3.0 and SDTMIG v4.0 is transformative. Rather than storing one row per non-standard variable, NS-- datasets store one row per parent observation, with each non-standard variable expressed as a true variable (column). This enables consistent metadata, improved readability, simpler merging, and more efficient review. It also resolves long standing pain points such as QVAL always being character, difficulty applying controlled terminology, and inconsistent value level metadata.

This paper presents a consolidated overview of how NSVs function, why they improve the SDTM data model, and how implementers can incorporate them successfully.

Background: Limitations Of Suppqual:

Background: Limitations Of Suppqual:

Structural Issues:

- One record per QNAM/QVAL pair results in redundancy and large datasets.
- Transposing is required to merge SUPPQUAL back to parent domains.
- Merging logic varies depending on IDVAR/IDVARVAL combinations, and even then it requires conversion of IDVARVAL (due to it being character) to merge w/ a numeric --SEQ.
- QVAL is always character—even when storing numeric values.



Background: Limitations Of Suppqual:

Metadata Limitations:

- SUPPQUAL relies on value-level metadata in Define-XML.
- Variable-level attributes such as length, type, codelist, and origin cannot be uniquely assigned.

Reviewer Burden:

- Reviewers must manually reassemble data to understand all variables associated with the parent domain.
- Data with many supplemental qualifiers becomes hard to interpret and increases review time.



Background: Limitations Of Suppqual:

Inconsistent Implementation:

- SUPPQUAL allowed different join keys (IDVAR = --SEQ, --TESTCD, --GRPID, etc.), which forced programmers to create multiple merge conditions.
- Validation inconsistencies were common due to ambiguous linkages.



Rationale for transition to NSVs

Rationale for transition to NSVs:

- Horizontal alignment enhances readability and interpretability.
- Variable-level metadata improves clarity and supports automation.
- Consistent merge logic based on --SEQ simplifies programming.
- Can also reduce the number of records needed, as the structure matches parent dataset structure (versus vertical SUPP).
- Regulatory agencies expressed a preference for simpler, more transparent data structures.
- Machine-readable metadata initiatives in SDTMIG v4.0 require a clearer metadata model that SUPPQUAL cannot support.

cdisc

Supp to NSV Transformation

- A significant structural overhaul in SDTM history.
- Vertical vs. Horizontal - Structural Differences: SUPPQUAL stored each variable in its own row linked back to the parent record, but saved values as text thus losing numeric types, and required transposing and complex joins to merge back.
- In contrast, NS-- uses a horizontal structure where each record contains all applicable non-standard variables for a parent record, and records without any values can be dropped.
- This simplifies merging, enhances readability, and supports variable-level metadata in Define-XML.

cdisc

Benefits of the NSV Structure

- Cleaner integration with parent domains using standard keys.
- Enhanced metadata handling, including variable type, length, codelists, and origin.
- Improved usability, allowing reviewers to view standard and non-standard variables side by side.
- Reduced dataset size, as multiple NSVs share a single parent-aligned record.
- Better consistency, since join logic uses a single key variable rather than multiple approaches

cdisc

The NSV Metadata Specification Table

#	Variable Name	Variable Label	Type	Codelist	Allowed Controlled Terms	Format	Role	Variable Group	Root Variable C-code	Root Variable Definition	Notes	Examples	Core
1	STUDYID	Study Identifier	Char				Identifier	Study	C83082	A sequence of characters used to identify or name the study.			Req
2	RDOMAIN	Related Domain Abbreviation	Char	(DOMAIN)			Identifier	Pointer	C83391	A variable that represents the domain abbreviation for the domain in which related records are located.			Req
3	USUBJID	Unique Subject Identifier	Char				Identifier	Observation Subject	C69256	A sequence of characters used to uniquely identify a subject across all studies for all applications or submissions involving the product.	See Use of "Subject" and USUBJID and SUBJID.		Exp
4	APID	Associated Persons Identifier	Char				Identifier	Observation Subject		A sequence of characters used to uniquely identify a single associated person, a group of associated persons, or a pool of associated persons.			Perm
5	POOLID	Pool Identifier	Char				Identifier	Observation Subject	C117053	A sequence of characters used to uniquely identify a group of subjects that have been pooled together.			Perm
6	SPDEVID	Sponsor Device Identifier	Char				Identifier	Observation Subject	C117060	A sequence of characters used by the sponsor to uniquely identify a specific device.	See Performers and Evaluators.		Perm
7	IDVAR	Identifying Variable	Char				Identifier	Pointer		The name of a variable, used in combination with other variables, to identify related records.	See NS Assumption 4.	"--SEQ"	Exp
8	IDVARVLN	Identifying Variable Numeric Value	Num				Identifier	Pointer		The numeric value of the variable named in IDVAR used to identify related records.	See NS Assumption 4.		Exp

Implementation rules

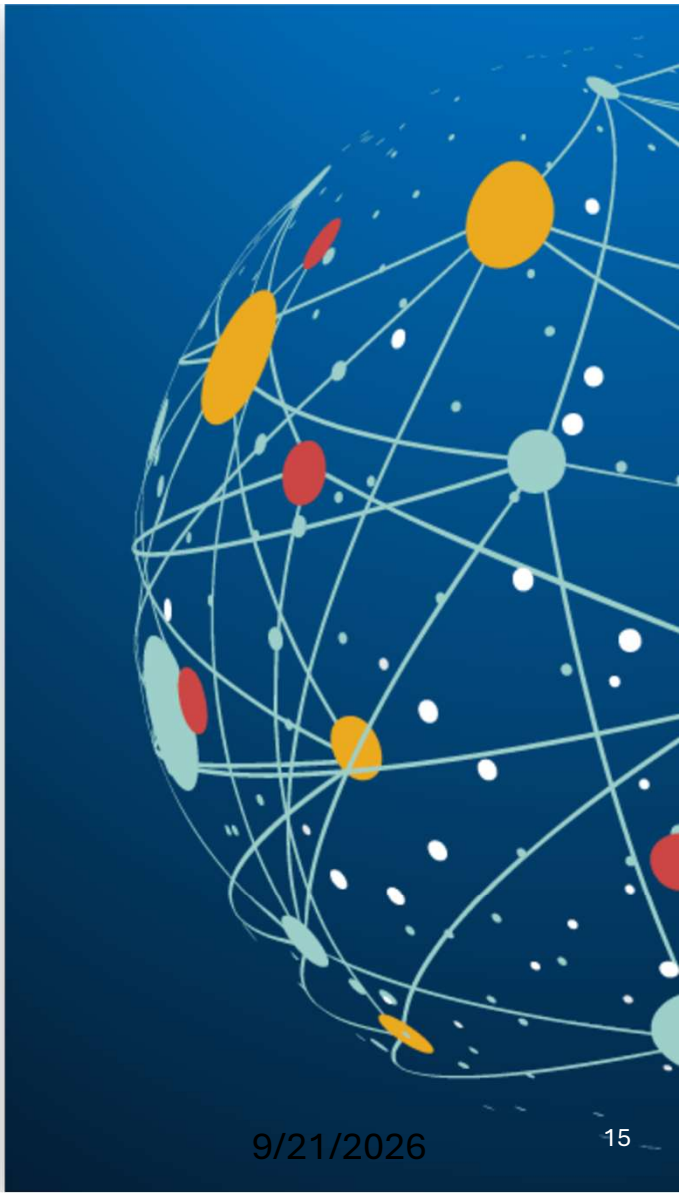
- NS-- datasets must include at least one NS variable populated for each record.
- Each NS-- record must relate back to a single, unique parent domain record.
- An appropriate identifier must be included in each NS-- record based on the type of entity represented by the parent record.
- IDVAR must equal –SEQ (except for NSDM), since --SEQ is the universal sequence variable used to uniquely identify each parent observation.
- Every SDTM dataset may have one corresponding NS– dataset, except when a domain is split as in FAMH the NS dataset mirrors the split: NSFAMH.
- NSVs must use the naming conventions outlined in SDTMIG Section 4.2.3.

Implementation rules

- NSVs follow the same rules regarding text strings exceeding 200 characters as described in section 4.3.6.
- When there are multiple values for qualifiers, rules from Section 4.3.5.1, “Multiple Values for a Non-result Qualifier Variable” would be used.
- CT provided for NSVs should be followed where applicable.
- All NSVs must be fully represented in Define-XML using variable-level metadata.
- An NSV must have 1 of the roles: NS Identifier / NS Timing / NS Qualifier / NS Rule.
- Certain data types should not be represented as NSVs, even if non-standard.

Assumptions:

- **Assumption 1** - One NS-- record per Parent record.
- **Assumption 2** - Only non-null NSVs are represented.
- **Assumption 3** - NSV datasets follow the same structural ordering as Parent datasets.
- **Assumption 4** - Variables must conform to SDTM typing rules.
- **Assumption 5** - Consistency across split domains.
- **Assumption 6** - Metadata must always reflect the actual content.



Examples:

Example 1: SUPPAE vs NSAE

ae.xpt Convert IDVARVAL from character to numeric

Row	STUDYID	DOMAIN	USUBJID	AESEQ	AETERM	AESEV	AESER	AESMIE	AESTDTC	AEENDTC
1	1996001	AE	99-401	1	UTERINE FIBROIDS	SEVERE	Y	Y	2023-01-05	2023-01-12
2	1996001	AE	99-567	1	FEVER	MILD	N		2023-09-25	2023-09-25

Transpose

suppae.xpt

Row	STUDYID	RDOMAIN	USUBJID	IDVAR	DVARVAL	QNAM	QLABEL	QVAL	QORIG	QEVAL
1	1996001	AE	99-401	AESEQ	1	AESOSP	Other Medically Important SAE	SPONTANEOUS ABORTION	CRF	
2	1996001	AE	99-401	AESEQ	1	AETRTEM	Treatment Emergent Flag	Y	Derived	
3	1996001	AE	99-567	AESEQ	1	AETRTEM	Treatment Emergent Flag	N	Derived	

nsae.xpt

Row	STUDYID	DOMAIN	USUBJID	IDVAR	IDVARVLN	AESOSP	AETRTEM
1	1996001	AE	99-401	AESEQ	1	SPONTANEOUS ABORTION	Y
2	1996001	AE	99-567	AESEQ	1		N

NSAE metadata

Variable	Label	Type	Codelist	Role	Origin Type	Origin Source
AESOSP	Other Medically Important SAE	Char		Non-Standard Qualifier	Collected	Investigator
AETRTEM	Treatment Emergent Flag	Char	(NY)	Non-Standard Qualifier	Derived	Sponsor

Example 2: SUPPDM vs NSDM

suppdm.xpt

Row	STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG	QEVAL
1	ABC789	DM	ABC789-010-047			RACE2	Race 2	ASIAN	CRF	
2	ABC789	DM	ABC789-010-047			RACE5	Race 5	WHITE	CRF	

SUPPDM represents multiple values in race vertically in different rows,

nsdm.xpt

Row	STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVLN	RACE2	RACE5
1	ABC789	DM	ABC789-010-147			ASIAN	WHITE

NSDM would represent this in one row as only one record in DM is being qualified with NSVs.

NSDM Metadata:

Variable	Label	Type	Codelist	Role
RACE2	Race 2	Char	(RACE)	Non-Standard Record Qualifier
RACE5	Race 5	Char	(RACE)	Non-Standard Record Qualifier

Example 3: NSSV

sv.xpt

Row	STUDYID	DOMAIN	USUBJID	SVSEQ	VISITNUM	VISIT	SVPRESP	SVOCCUR	SVREASOC
1	1996001	SV	99-401	1	1	SCREENING	Y	Y	
2	1996001	SV	99-401	2	2	CYCLE 1 DAY 1	Y	Y	
3	1996001	SV	99-401	3	3	CYCLE 1 DAY 2	Y	N	FIRST 200 CHARACTERS OF REASON FOR OCCUR VALUE...

NSSV dataset adds qualifying information to Subject Visits data where RDOMAIN = "SV".

nssv.xpt

Row	STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVLN	SVREASO1
1	1996001	SV	99-401	SVSEQ	3	SECOND 200 CHARACTERS OF REASON FOR OCCUR VALUE

NSSV Variable Metadata [SVREASO1]

Variable	Label	Type	Codelist	Role
SVREASO1	Reason for Occur Value	Char		Non-Standard Qualifier

Example 4: NSHO

ho.xpt

Row	STUDYID	DOMAIN	USUBJID	HOSEQ	HOTERM	HOSTDTC	HOENDTC	HODUR
1	1999001	HO	0001	1	HOSPITAL STAY	2004-01-05	2004-01-12	P1W
2	1999001	HO	0001	2	HOSPITAL STAY	2004-01-23	2004-02-07	P15D
3	1999001	HO	0002	1	HOSPITAL STAY	2004-01-21	2004-01-22	P1D

The HO dataset includes a subset of standard variables.

nsho.xpt

Row	STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVLN	HOAERPFL	HOMEDSFL	HOPROCFL	HONAM	HOSPUTY	HOSPUFL	HORLCNDF
1	1999001	HO	0001	HOSEQ	1	Y	Y	Y	GENERAL HOSP	ICU	Y	Y
2	1999001	HO	0001	HOSEQ	2	Y	Y	N	UNIV HOSP	CCU	Y	Y
3	1999001	HO	0002	HOSEQ	1	Y	N	Y	ST. MARY'S	ICU	N	Y

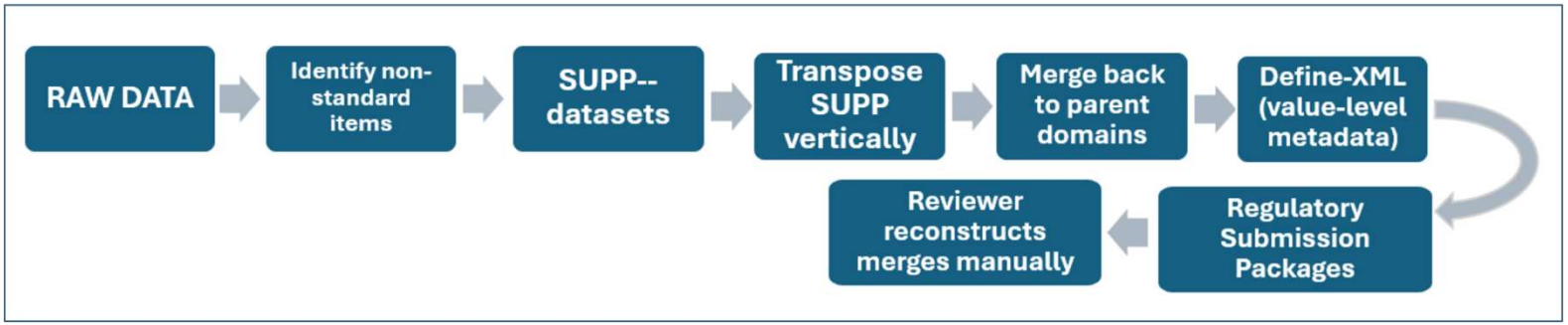
NSHO Metadata:

NSHO Variable Metadata

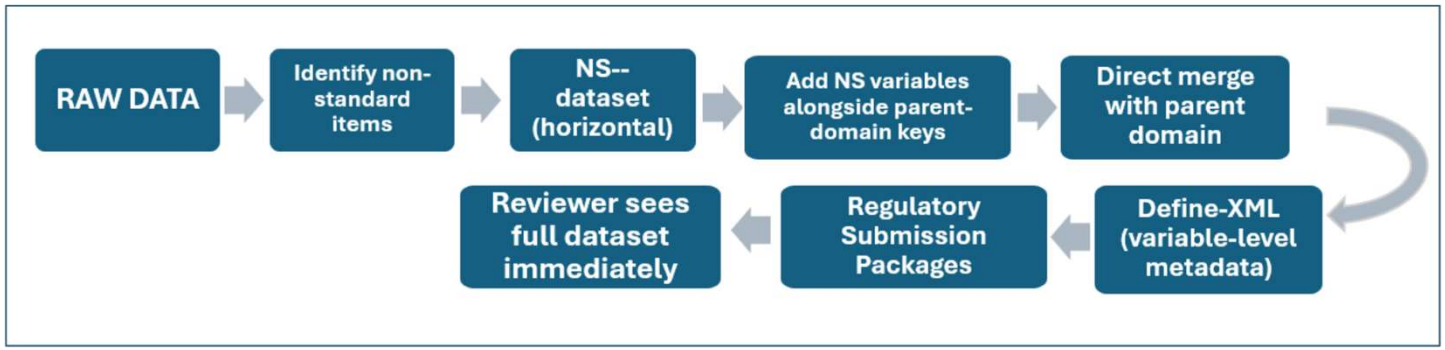
Variable	Label	Type	Codelist	Role
HOAERPFL	AE Reported This Episode	Char	(NY)	Non-Standard Qualifier
HOMEDSFL	Meds Prescribed	Char	(NY)	Non-Standard Qualifier
HOPROCFL	Procedures Performed	Char	(NY)	Non-Standard Qualifier
HONAM	Provider Name	Char		Non-Standard Qualifier
HOSPUTY	Specialized Unit Type	Char		Non-Standard Qualifier
HOSPUFL	Any Time in Spec. Unit	Char	(NY)	Non-Standard Qualifier
HORLCNDF	Visit Related to Study Med Cond.	Char	(NY)	Non-Standard Qualifier

Workflow comparison:

- SuppQual Workflow:



- NS-- Workflow:



Structural Differences:

Feature	SUPPQUAL (SUPP--)	NSV (NS--)
Structure	Vertical (one row per non-standard variable)	Horizontal (one row per parent record)
Metadata	Value-level only	Variable-level (codelists, length, type, origin)
Merge Complexity	Requires transpose + join	Direct join via STUDYID/RDOMAIN/USUBJID/--SEQ
Data Types	QVAL = character always	True numeric/character typing
Review Burden	High (multiple files, reassembly needed)	Low (directly readable)
Alignment	Deprecated in SDTM v3.0	Required in SDTM v3.0

Technical Comparison

- **SUPP-- Merge Logic (Old)**

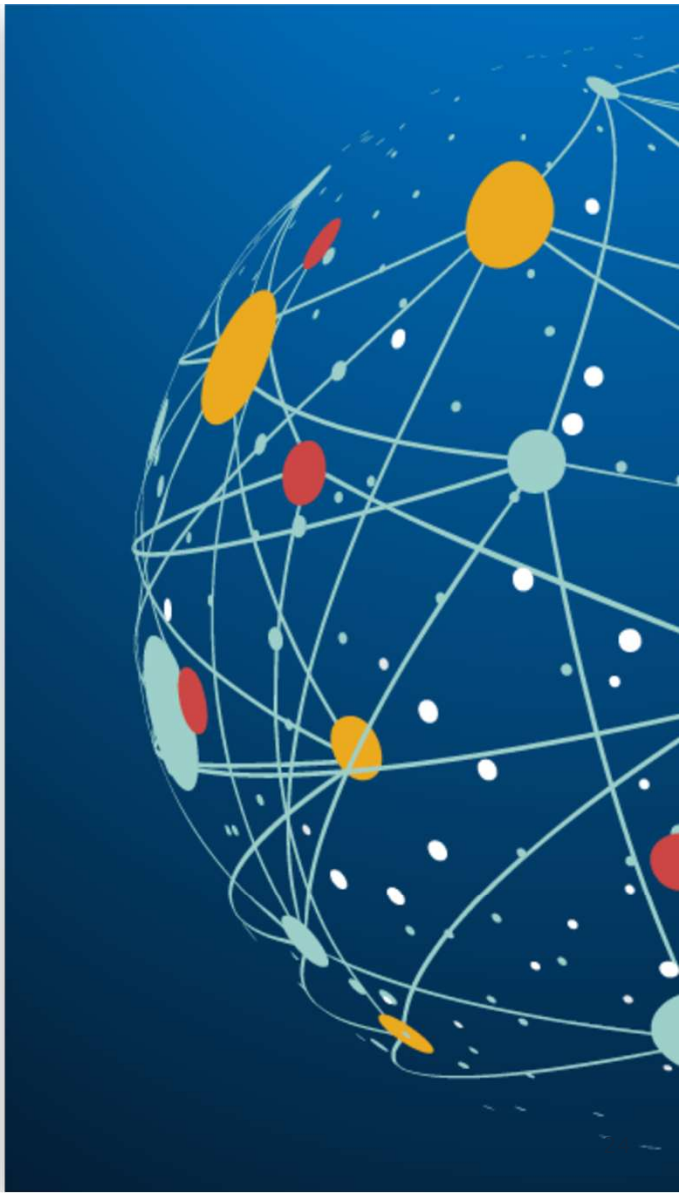
```
proc transpose data=suppae
out=work.ae_supp;
    by usubjid idvar idvarval;
    id qnam;
    var qval;
    run;
data ae_merged;
    merge ae ae_supp;
    by usubjid aeseq;
    run;
```

- **NS-- Merge Logic (New)**

```
data ae_merged;
    merge ae nsae;
    by usubjid aeseq;
    run;
```

Practical considerations for Sponsors:

- Updating macro libraries to support NS-- creation
- Validating all IDVAR and IDVARVLN linkages
- Ensuring Define-XML metadata is complete
- Updating reviewer-friendly outputs (e.g., SDRG notes)
- Training statisticians and reviewers on new structures



Additional clarification / discussion:

- The transition to NSVs significantly simplifies both programming and review.
- The metadata improvements in SDTMIG v4.0 would enhance these benefits.
- Regulatory agency reviewers would benefit from having a complete representation of variables within a single dataset structure.

Conclusion:

- NSVs resolve many long standing SUPPQUAL limitations.
- Benefits of the adoption of NS-- datasets:
  - » Cleaner, more intuitive data structures
  - » Better metadata quality and interpretability
  - » Reduced programming complexity
  - » Improved reviewer efficiency
  - » Alignment with modern regulatory expectations



References:

SDTM v3.0 and SDTMIG v4.0

- SDTMIG v4.0 Home - SDTMIG v4.0 – CDISC Wiki
- SDTM | CDISC
- SDTMIG | CDISC
- Non-Standard Variables Registry | CDISC



Acknowledgements

Special thanks to my esteemed colleague and co-lead of the 'SUPP to NSV' sub-team, **David Neubauer**, Sr. Manager, Clinical Data Assets, Repositories, & Standards, AbbVie, for his support and review of this paper.

Since appreciation for the collaboration on the sub-team, to all the members of "Supp to NSV": Kapila Patel, Michael Wise, and Maria Sekac.





Questions?

Your comments and questions are valued and encouraged.

Contact the author at:

- Name: Soumya Rajesh, Sr. Standards Engineer, Global Data Standards
- Organization: IQVIA
- Email: soumya.rajesh@iqvia.com
- Web: linkedin.com/in/soumya-rajesh-2b7a074

Thank You!

