PharmaSUG2015 - Paper AD04

Accelerate define.xml generation using defineReady

Senthilkumar Karuppiah, Navitas LLP, Chennai, India

Georgina Wood, Navitas LLP, Princeton, New Jersey, USA

ABSTRACT

With FDA & other regulatory agencies mandating electronic submissions and standardized data, Pharmaceutical companies & sponsors are gearing up for their submissions in CDISC SDTM and ADaM based standards. An important component of such submissions is the generation of the CRT-Data Definition Document (define.xml). Generating define.xml - based on metadata document gives more challenges to SAS programmers in the space of automation. This paper will discuss the detailed mechanism/automated way of generating and validating the define.xml from a user interface (GUI) and back end with the SAS® integrated modules and pre-loaded metadata file. Also this paper takes a quick look on, how the defineReady will automatically generate and validate the components of define.xml like dataset metadata, variable level metadata, codelist, valuelist, origin and comments. defineReady will also ensure the projects are organized SPONSOR/THREAUPTIC/STUDY wise and support multi-user capabilities along with audit trail and validation reports in a regulated environment.

INTRODUCTION

The FDA reviewer requires a document which navigates through the study data presented in individual dataset. CDISC define.xml is the one such document which describes the structure and content of clinical data in a machine readable format for electronic submissions of CDISC datasets, such as SDTM and ADaM.

The define.xml file is required to clearly describe the structure, content and the relationship between various components of collected data in XML format and help reviewers easily access and understand the clinical data they review.

The define.xml describes metadata in 5 Levels:

1. Table Level Metadata:

Datasets for Study

Dataset	Description	Class	Structure	Purpose	Keys	Location
AE	Adverse Events	Events	One record per adverse event per subject	Tabulation	STUDYID, USUBJID, AETERM, AESTDTC	AE.xpt

2. Variable Level Metadata:

Variable	Label	Type	Controlled Terminology	Origin	Role	Comment
STUDYID	Study Identifier	text		Protocol, CRF Page 1	IDENTIFIER	The STUDYID variable has a fixed format: 'XXXX-YYYY', where 'XXXX' indicates the 4-digit compound code and the 'YYYY' the 4-digit study code
DOMAIN	Domain Abbreviation	text	DOMAIN	Assigned	IDENTIFIER	
USUBJID	Unique Subject Identifier	text		Derived	IDENTIFIER	The USUBJID variable has a fixed format: 'XXXX-YYYY-ZZZZZ', where 'XXXX' indicates the 4-digit compound code, 'YYYY' the 4-digit study code and 'ZZZZZ' the 5-digit patient code

3. Value Level Metadata:

Value Level Metadata (ValueList.VS.VSTESTCD)							
Source Variable	Value	Label	Type	Controlled Terminology	Origin	Role	Comment
VSTESTCD	BMI	BODY MASS INDEX	text		Derived		See Computational Method: COMPMETHOD.BMI
VSTESTCD	DIABP	DIASTOLIC BLOOD PRESSURE	text		CRF Page 24		
VSTESTCD	HEIGHT	HEIGHT	text		CRF Page 22		
VSTESTCD	PULSE	PULSE RATE	text		CRF Page 24		
VSTESTCD	SYSBP	SYSTOLIC BLOOD PRESSURE	text		CRF Page 24		
VSTESTCD	WEIGHT	WEIGHT	text		CRF Page 22		

4. Controlled Terminology Metadata:

SEX, Reference Name (SEX)	
Code Value	Code Text
F	FEMALE
M	MALE

5. Computational Algorithm Metadata:

Computational Algorithms (COMPMETHOD.AGE)	
Reference Name	Computation Method
COMPMETHOD.AGE	Equals to (DM.RFSTDTC-DM.BRTHDTC)/365.25

WHY DEFINE-READY

In the space of automation, generating define.xml requires multiple eyes on each area like metadata repository building, handling sponsor defined data along with standards, identify CRF origins which are very daunting and also time consuming jobs. It requires a lot more in-depth understanding in terms of other technologies like XML, Style Sheets, SAS®, interacting with PDF/RTF/XLS and also important to adhere standards of CDISC/FDA.

Under the umbrella of define.xml are many sub-challenges, the collection of the above mentioned metadata, which can make the generation of define.xml an intimidating task. For starters, there's the organization of metadata, something that may have received minimal attention at most, in the past.We are now not only forced to think about it, but we have to keep track of it and organize it. We have to consider not only simple dataset and variable attributes such as lengths and labels, but also those that CDISC has deemed important such as the role of variables and their expected values, documentation of computational algorithms, and many more. We need to plan a database for it and consider all the different ways to collect this information. Hence lot more companies are ready to invest their resources in understating the define xml requirements and also complexity of generation.

Hence NAVITAS is planning to create solution in the name of defineReady, which will help generate the define.xml to meet the FDA requirement in a regulated environment. defineReady will have a predefined process with the minimal user interaction togenerate the desired output. It is a part of a suite of tools which can be implemented to provide data for your submission that is compliant to regulatory requirements. This automated process delivers data with greater accuracy and integrity.

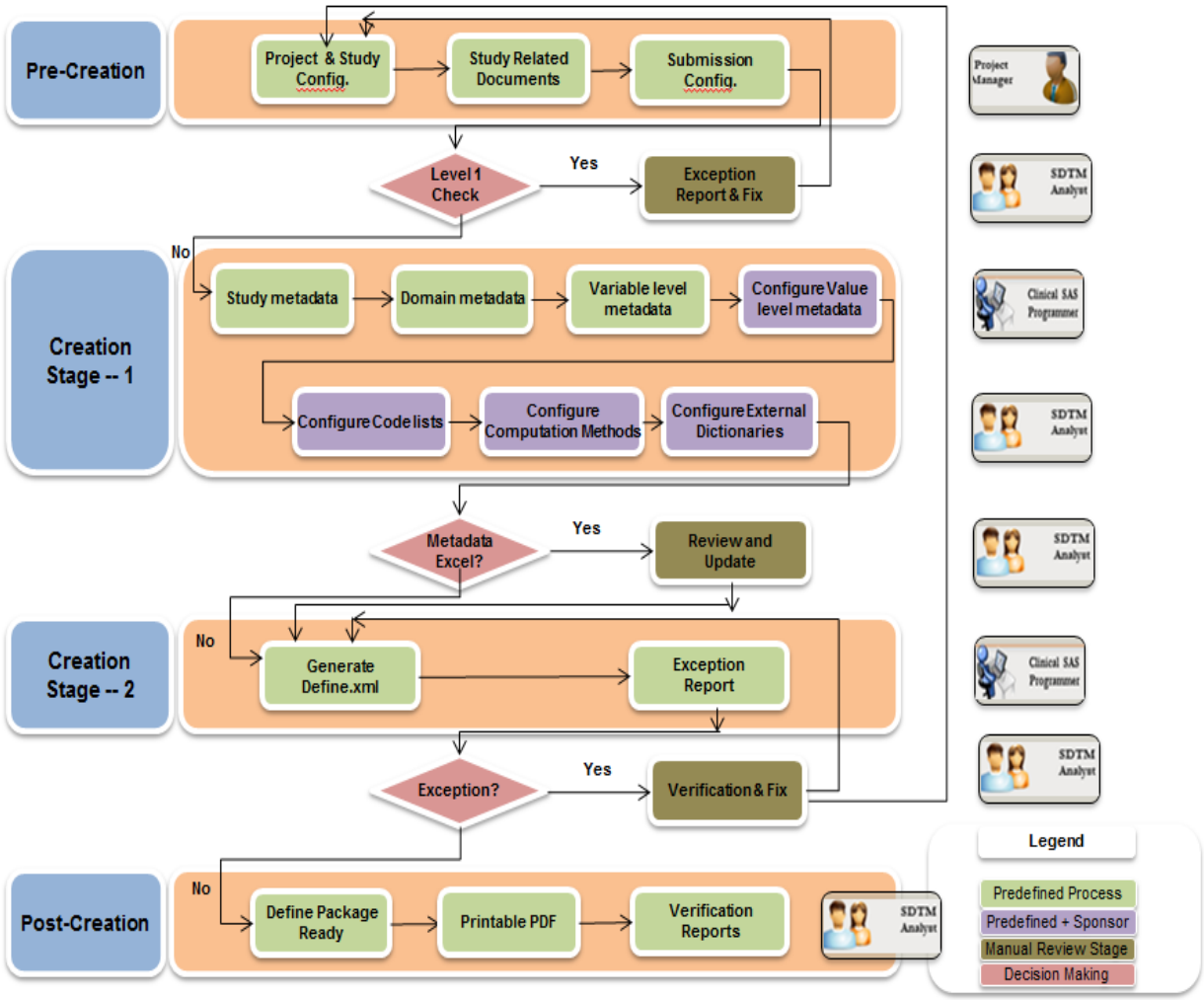
DEFINE XML CREATION

NAVITAS'defineReady is a SAS® and .NET integrated solution to generate and validate define.xml in a regulated environment.

The prerequisites of defineReady are

- Datasets/XPTs
- Annotated Case Report Form
- Date Guide & Supplemental Documents based on the SDTM/ADaM Submission
- Derivation Documents in standard template WORD/XLS

defineReady -- Process flow 1



KEY FEATURES OF DEFINE-READY

1. Pre-loaded Global Metadata Repository to support all SDTM/ADaM versions.
2. Able to handle the multiple versions of define.xml (various versions of style sheets for the respective submissions of ADaM and SDTM).
3. System will support the sponsor defined code lists along with pre-loaded standard codelist provided by National Cancer Institute.
4. System will auto import the derivation documents and annotated CRFs.
5. Handling of CUSTOM domains and its corresponding variables, SPLITTING domains and SUPPQUAL domains.
6. Able to support multiple versions of Input excel and also multiple versions of define xml.

OVERVIEW OF THE GENERATED REORTS

1. Able to generate the compliance report with standards.
2. User friendly reports at various levels (like Domain Level, Variable Level, Value Level and Codelist).
3. Versioned Input EXCELs/Metadata EXCELs reports.
4. Validation Reports (against Standards & against DATA).
5. Printable PDF file.
6. Project status report by different stages by respective users. (for audit-trail).

CONCLUSION

The defineReady tool will help in generating a validated define.xml in a regulated environment which adheres to the respective submission standards of SDTM/ADaM. Also this tool optimizes the whole end-end process of define xml creation and is very user friendly.

ACKNOWLEDGMENTS

The authors would like to acknowledge the help of all the NAVITASpersonnel involvedin preparing this paper and PharmaSUG for providing the opportunity to present the same.

RECOMMENDED READING

- CDISC Define-XML Specification and Implementation Guides.
- CDISC SDTM Standard and Implementation Guide.
- ADaM standard and Implementation Guide.

CONTACT INFORMATION

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

Name : Senthilkumar Karuppiah
Enterprise : Navitas LLP
Address : No 16, GST Road, Perungalathur
City,StateZIP: Chennai, India 600063
Work Phone:+91 44 4590 9105
Fax :+91 44 4590 9099
E-mail :senthilkumark@navitas.net