PhUSE 2012

Paper CD07

SDTM and ADaM: Metadata Playing Together Nicely

Ken Stoltzfus, Octagon Research Solutions, Wayne, USA

ABSTRACT

As the use of the CDISC standards SDTM and ADaM becomes more and more embedded in sponsor organizations across the industry, the opportunities for implementing standards-based, metadata-driven processing grows and grows. Metadata from the CDISC SDTM and ADaM standards can serve as the consistent consumable for processes that are part of the clinical data lifecycle, with the goal being more efficient mechanisms to produce regulatory compliant deliverables.

From my standards experience on both sides of the pharmaceutical industry - pharmaceutical companies and contract research organizations - it is apparent that the industry is still in the early stages on the learning curve of using metadata to drive clinical data processes. But it is my objective here to share that experience and insight into describing how the use of SDTM and ADaM metadata can be productive.

The example included in this presentation will involve the SDTM AE domain and the ADaM ADAE dataset to demonstrate a simple example of Adverse Events metadata working together in an integrated fashion for a more seamless data flow across the multiple processes and functional areas. It does not intend to establish that this is the only structure or method to handle metadata.

INTRODUCTION

As more sponsor organizations follow the trend of bringing the production of SDTM and ADaM in-house rather than with outsourced external partners, they face both challenges and opportunities regarding how to implement them into their existing environment. A key factor in how companies effectively address the challenges and embrace the opportunities is how they use their standards metadata. How integrated is their use of SDTM and ADaM metadata, and what underlying processes and technologies are in place to make this happen?

METADATA FOR CDISC STANDARDS

BACKGROUND

Within many sponsor organizations, the management and usage of standards metadata is still narrowly confined to individual functional areas and/or processes. For example, the data management function might be responsible for production of SDTM datasets and therefore are the only ones overseeing the SDTM metadata. Conversely, the statistical programming function responsible for the production of ADaM analysis datasets and results holds the oversight of the ADaM metadata in a completely separate location and manner.

In a sponsor's early usage of CDISC standards, this approach might facilitate the development of expertise with the individual models of SDTM and ADaM, but it does not help with the overall advantageous utilization of the standards metadata. The desired approach would involve integrated management and employment of the metadata into a single metadata model that prevents the problems and inconsistencies often seen at "handoff points" in the clinical data lifecycle, such as data meanings, visit structures, and baseline derivations.

The metadata will be represented here in tabular format for ease in viewing and comprehension, as opposed to the relational database model that is increasingly being used in the industry to facilitate more robust functionality of a metadata registry (MDR). Each column will be briefly described in advance of the table display. Due to space constraints, both the number of data elements (rows) and metadata attributes (columns) will be limited to a fraction of the overall metadata. It is acknowledged here that the number of attribute columns typically used in the effective storage and usage of metadata is at least double the number displayed here.

This metadata displayed in both Table 1 and Table 2 are representative of the metadata that is both being stored in the MDR and being exported for use across the clinical data lifecycle. It contains a blend of information that is straight from CDISC and from sponsor-generated sources. Each sponsor must determine the process for their own

environment by which CDISC and sponsor metadata is integrated within the MDR, balancing the import capabilities of the tool and the labor-intensive, hands-on metadata management tasks.

SDTM METADATA

SDTM AE Domain – Adverse Events

- Variable Name: CDISC Variable Name
- Variable Label: CDISC Variable Label
- Type: CDISC character or numeric
- Length: Sponsor-defined length of variable
- Controlled Terminology: Terminology that is managed either by CDISC or the sponsor
- Format: Sponsor-defined display formats
- CDISC: Name of the CDISC model for this data element
- IG Version: Version of the Implementation Guide for the CDISC model for this data element
- Concept Group: Sponsor-defined grouping of medical concepts
- Concept ID: Sponsor-defined individual medical concepts

Table 1

Variable Name	Variable Label	Type	Length	Controlled Terminology	Format	CDISC	IG Version	Concept Group	Concept ID
USUBJID	Unique Subject Identifier	Char	30		$30	SDTM	3.1.2	Global	Unique Subject Identifier within Submission
AESEQ	Sequence Number	Num	8		3.0	SDTM	3.1.2	Global	Subject Sequence
AETERM	Reported Term for the Adverse Event	Char	200		$200	SDTM	3.1.2	Adverse Event	Verbatim Adverse Event
AEDECOD	Dictionary-Derived Term	Char	200		$200	SDTM	3.1.2	Dictionary	Dictionary Preferred Term
AECAT	Category for Adverse Event	Char	100		$100	SDTM	3.1.2	Adverse Event	Category
AEBODSYS	Body System or Organ Class	Char	200		$200	SDTM	3.1.2	Dictionary	Dictionary Body System
AESEV	Severity / Intensity	Char	20	AESEV	$20	SDTM	3.1.2	Adverse Event	Severity
AESER	Serious Event	Char	1	NY	$1	SDTM	3.1.2	Adverse Event	Serious
AEACN	Action Taken with Study Treatment	Char	30	ACN	$30	SDTM	3.1.2	Adverse Event	Action
AEREL	Causality	Char	20	REL	$20	SDTM	3.1.2	Adverse Event	Causal Relationship
AEOUT	Outcome of Adverse Event	Char	40	OUT	$40	SDTM	3.1.2	Adverse Event	Outcome
AESTDTC	Start Date/Time of Adverse Event	Char	30	ISO 8601	$30	SDTM	3.1.2	Adverse Event	AE Onset
AEENDTC	End Date/Time of Adverse Event	Char	30	ISO 8601	$30	SDTM	3.1.2	Adverse Event	AE Resolved

In Table 1, the key messages include:

- The separation of controlled terminology from display formatting into 2 distinct metadata attribute columns allows a single metadata model to support both SDTM and ADaM metadata. Also, this helps to address the common misunderstanding that display formats and controlled terminology are one and the same.
- The explicit inclusion of the CDISC model and its version as metadata attribute columns supports the import functionality of the MDR in how it might process the different SDTM and ADaM metadata in a single pathway, but also allows for both SDTM and ADaM metadata to coexist together within the MDR if the sponsor so chooses.
- Both the Concept Group and Concept ID attribute columns introduce the idea of a concept-based model, which is essentially a model driven by medical concepts that increase clarity of meaning instead of by use cases such as SDTM. A concept-based model includes data definitions and relationships into individual concepts that have multiple related data elements (e.g., blood pressure concept: result, units, body position, method, etc.). While the metadata model for a concept-based MDR would likely not be stored in the manner displayed in Table 1, the table does serve as a “bridge” of sorts between a CDISC domain-based model and a concept-based model.

ADAM METADATA

ADaM ADAE Dataset – Adverse Events Analysis Dataset

- Variable Name: CDISC Variable Name
- Variable Label: CDISC Variable Label
- Type: CDISC character or numeric
- Length: Sponsor-defined length of variable
- Controlled Terminology: Terminology that is managed either by CDISC or the sponsor
- Source: Variables that are used in the derivation of the variable defined in that row
- Derivation: The detailed information that explains how the source variable(s) were used to compute the value of the variable defined in that row

Variable Name	Variable Label	Type	Length	Controlled Terminology	Format	Source	Derivation
USUBJID	Unique Subject Identifier	Char	30		$30	AE.USUBJID	
AESEQ	Sequence Number	Num	8		3.0	AE.AESEQ	
AETERM	Reported Term for the Adverse Event	Char	200		$200	AE.AETERM	
AEDECOD	Dictionary-Derived Term	Char	200		$200	AE.AEDECOD	
TRTEMFL	Treatment Emergent Analysis Flag	Char	1	Y	$1	ADSL.TRTSDT, ASTDT, ADSL.TRTEDT	If ADSL.TRTSDT <= ASTDT<=(ADSL.TRTEDT +21) then TRTEMFL='Y'
AESTDTC	Start Date/Time of Adverse Event	Char	30	ISO 8601	$30	AE.STDTC	
ASTDT	Analysis Start Date	Num	8		yymmdd10.	AE.STDTC	Macro [DTCN,AE.AESTDTC]
AESEV	Severity /Intensity	Char	20	AESEV	$20	AE.AESEV	
AESEVN	Analysis Severity /Intensity (N)	Num	1	AESEVN	1.0	AE.AESEV	Macro [CODEN, AE. AESEV, AESEVN]
AOCCPFL	1st Occurrence of Preferred Term Flag	Char	1	Y	$1	TRTEMFL USUBJID AEDECOD ASTDT AESEQ	(TRTEMFL='Y') Sort Order (USUBJID), (AEDECOD), (ASTDT) If FIRST.AEDECOD, then (AOCCPFL ='Y')
TRTA	Actual Treatment	Char	20	TREAT	$20	ADSL.TRT01A	
TRTAN	Actual Treatment (N)	Num	8	TREATN	1.0	ADSL.TRT01A	
TRTSDT	Date of First Exposure to Treatment	Num	8		yymmdd10.	ADSL.TRTSDT	
TRTEDT	Date of Last Exposure to Treatment	Num	8		yymmdd10.	ADSL.TRTEDT	
AGEGR1	Pooled Age Group 1	Char	20	AGEG1	$20	ADSL. AGEGR1	
SEX	Sex	Char	2	SEX	$2	DM.SEX	
RACE	Race	Char	40	RACE	$40	DM.RACE	

In Table 2, the key messages include:

- The ADAE ADaM dataset does not subscribe to ADaM's Basic Data Structure that is centered around Parameter (PARAM) and Analysis Value (AVAL). This allows for a more focused purpose of performing an analysis of counting subjects with a dictionary term. This analysis is more appropriately executed using an SDTM structure with added analysis variables.
- The Source and Derivation attribute columns for ADaM metadata helps to facilitate downstream processes involving analysis datasets by providing metadata that can both drive the programs that generate analysis data and also aid in the automated validation of analysis data produced externally.
- Within the Derivations attribute column, sponsors must define the rules and conventions of the content contained within that attribute. This metadata content can range from instructions to macro calls to explicit machine-executable code. While many sponsors currently generate ADaM metadata with pseudo-code as the provisional content, there is increasing traction across the industry for the derivations content to be machine-readable and executable to maximize the automation of the analysis and reporting processes.

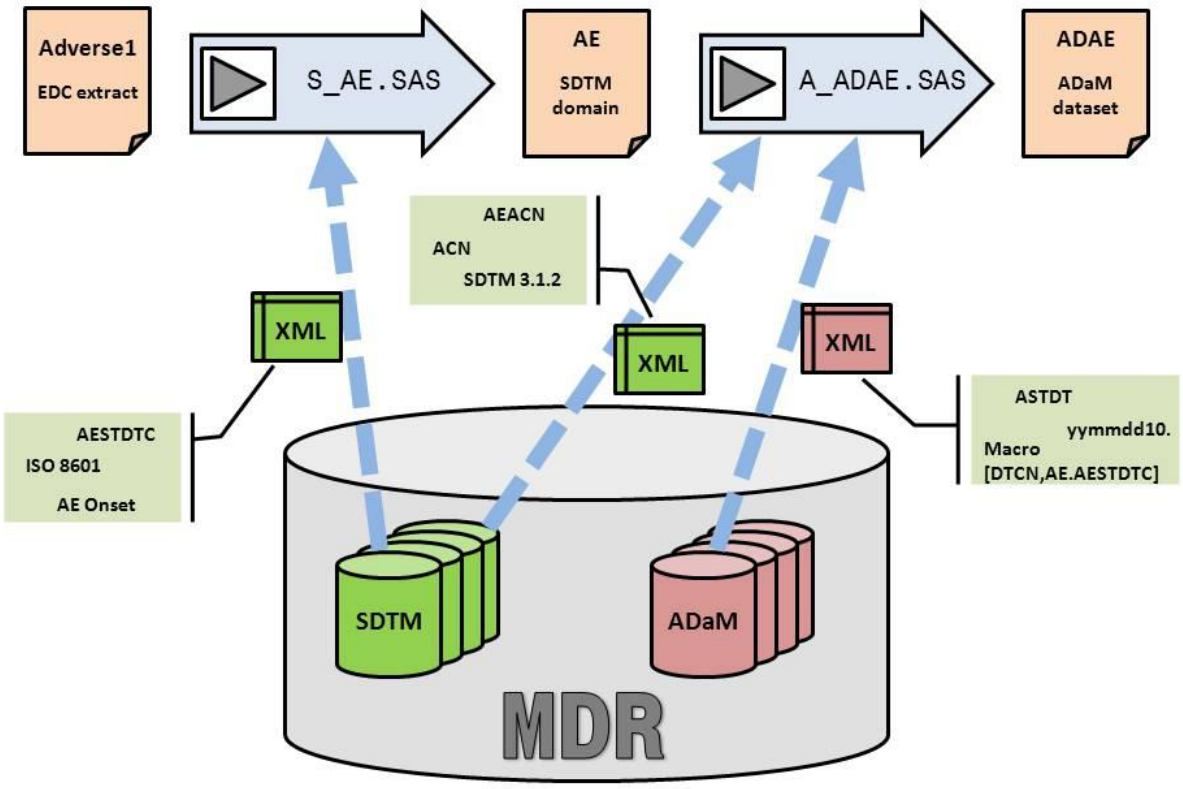
USAGE OF METADATA

BACKGROUND

The use of metadata within clinical processes has been steadily growing across the industry as sponsors try to become more efficient and consistent in executing clinical trial data flow. The efficiencies are realized through the consumption of metadata that can automate processes that were previously more labor-intensive, while the consistency is gained in having the processes from across the clinical data lifecycle execute upon centrally-managed metadata.

Diagram 1 below depicts the flow of metadata from an MDR to support the generation of SDTM and ADaM datasets.

Diagram 1



MAINTENANCE OF METADATA

BACKGROUND

Across the industry there exists a broad range of practices on how sponsor organizations are managing their metadata. It is still very common for metadata to be stored and managed in what has been labeled as a "first-generation metadata repository": Excel spreadsheets. While this tool does provide a favorable comfort level to clinical staff, it is limited with regards to the operational maintenance, elemental-level versioning, and effective utilization. Sponsors on the leading edge in the industry are beginning to implement technology solutions that allow their standards governance organizations to consolidate their metadata in a central location, to efficiently manage it according to company guidelines, and finally to deliver it for use by their clinical systems and processes to consistently produce regulatory-compliant data.

METADATA RESIDING IN A METADATA REGISTRY

Once a sponsor's metadata has been centrally located within an MDR, several things suddenly become possible:

- Storage in a single location that facilitates consistent usage of the same metadata for multiple uses
- Linkages between metadata that can be inherited within the MDR regardless of taxonomy or structure
- Application of medical concepts to metadata that bind together relevant data elements within a single medical meaning that ensures clarity of data meaning across the clinical data lifecycle
- Centralized standards governance and maintenance of metadata according to SOP-compliant workflows and approvals
- Export of centrally-managed, consistent metadata to multiple systems and processes across the clinical data lifecycle

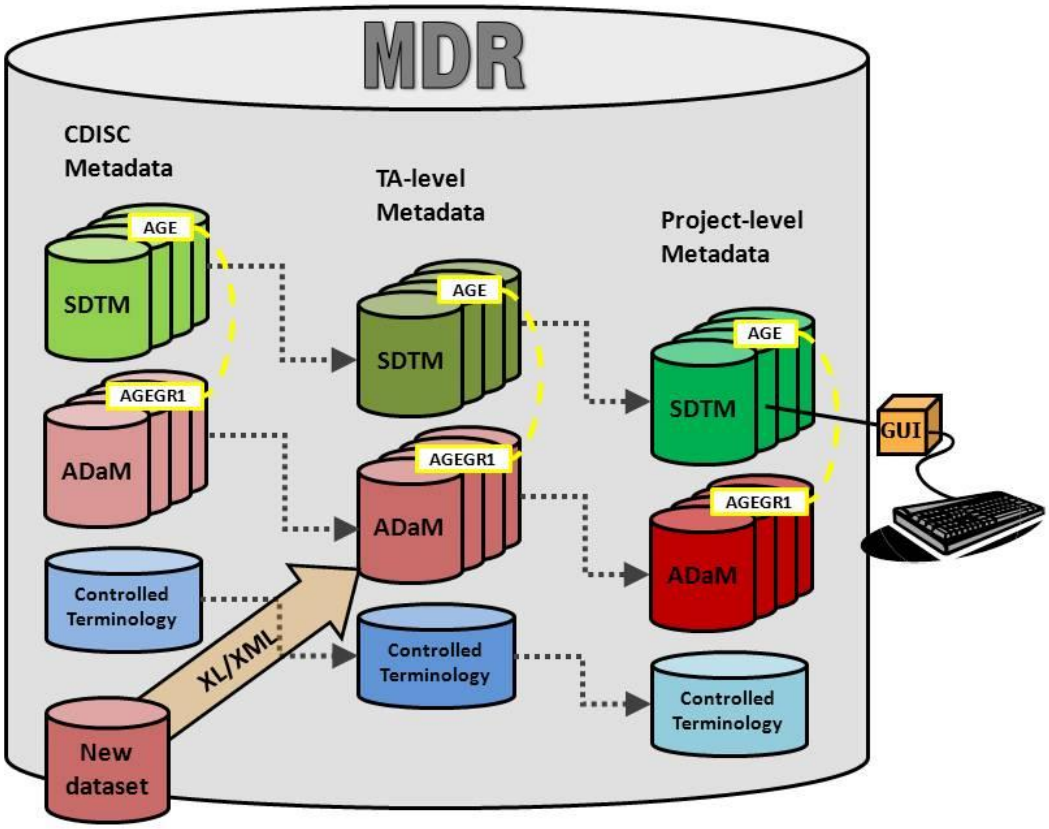
Diagram 2 below depicts an example of metadata storage in an MDR within a CDISC domain-based model that consists of multiple levels of governance at the CDISC, TA, and Project levels. The 3 levels shown here are each based on the relevant CDISC model, with allowable variations within the model and additional sponsor-specific domains/datasets.

Within each level are distinct areas for SDTM and ADaM, along with the Controlled Terminology that supports them. There is direct inheritance from the CDISC level downward through the TA and Project levels, including the linkages in the metadata initially provided by the MDR at the CDISC level (e.g.: AGE in SDTM linked with AGEGR1 in ADaM).

Maintenance of the metadata in the MDR scenario in Diagram 2 is performed by standards governance via:

- Hands-on modifications approved by governance and entered through a graphical user interface (GUI).
- Executable automated uploads of entire metadata domains in either XL or XML that are new to the MDR through an import utility.

Diagram 2



CONCLUSION

Metadata for both of these CDISC standards, SDTM and ADaM, can together be employed appropriately to provide efficient and consistent processing. If the metadata model has the flexibility to accommodate multiple standards domain models and the MDR tool has the functionality to support metadata maintenance and metadata export, the systems and processes of the clinical data lifecycle can be successfully optimized in their production of regulatory compliant deliverables.

ACKNOWLEDGMENTS

I wish to acknowledge my colleagues at Octagon Research Solutions, David Evans, Barry Cohen, and Abi Velurethu, alongside whom I have provided services for clients in the areas of CDISC standards implementation and metadata management.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Ken Stoltzfus
Octagon Research Solutions
585 East Swedesford Road
Wayne, PA 19087 USA
Work Phone: 1-610-535-6500 (Ext. 5891)
Fax: 1-610-535-6515
Email: kstoltzfus@octagonresearch.com
Web: http://www.octagonresearch.com