PharmaSUG 2018 - Paper SS-12

Improving Metadata Compliance and Assessing Quality Metrics with a Standards Library

Veena Nataraj, Erica Davis, Shire

ABSTRACT

Establishing internal Data Standards helps companies ensure a consistent interpretation of CDISC standards in submissions to health authorities around the globe. This presentation discusses our experience, various methodologies and solutions taken to improve the quality of SDTM metadata standards.

INTRODUCTION

Shire has a Standards Governance framework that includes a Clinical Standards Board and Standards Governance Team for each of the CDISC Foundation standards, starting from data collection through to submission and Table, Figure and Listings (TFLs). The SDTM Governance Team is responsible for interpreting the CDISC SDTM concepts and representing the standards metadata as specifications. Shire SDTM Toolkit is the name of the SDTM Standards Library created within Shire. This SDTM Toolkit is part of the end-to-end implementation of standards. It maps the collected information and is the source for ADaM datasets.

As part of the internal Standards Development cycle, a vendor-neutral SDTM Toolkit is updated and published each year. The SDTM Toolkit consists of metadata specifications, business rules and best practices for study standards implementations. It is provided to multiple strategic CRO Partners, who implement these standards' within their environment.

Quality checks for metadata consistency are performed at different points during the development of SDTM Toolkit.

- Prior to providing it to the strategic CROs Partners so to ensure consistency and conformance with the CDISC SDTM Implementation Guide (SDTMIG), Therapeutic Area User Guide (TAUG), Shire internal business needs and other related documents.
- After the strategic CRO Partners implements the SDTM specifications The resulting SDTM Toolkit deliverables are checked to ensure conformance to the Shire Standard specifications.

When Shire receives study metadata from a strategic CRO Partner(s) as part of the SDTM delivery, the metadata will be checked to summarize quality metrics. This helps Shire to oversee how studies use these defined standards and ensures a consistent interpretation of Shire Standards across studies.

SDTM TOOLKIT

"Design is not just what it looks like and feels like. Design is how it works." -Steve Jobs

When developing data standards it is important to think about the definitions rather than the tools that will be used to store the standards. To support Shire's vendor-neutral SDTM standards, three Microsoft® Excel® utilities were developed, which along with other materials make up the Shire SDTM Toolkit:

- The Shire Study Start tool
- The Shire Trial Design tool
- The Shire SDTM Interpretation Metadata

THE SHIRE STUDY START TOOL:

This tool acts a specification, allowing Shire study team members to provide information about the study to the SDTM implementers, whether they be internal or at a CRO. It contains such details, as the version of the SDTMIG, controlled terminology and medical coding dictionaries. It also identifies the validation ruleset version and the supplemental/custom domains. There are two sections in this document:

Study Level information:

The study team completes this section to specify the study level information that can be communicated to the CRO or internal implementer. This includes identifying the standard versions, validation rulesets and external dictionaries.

Study-level Parameters			
Parameter	**Value**	**Example**	**Notes**
Standard Name	SDTM-IG	SDTM-IG	
Standard Version	3.2	3.2	
Supplemental Standards		Device IG 1.1; Cardiovascular TA 1.0	Include any extensions to version
Define Version		1.0.0	
CDISC CT Version		2015-09-25	Based on publication date as identified by CDISC
Validation Ruleset		P21Community2.2.0, Ruleset as of 01-MAR-2106	Include both the software version and ruleset publication date as appropriate.
MedDRA Dictionary Version		8.0	
WHO Drug Dictionary Version		200204	

Fig 1: Study Level parameters

Parameter Level information:

This section captures the study conventions to be used. For example, the Trial Arm Names convention allows a study team to specify how the CRO or internal implementer should establish trial arm values for the study. This information can be obtained from the study protocol. The Notes column provides guidance on how to complete the cell. The Value column is filled out by the study team to identify the information sent to the CRO or internal implementer.

Study Conventions			
Convention	**Value**	**Example**	**Notes**
Trial Arm Names		Alphabetic	Specify the Value column either as categorical terms ("Alphabetic" or "Numeric"), or list the arm values to be proposed for the study.
Visit Names		Initial Caps	Specify the naming convention, such as "Initial Caps", "ALL CAPS", "per source data", etc.

Fig 2: Parameter Level parameters

THE SHIRE TRIAL DESIGN TOOL:

This tool allows the study team to specify a study's trial design metadata based on study materials such as the protocol. This information needs to be reviewed critically by all the contributors prior to starting the SDTM generation since the Trial Design components will be used to develop the study's mapping specification. The Trial Design domains under SDTMIG version 3.2 are:

- Trial Summary (TS)
- Trial Disease Assessment (TD)

- Trial Inclusion (TI)
- Trial Element (TE)
- Trial Arms (TA)
- Trial Visits (TV)

THE SHIRE SDTM INTERPRETATION METADATA:

This contains Shire's interpretation of the CDISC SDTM Implementation Guide as defined by Shire's SDTM Governance Team. It does not supersede CDISC's published documentation, but rather helps to clarify and provides guidance for use at Shire.

It is a metadata document that represents both CDISC-standard and Shire-custom SDTM domains. It includes variable attributes and programming notes for the specification of SDTM datasets. Additional attributes, such as standard derivations and comments, are represented in the Define-XML document. This can also serve as the source to represent SDTM and Define-XML for studies.

Two criteria are used in the document. They are:

- Requirements - These are metadata concepts that must be followed to be compliant to Shire SDTM Standards. When not followed a standards waiver must be obtained by the study team.
- Recommendations - These are best practices suggested by Shire's SDTM Governance Team.

Some of the information contained as part of metadata includes information on Library, Domains, Variables, Value Level metadata (variable and SUPPQUALs), which appears as its own tab are color coded with the two criteria.

Library Tab:

This tab contains the information of the name of SDTM Toolkit, versions of SDTM and Define, the Shire controlled terminology version and the validation ruleset.

Attribute	Value	Example	Usage Notes
StudyName		CDISC01	Simple reference for study, e.g., protocol name
StudyDescription		CDISC Test Study	Brief description, can be the description or sub-title of the protocol.
ProtocolName		CDISC01	Can be the title of the protocol.
StandardName	SDTM-IG	SDTM-IG	
StandardVersion	3.2	3.2	
Author		Shire	
ODMVersion	1.3.1	1.3.1	Version 1.3.1 is the base version for define.xml v1. (Check v1 vs v2)
DefineVersion		1.0.0	
CDISC CT Version			Based on publication date as identified by CDISC
Pinnacle 21 Version			Include both the Pinnacle software version, and ruleset publication date as appropriate.
External Dictionaries	See the Dictionaries tab in this document		
Trial Design External Dictionaries	See the Trial Design content specified at the study level		

Fig 3: shows Study Tab with Study Information

Domains Tab:

The Domains Tab lists the domains from SDTMIG version 3.2 and custom domains which have been standardized at Shire. This tab shows the information for each domain, with a description and other attributes such as Shire Usage. Shire Usage indicates if a domain has an established use case in the current implementation. Additional details such as default key variables, comments, derivations and notes to support the CRO or implementer are also provided.

Dataset	Description	Shire Usage	Repeating	Reference Data	Structure	Key Variables	Comment / REFC:Comme	Additional Notes
AE	Adverse Events		Yes	No	One record per adverse event per subject	STUDYID, USUBJID, AEDECOD, AESTDTC	See Reviewer's Guide, Section	
CM	Concomitant Medications		Yes	No	One record per recorded medication occurrence or constant-dosing interval per subject	STUDYID, USUBJID, CMSTDTC, CMENDTC, CMCAT, CMTRT, CMDOSTXT, CMDOSU, CMINDC, CMDOSFRQ	See Reviewer's Guide, Section 3.3, Concomitant Medications	This domain is for use in capturing Prior/Concomitant Medications and Therapies that are appropriate to code with the WHO Drug Dictionary. Therapeutic procedures
CO	Comments	Not Used	Yes	No	One record per comment per subject	STUDYID, USUBJID, COSEQ	See Reviewer's Guide, Section	

Fig 4: Domains Tab with Notes

Variables Tab:

The Variables tab specifies variables for each domain, with attributes including if they are Required, Permissible or Expected. Some of the additional attributes have been standardized to help the CRO or internal implementer, as well as helping the study teams oversee the SDTM and Define-XML development processes. They are:

- Shire Usage item identifies how the variable has to be used within Shire environment. The variable could be "Required", "Not Used", "Expected" or "Conditionally Expected".
- References to indicate Define attributes such as Datatype, Derivation or Comments.
- References to controlled terminology as applicable.
- Additional notes with programming guidance.

Seq. For Or	Dataset	Variable Nam	Variable Label	Data Type	Role	CDISC Notes (for domains) Description (for Genera	Core	Shire Usag	Define datatype	Controlled Term / Codelist / E: External Codeli	Origin	Derivation / REFM: Method Reference	Variable Comment / REFC: Reference	Additional Notes
1	AE	STUDYID	Study Identifier	Char	Identifier	Unique identifier for a study.	Req		text		Protocol			Value from Protocol. Character 10 -- SHPXXXYYY - With no hyphens,
2	AE	DOMAIN	Domain Abbreviation	Char	Identifier	Two-character abbreviation for the domain.	Req		text	(C_DOMAIN_AE)	Assigned			CDISC has established a Domain codelist. Each domain variable should be linked to a subset
3	AE	USUBJID	Unique Subject Identifier	Char	Identifier	Identifier used to uniquely identify a subject across all studies	Req		text		Derived	REFM:USUBJID		
4	AE	AESEQ	Sequence Number	Num	Identifier	Sequence Number given to ensure uniqueness of subject records within a domain. May be any valid	Req		integer		Derived	Sort records in ascending order by the variables identified in the dataset key. Generate AESEQ by starting at		The variables identified in the sort order as determined by the --SEQ derivation must agree with the key variables established for

Fig 5: Variables Tab

Value Level Tab:

This tab has Value Level metadata. Below is an example for the ECG domain (EG).

Identifying Variable	--TEST Value (Variable Labe	Where Clause	Shire Usage	Data Type	Numeric Leng	Numeric Significant Digits	Codelist / E: External Codeli	Origin	REFM: Reference	Comment / REFC: Reference	Additional Notes
EGORRES		EGTESTCD = 'EGHRMN'		integer	3			CRF			
EGORRESU		EGTESTCD = 'EGHRMN'		text				CRF			
EGSTRESU		EGTESTCD = 'EGHRMN'		text				CRF			
EGORRES		EGTESTCD = 'INTP'		text			(C_EGSTRESC)	CRF			
EGORRES		EGTESTCD = 'PRAG'		integer	3			CRF			
EGORRESU		EGTESTCD = 'PRAG'		text				CRF			
EGSTRESU		EGTESTCD = 'PRAG'		text				CRF			
EGORRES		EGTESTCD = 'QRSAG'		integer	3			CRF			
EGORRESU		EGTESTCD = 'QRSAG'		text				CRF			
EGSTRESU		EGTESTCD = 'QRSAG'		text				CRF			
EGORRES		EGTESTCD = 'QTAG'		integer	3			CRF			
EGORRESU		EGTESTCD = 'QTAG'		text				CRF			
EGSTRESU		EGTESTCD = 'QTAG'		text				CRF			
EGORRES		EGTESTCD = 'RRAG'		integer	3			CRF			
EGORRESU		EGTESTCD = 'RRAG'		text				CRF			
EGSTRESU		EGTESTCD = 'RRAG'		text				CRF			

Reference Domain	Qualifier Variable Name	Qualifier Variable Label	Shire Usage	Qualifier Variable Data Type	Qualifier Variable Length	Qualifier Variable Significan	Controlled Term / Codelist / E: External Codelist / Format	Qualifier Variable Origi	Qualifier Variable Evaluator	Variable Derivation / REF: Reference	Qualifier Variable Comment / REF: Reference	Additional Notes
EG	EGAEYN	Abnormality an		text			(C_NY_NY)	CRF				
EG	EGCLSIG	Clinically Significant		text			(C_NY_NY)	CRF				

Fig 6: Details of Value Level Metadata defined

Interpretation Guidance:

In addition to the detailed metadata, Shire has developed guidance on a range of topics to facilitate consistent solutions across studies, therapeutic areas, and CROs. Broadly, the topics covered are:

- CRF Mapping and Annotation Principles
- General SDTM Mapping Conventions
- Handling of Rollover and Extension Studies
- Trial Arm Assignments for Non-Treated Subjects
- EPOCH Representation
- Multiple Subject Participations
- Domain-Specific Considerations (e.g., details specific to AE, LB)
- Laboratory Transfer Specifications
- Define-XML Considerations
- Preparation of the Study Data Reviewer's Guide (cSDRG)
- Pinnacle 21 Usage and Guidance

QUALITY CHECKS

"Quality is the standard of something as measured against other things of a similar kind; the degree of excellence of something." – Oxford Dictionary

The quality of the SDTM specifications can be improved by ensuring that the metadata represented in the Shire SDTM Metadata Specification is consistent and aligns with other CDISC standards, such as CDASH and ADaM. The additional notes and guidances within the metadata specification and other documents provide details on how a CRO implements the SDTM and Define in their environment. The Quality Checks are checks created in SAS®. When a final draft of the study metadata specification is available, the following metadata quality checks are performed:

- Shire Metadata Consistency Checks
- Shire Business Checks

SHIRE METADATA CONSISTENCY CHECKS:

After drafting the Shire SDTM Toolkit's metadata for the next annual release but prior to sending it to strategic partner CROs, the metadata is checked for consistency against Shire's Standards, such as CDASH, ADaM, plus Shire defined controlled terminology.

The following are some examples of the checks performed:

- The Controlled Terminology used in the Shire metadata specification has been defined in the Shire Terminology file using the same attributes.

- Variables defined across domains have the same definitions when appropriate.

SHIRE BUSINESS CHECKS:

Shire Business checks are specific to Shire's metadata development principles and ensures that any Shire specific metadata such as custom domains, variables or controlled terminologies, follow the Shire's CDISC metadata conventions so that there are no conformance issues when used in studies.

The following are some examples of the checks performed:

- Length of Label is greater than 40 characters
- Missing Derivation when Origin is Derived
- Marked as Not Used in a study but variable is required
- Inconsistent definition of controlled terminology
- Incorrect format for date variable
- Inconsistent mapping of SDTM Datatype and Define Datatype
- Length of a variable when greater than 200
- Comparison against CDISC definitions to make sure they match

METADATA COMPLIANCE

Compliance is a state that is reached when the rules are followed. Metadata Compliance is a state that is achieved when received metadata meets the expectations of the defined metadata specifications.

Shire's SDTM Governance uses metadata compliance to compare the defined specifications to the CRO implemented Standards Library, and study metadata is compared to the implemented Standards Library. This helps Shire to oversee the metadata quality and compliance to Shire Standards, when the CRO provides SDTM deliverables for SDTM Toolkit and study. Following are some instances where metadata compliance checks are used at Shire to ensure efficiencies are gained using SDTM standards:

- Metadata Checks: When a CRO implements SDTM and Define-XML in their own tools and processes (e.g., in their own library), the resulting metadata is compared against Shire's internal library. Meta Checks are performed to confirm the quality of the resulting library.
- Shire Standards Compliance Checks: When the Shire SDTM Standard Library implemented at the CRO is used by a study, the resulting study metadata is compared to Shire Standards Library. The study metadata is checked using the Shire Standards Compliance Checks. These checks allow Shire to determine how studies have used the Shire SDTM Toolkit and understand usage of how much metadata is compliant to Shire Standards.

METADATA CHECKS:

Metadata quality checks are performed on the delivered SDTM domain and Define-XML, see below:

- Metadata Check of Define XML – The Define-XML is the important component of the submission as it allows a reviewer to navigate through the study components. The regulatory agency reviewer checks Define-XML for compliance with the SDTM Metadata Submission Guidance (SDTM-MSG) concepts. Following are some of the checks performed:
    - Description of Check: This represents the item to be checked in the Define XML.

- Location of Information: This represents where the information is documented in the Shire Specification Document and where it is displayed in the Define XML file.

Pre-programming	
Description of Check	Location of Information
Domain description makes sense. (e.g. The description for all supplemental variables is "Supplemental Qualifiers for XX").	iSpec: Domains tab XML: Tabulation Datasets table
Sorting derivations make sense with domain's structure and key variables (Note: Key variables can be different depending on the study needs) (--SEQ is not required to match a domain's keys).	iSpec: Domains tab iSpec: Variables tab, --SEQ derivation XML: Tabulation Datasets table XML: Variable Level table, --SEQ derivation
Variable derivations are accurate and written in proper English. Detailed pseudo-code should not be present, although simple mathematical statements are allowed (e.g., the calculation for --DY variables).	iSpec: Variables tab, [Derivation / REFM: Method Reference] XML: Variable-Level Table, [Derivation/Comment]
Programmatic Checks	
Ensure all variables have an origin. (**Note:** Variables that have a value-level section might have multiple origins and thus their variable level origin definition will be missing.)	XML: Variable Level table, Origin XML: Value Level table, Origin
Ensure all variables with an origin of Derived also have a computation method.	XML: Variable Level table, Variable XML: Computational Algorithms table
XML Review	
Ensure all hyperlinks are active and direct reviewer to the intended location.	XML: All tables
Ensure CRF Pages correctly describe the origin of the variables being explained.	XML: Variable Level table, Origin XML: Value Level table, Origin
Ensure value-level links goes to the correct section (e.g. Split domains).	XML: Variable Level table, Variable
Ensure the order of codelist values matches CRF ordering. Also, timing variables should be ordered accordingly.	XML: Controlled Terminology table

Fig 7: Metadata Checks Sample

SHIRE STANDARDS COMPLIANCE CHECKS:

The purpose of the Shire Standards Compliance Checks is to ensure that Shire SDTM standards requirements have been correctly implemented for a study as compared to the Shire Toolkit.

Any deviation from this will be flagged as a Violation or an Exception in a generated report which is sent to both the Shire study programmer and Standards Team for review.

- Violations are defined as significant departures from the Shire SDTM standards requirements.
- Exceptions are changes to the default values identified in the Shire SDTM standards.

This review may lead to a change or the need for a standards waiver request to be submitted to from Shire's Standards Governance. Shire has developed requirements to partners and vendors for the development of a SDTM Compliance Report. In doing so, Shire identified the need for a more consumable and machine-readable metadata from the defined SDTM Specifications. Following are some examples of the Shire Standards Compliance Checks:

2.1.1 General Requirements

Req#	Requirement
G01	The compliance report identifies important differences in metadata between a specific implementation (studies, library updates, etc.) and Shire's standard library.
G02	There are two categories of interest in a compliance report: • Violations • Exceptions
G03	Violations are significant violations of Shire's standard library. Differences in this category are expected to be corrected, or may be the basis for a Waiver request.

2.1.2 Functional Requirements

Req #	Requirement
F01	There are currently 6 classes of metadata to be used in the comparison: • Dataset/Domain • Variable • Value-level per Variable • Value-level per SUPPQUAL • External Dictionaries • Methods
F02	Dataset/Domain metadata identify dataset-level attributes for comparison. The key variable is the dataset name. Once matched on this key, any differences in attributes identified as Violations/Exceptions would be captured in the compliance report.

Fig 8: Compliance Checks Requirements Sample

CONCLUSION

This paper discusses how Shire standardized metadata is represented as an SDTM Toolkit, and how that metadata is leverage to check for compliance with Shire Standards specifications.

The SDTM specification represents a vendor-neutral presentation of metadata that aids in distributing them across multiple strategic CRO Partners. Developing a metadata SDTM Toolkit also helps to shorten the study development and ensures consistent implementation. Quality checks are performed to validate updates to the toolkit and communicate changes made in the release.

By assessing a study's compliance to Shire Standards early in the development process, helps to reduce potential re-work and impact to study timelines or budget. Agreement on items that need careful consideration is achieved prior to the SDTMs being created. This helps Shire to oversee and manage standards. When a study deviates from the Shire standard, a standards waiver request is sent through the Shire Governance Model.

Quality metrics can be assessed about the SDTM Toolkit development to help assess the quality of the process as well as offering insights into the robustness of the SDTM Toolkit and its implementation.

Metadata compliance with an automated compliance report generation helps Shire to assess trends on how much Shire SDTM standards are modified within studies. This provides Shire information about their clinical development process starting with protocol development. It also provides insight on topics that need additional guidance to the CRO and Shire Study teams through internally offered SDTM training clinics and training materials.

This overall SDTM standards strategy helps us to receive quality deliverables such as SDTM domains and Define-XML.

REFERENCES

1. CDISC SDTM Model 1.4 - https://www.cdisc.org/system/files/members/standard/foundational/sdtm/study_data_tabulation_model_v1_4.pdf
2. CDISC Study Data Tabulation Model Implementation Guide (SDTMIG) Version 3.2
3. Definition of Quality – Oxford Dictionaries https://en.oxforddictionaries.com/definition/quality

ACKNOWLEDGMENTS

I would like to acknowledge Diane Piper who provided a framework to establish data standards within Shire and SDTM Governance Team for their continued support.

CONTACT INFORMATION

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

Veena Nataraj
Shire
vnataraj@shire.com
www.shire.com

Erica Davis
Shire
erdavis@shire.com
www.shire.com