

Streamlining the Metadata Management Process Using SAS® Life Science Analytics Framework

Alex Ford, SAS Institute

ABSTRACT

It was not long into my clinical programming career before I discovered that CDISC is truly an acronym for “Can Do It Somewhat Correctly”. Each run of a validation report uncovered new warnings or errors followed by tracking down the source of those issues to log and report for a define.xml. The latest release of the SAS® Life Science Analytics Framework (LSAF) provides a centralized framework where standards can be imported and live alongside a study and its data, managed by a graphical user interface. By associating a data standard, controlled terminology, and dictionaries with a study, team leads have the data and information necessary to produce a define.xml at the click of a button. Join us as we explore the metadata management features available in LSAF 5.1 which enable programmers of all levels to manage data standards correctly the first time, saving studies both time and money.

INTRODUCTION

LSAF is an end to end clinical platform that was specifically designed for life science customers and partners to help streamline the statistical progress of clinical trials. This is accomplished by housing all study related artifacts, providing a native code editor, establishing individual workspaces, mitigating inherent risks through audit trail and versioning, and in this latest release - providing a metadata management framework. This paper will take a deep dive into the metadata aspect and discuss how it assists clinical research teams to adhere to CDISC standards and enable Define.xml production at the click of a button.

The metadata management framework takes CDISC formats, controlled terminology, and external dictionaries and associates them with individual studies. This enables LSAF to run checks against the standards to report findings for reconciliation prior to define creation. LSAF also generates value level metadata and analysis results metadata which are combined with supporting documents in the define. From an admin perspective, metrics are provided giving team leads a high-level overview of associations between studies and standards for impact analyses. In the next sections we will walk through the look and feel of the framework and highlight how it makes your life easier as a programming team lead.

CLINICAL METADATA MANAGEMENT

Within LSAF, there is a specific section dedicated to Clinical Management which is broken down into the subsections Data Standards, Controlled Terminology, and External Dictionaries. These three subsections house the entirety of the metadata management materials which are associated with studies and used in the creation of a define package. The Data Standards consist of CDISC standards, while Controlled Terminology and External Dictionaries are flexible enough for manual definitions. Each entry is classified by type, has information concerning the standard on which they are based, and are versioned. The published state of a standard indicates whether that standard is in development, production, or retired. Only standards in the production state can be associated with new studies and only individuals with proper write access can develop new standards.

The screenshot shows the 'Data Standards' section of the SAS Life Science Analytics Framework. A table lists six standards with the following columns: Name, Type, Standard Model, Base Standard Name, Base Standard Version, Published State, Version, and Date Modified.

	Name	Type	Standard Model	Base Standard Na...	Base Standard ...	Published State	Version	Date Modified
<input type="checkbox"/>	CDISC ADaM 1.1	Analysis	Analysis Standard	ADaM	1.1	Production	4	Sep 14, 2018 09:52 P...
<input type="checkbox"/>	CDISC SDTM 3.2	Tabulation	Tabulation Standard	SDTM	3.2	Production	3	Apr 26, 2018 07:32 AM...
<input type="checkbox"/>	CDISC SDTM 3.2 Dermatitis	Tabulation	Tabulation Standard	SDTM	3.2	Production	1	Aug 15, 2018 09:59 AM...
<input type="checkbox"/>	CDISC SDTM 3.2 Pneumococcal	Tabulation	Tabulation Standard	SDTM	3.2	Production	3	Oct 21, 2018 11:25 PM...
<input type="checkbox"/>	CDISC SDTM 3.2 Vaccines	Tabulation	Tabulation Standard	SDTM	3.2	Production	3	Oct 04, 2018 01:46 AM...
<input type="checkbox"/>	CDISC SEND 3.1	Tabulation	Tabulation Standard	SEND	3.1	Production	3	Apr 26, 2018 07:26 AM...

Figure 1. Data Standards

After picking a standard from the list in Figure 1 and associating it with a new study, study teams would begin by filling in the pertinent study information as it would be displayed in a CSR submitted to the FDA for review.

The screenshot shows the 'Study Details' for the 'CDISC SDTM 3.2 Vaccines' standard. The details are organized into sections: Details, Column Groups, Tables, Validation Results, and Studies. The 'Studies' section contains the following information:

Standard model:	Tabulation Standard
Standard type:	Tabulation
Base standard name:	SDTM
Base standard version:	3.2
Current state:	Production
State change comment:	
Published state:	Production
Published version:	3
Date created:	Apr 26, 2018 07:29 AM GMT-04:00
Created by:	Melissa Martinez
Date modified:	Oct 04, 2018 01:46 AM GMT-04:00
Modified by:	Melissa Martinez

Figure 2. Study Details

Within each data standard, column groups are defined. The column groups house the variable names, labels, order, requirement, and computational methods. For example, here is the Associated Persons group:

The screenshot shows the 'Column Groups' tab in the SAS Life Science Analytics Framework. The main content area displays a table with 4 rows and 10 columns. The columns are: Column Group, Name, Description, Order, Core, Mandatory, Origin Type, Code List Reference, Computational Method, and Submission Data Type. The rows represent different associated persons and their relationships to the subject.

Column Group	Name	Description	Order	Core	Mandatory	Origin Type	Code List Reference	Computational Method	Submission Data Type
Associated Persons	APID	Associated Pe...	1						Char
Associated Persons	RSUBJID	Related Subject	2						Char
Associated Persons	RDEVID	Related Device	3						Char
Associated Persons	SREL	Subject, Devic...	4						Char

Figure 3. Column Groups

The Tables tab contains the name of each of the tables that is required as part of a particular data standard as well as its description, structure, order, and purpose.

The screenshot shows the 'Tables' tab in the SAS Life Science Analytics Framework. The main content area displays a table with 51 rows and 12 columns. The columns are: Name, Description, Structure, Order, Required for Standard, Comment, Domain, Domain Description, Class, Purpose, and Custc. The rows list various tables such as AE, APDM, APRELSUB, CE, CM, CO, DA, DD, DM, DS, DV, EC, EG, EX, FA, and HO.

Name	Description	Structure	Order	Required for Standard	Comment	Domain	Domain Description	Class	Purpose	Custc
AE						AE		Events		
APDM						APDM		Associa...		
APRELSUB						APRELSUB		Associa...		
CE						CE		Events		
CM						CM		Interven...		
CO						CO		Special...		
DA						DA		Findings		
DD						DD		Findings		
DM						DM		Special...		
DS						DS		Events		
DV						DV		Events		
EC						EC		Interven...		
EG						EG		Findings		
EX						EX		Interven...		
FA						FA		Finding...		
HO						HO		Events		

Figure 4. Tables

LSAF is able to run validation checks using the data standards that have been associated. The results of these checks are compliance reports which are exported to the study folder.

Severity	Object Type	Domain	Column	Column Group	Attribute	Problem
①	Class	APDM			Class	Undefined value Associated Persons for column Class.
①	Class	APRELSUB			Class	Undefined value Associated Persons for column Class.
①	Class	CO			Class	Undefined value Special-Purpose for column Class.
①	Class	DM			Class	Undefined value Special-Purpose for column Class.
①	Class	SE			Class	Undefined value Special-Purpose for column Class.
①	Class	SV			Class	Undefined value Special-Purpose for column Class.

Figure 5. Validation Results

Under the studies tab there is additional information that is populated for the creation of the define package. All tables, the associated CT, and the relevant external dictionaries are defined at this level.

Standard	Name	Description	Structure	Order	Required for Standard	Comment
CDISC ADA...	ADAE	Adverse Events A...	one record per subject per adverse e...	3		See SAS pro...
CDISC ADA...	ADQSA...	ADAS-Cog Analysis	One record per subject per paramet...	2		See referenc...
CDISC ADA...	ADSL	Subject Level An...	one record per subject	1		Screen Failu...

Figure 6. Study Definitions

The study specs, programs, and outputs are housed within LSAF, so the framework can derive Value Level Metadata and Analysis Results Metadata within the study level.

Table	Column	Where Clause	Where Clause Comment	Description	Order	Submission Data Type	SAS Field Name
ADQSA...	QSSEQ	PARAMCD NE "...		Sequence Nu...	1	Num	QSSEQ
ADQSA...	AVAL	PARAMCD EQ "...		Analysis Value	2	Num	ACTOT
ADQSA...	QSSEQ	PARAMCD EQ "...		Sequence Nu...	2	Num	ACTOT
ADQSA...	DTYPE	PARAMCD EQ "...		Derivation Type	2	Char	ACTOT
ADQSA...	DTYPE	PARAMCD NE "...		Derivation Type	1	Char	DTYPE
ADQSA...	AVAL	PARAMCD IN ("...		Analysis Value	1	Num	AVAL

Figure 7. Value Level Metadata

Lastly, all relevant supporting documents are packaged at the study level. This could include the protocol, CRF, study specifications, SAP, computational algorithms, etc.

Table	Column	Where Clause	Document Type *	Title	HRef	Page Reference Type	Page Reference	Analysis
ADAE			COMMENT	adae.sas	adae.s...			
ADQSA...			COMMENT	Analys...	adrg.pdf	NamedDestination	Section2.1	
ADQSA...			COMMENT	adgsa...	adgsad...			
ADQSA...	AVAL	PARAMCD EQ "...	METHOD	Analys...	adrg.pdf	PhysicalRef	3	
ADSL			COMMENT	Analys...	adrg.pdf	PhysicalRef	6	
			DISPLAY	Table ...	dummy...	PhysicalRef	3	RD Table
			RESULTDOC	Statisti...	dummy...	PhysicalRef	4	RD Table
			RESULTDOC	Analys...	adrg.pdf	PhysicalRef	6	RD Table
			RESULTCODE	at14-5...	at14-5...			RD Table
			RESULTDOC	Statisti...	dummy...	PhysicalRef	4	RD Table
			DISPLAY	Table ...	dummy...	PhysicalRef	2	RD Table
			STUDY	Statisti...	dummy...			
			RESULTDOC	Statisti...	dummy...	PhysicalRef	5	RD Table
			RESULTDOC	Clinica...	dummy...	PhysicalRef	3	RD Table
			STUDY	Analys...	adrg.pdf			
			RESULTCODE	at14-5...	at14-5...			RD Table
			STUDY	Clinica...	dummy...			
			RESULTS	at14-5...	at14-5...			RD Table

Figure 8. Supporting Documents

Once all study level assets have been defined and statistical development of tables, listings, and figures is complete, LSAF has a GUI interface that allows for creation of the define.xml. This GUI bundles together the data standard, VLM, ARM, CTs, External Dictionaries, and all manually input study level information and generates a Define.xml package.

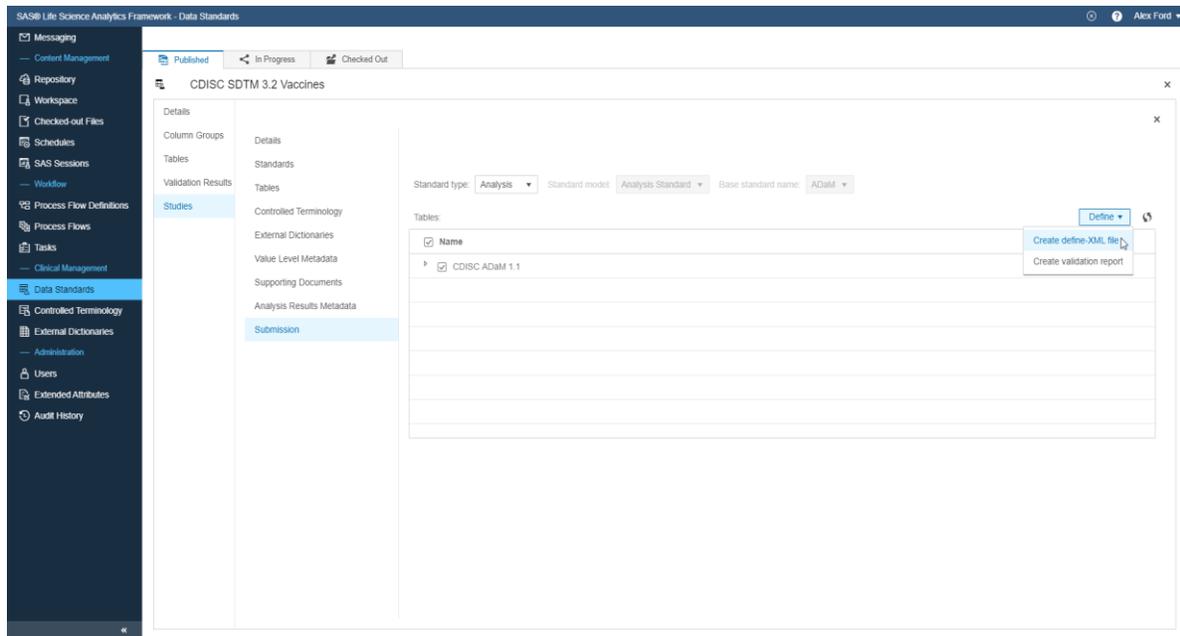


Figure 9. Define Creation

CONCLUSION

LSAF is intended to be an end to end clinical platform that assists in governing and streamlining the submission process. LSAF V5.1 has extensive metadata management capabilities which are designed to assist programming teams in meeting CDISC data standards prior to submission. The metadata management portion of the framework is capable of housing and versioning standards as they are established by a standards governor. Those standards are then associated with studies and used for compliance checks and define.xml generation. Traditionally, define creation has been a manual and disjointed process. LSAF empowers statistical teams to contain their entire development and QC process within a single unified system to save time and to enable users to complete CDISC compliant datasets correctly the first time.

REFERENCES

https://www.sas.com/en_us/software/life-science-analytics-framework.html

CONTACT INFORMATION

Any questions or comments are encouraged. Contact the authors at:

Alex Ford

SAS Institute

100 SAS Campus Drive

Cary, NC, 27513

Email: Alex.Ford@sas.com

Brand and product names are trademarks of their respective companies.