

PharmaSUG 2012 - Paper DS15

Building Flexible ADaM Analysis Variable Metadata

Songhui Zhu, K & L Consulting Services, Fort Washington, PA

Lin Yan, Celgene Corporation, Summit, NJ

ABSTRACT

Building variable metadata for a clinical study is tedious and time-consuming in the process of ADaM development. Especially, changes in variables and analysis rules from various sources make harmonization of metadata for multiple parallel studies challenging. There is a need to build meta data with great flexibility to accommodate various dynamic changes in variables and analysis rules.

In this paper, methods for creating variable metadata with various flexibilities are presented. The flexibilities include that to accommodate the changes in attributes from SDTM data, that to accommodate the changes in attributes of variables in common variables, that to accommodate the changes in analysis time points, and that to accommodate the changes in analysis derivation rules. In addition, principles for grouping variables are also proposed.

1. INTRODUCTION

In Version 2.1 of the CDISC Analysis Data Model, there are three levels of metadata: analysis dataset metadata, analysis variable metadata (including analysis parameter value-level metadata), and analysis results metadata. In the process of building those metadata, variable metadata is a familiar but time-consuming task. The common practice in pharmaceutical industry is that a company creates its company-wide standard ADaM variable metadata, then each study copy that metadata and modify it to set study-level variable metadata. By doing this, the time used to create study-level variable metadata can be greatly reduced and consistency of variable metadata among studies can be achieved to some degree. However, due to the generality and stability of a company-wide variable metadata, if a study-level variable metadata is based only on a company-wide metadata, it would not be flexible enough to handle the changes

from SDTM data and special needs of analysis for a study. For example, if a new variable is added to study SDTM data, or the length of a SDTM variable in main domain is changed, the changes need to be manually accommodated in ADaM variable metadata. If an analysis rule which applies to multiple analysis datasets is modified, a corresponding modification need to be done manually to each affected dataset.

In this paper, authors present methods of building ADaM analysis variable metadata with flexibility needed to accommodate the changes in variables from SDTM data, changes in variables from ADSL, and changes in analysis rules. When computer programs are developed, those methods can give the developers of metadata the flexibility needed to handle the changes in SDTM data, changes in common variables in ADSL, changes in analysis time points and changes in analysis derivation rules. Furthermore, principles about how to divide metadata into segments are also proposed.

2. METHODOLOGY

One may divide the process of metadata generation into two stages. At the first stage, one need:

- a) copy common metadata to draft study variable metadata ;
- b) analyze the source variability for a study and divide the metadata into several parts according to the source variability;
- c) set the metadata fields which would be affected by the changes in the source blank.

Then at the second stage, one may:

- d) feed the affected metadata fields with corresponding values from the source;
- e) integrate metadata parts dynamically and automatically.

This method will be explained in details using examples.

2.1 STAGE 1: DEVELOPING DRAFT METADATA TO EMBRACE SOURCE VARIABILITY

After copying the standard metadata to generate draft study-level metadata, the metadata developer need to analyze which variables will be changed due to the source variable changes. The change source could be 4 parts. The attributes change in SDTM,

the attributes change in ADSL for non-ADSL datasets, the analysis time point change and the derivation rule change.

2.1.1 Embracing changes in attributes of variables from SDTM data

According to CDISC Analysis Data Model Version 2.1, ADaM variables from SDTM must be “same name, same meaning, same values ”[2]. This means that the attributes of variables from SDTM data must be kept in ADaM. If there are any changes in the attributes of variables from SDTM data, the corresponding changes must be made in ADaM variable metadata. Further, the changes in attributes of variables from SDTM data may happen at any time. This makes the harmonization of attributes of variables between ADaM and SDTM more difficult. In order to comply with ADaM standards, it is essential that the variable metadata be flexible enough to accommodate these changes and easy to update dynamically.

In draft ADaM variable metadata, one may keep only the variable names and necessity status for all variables from SDTM, while other descriptions such as labels, types, lengths, and sources will be filled in from SDTM metadata dynamically and automatically. This can be illustrated by the following example.

Table 2.1.1 Example of Metadata to Embrace the changes in SDTM data

Dataset Name	VariableName	Variable Label	Variable Type	Length	Display Format	Codelist /Controlled Terms	Core	Source	Derivation
ADSL	STUDYID						Req	SDTM DM	
ADSL	USUBJID						Req	SDTM DM	
ADSL	SITEID						Req	SDTM DM	
ADSL	SUBJID						Req	SDTM DM	
ADSL	SEX						Req	SDTM DM	
ADSL	SEXN	Sex (N)	integer	3		1='M', 2='F', 3='U', 4='UN';	Perm		If DM.SEX='M' then SEXN=1;

In the example above, variables except SEXN are from SDTM. The label, type, and length of each of those variables is intentionally set to blank in the draft variable metadata but can be read from SDTM metadata or working specs dynamically by computer programs [4]. When there is any change in the attributes of any SDTM

variable, rerun of the computer programs will keep ADaM metadata consistent with SDTM data.

2.1.2 Embracing the changes in attributes of common variables

Some variables in ADSL such as basic demographic variables and treatment variables are need in some non-ADSL data sets. In addition, number of common variables and attributes of some common variables may change at any time in the process of metadata development. In those cases, if those common variables are fixed in each variable metadata, it would be time-consuming to modify them in each dataset every time when any attributes of any common variables have been changed. To harmonize the common variables in each dataset with ADSL, one may adopt an approach illustrated in Table 2.1.2.1.

For a variable specifically needed for an analysis dataset, one may insert only the variable name into the variable metadata of the non-ADSL dataset. Other information about this variable can be copied from ADSL metadata dynamically. For example, in ADLB, variables such as subgroup analysis variables AGEGR1 and AGEGR1N shown in Table 2.1.2.1 need to be copied from ADSL. When the attributes for these two variables in ADSL changes, these two variables need be changed accordingly. To keep consistency between ADSL and other ADaM datasets, one may keep only the variable names in metadata for each ADaM dataset when first creating the study-level variable metadata and other information about these variables can be read from ADSL dynamically by computer programs. However the info of Core, Permissible, or Conditional Permissible of the variable need to be specified in the metadata even in the first round of creation.

Table 2.1.2.1 Example of Metadata to Embrace the changes in ADSL variables - 1

Dataset Name	Parameter Identifier	Variable Name	Variable Label	Variable Type	Length	Display Format	Codelist / Controlled Terms	Core	Source	Derivation
ADLB	*ALL*	ADT	Analysis Date	integer		Date9.		Perm		Date part of LB.LBDTC
ADLB		AGEGR1						Perm	ADSL	
ADLB		AGEGR1N						Perm	ADSL	

For variables in ADSL needed for each other dataset, one may mark those variables in ADSL and copy them into non-ADSL metadata using programs as illustrated in Table 2.1.2.2.

Table 2.1.2.2 Example of Metadata to Embrace the changes in ADSL variables - 2

Dataset Name	Variable Name	Variable Label	Variable Type	Length	Display Format	Codelist /Controlled Terms	Core	Source	Derivation	Common Variable	Sub Common Variable
ADSL	FASFL	Full Analysis Set Populatio	text	1	\$1	Y, N	Cond		Y when took at least one study drug		ADEFF, ADTE
ADSL	TRTSDT	Date of First Exposure	integer	8	YYMM DD10.		Cond		Date part of SDTM DM.RFSTD	Y	

In the example above, a column “CommonVariable” with value Y is used to mark those variables will be copied into each analysis dataset. If there are certain variables needed by a few specific analysis datasets, not each analysis dataset, for example, FASFL (Full Analysis Set Population Flag) for ADEFF (Efficacy analysis dataset), an additional column “Sub Common Variable” marked with the names of datasets which need those ADSL variables could be added to draft variable metadata. Should any change occur there is no need to perform the corresponding modifications in each dataset. What you need is to just re-run the metadata creation programs.

2.1.3 Embracing the change in analysis time points

It is very often that an analysis visit window is defined for BDS such as ADLB, ADVS, and efficacy datasets. From one study to another, the analysis time points may be

different. If study-level metadata are built completely from company-wide metadata, the changes in analysis time points may need to be made in each BDS data set. It is better to keep the definition of analysis window in a centralized metadata and give a link from each BDS to this metadata. For example, one may design the metadata for visit window as shown in Table 2.1.3. Thus, the metadata is easy for reuse.

Table 2.1.3 Example of Metadata to Embrace the changes in analysis time points

Variable Name	Identifier	Code	Decode
AVISIT	*DEFAULT*	SCREENING	< Day 1
		BASELINE	Day 1
		WEEK 2	2- 21 Days
		WEEK 4	>= 22 Days
AVISITN	*DEFAULT*	-99	SCREENING
		0	BASELINE
		2	WEEK 2
		4	WEEK 4
AVISIT	ADEG	BASELINE	Day 1
		WEEK 4	> Day 1
AVISITN	ADEG	0	BASELINE
		4	WEEK 4

In the analysis time point metadata, time point variables are listed in the first column. The values of the those variables and derivation rules are shown in the third and fourth columns. Different endpoints may have different time schedule which can be indicated using IDENTIFIER in the second column. The value DEFAULT in INDENTIFIER means that the time schedule and derivation rules will apply to each endpoint except those specified in the column, ADEG in this case. The time points in dataset ADEG (in the example) are different from other endpoints due to different trail schedule.

2.1.4 Embracing the changes in analysis derivation rules

Specifying analysis derivation rules in variable metadata is another tedious and time-consuming part. Derivation rules are always in change during the creation of metadata and whole analysis reporting process. To accommodate the changes in derivation rules, one may create separate metadata and link it to variable metadata dynamically. An example is given in table 2.1.4.1.

Table 2.1.4.1 Example of Metadata to embrace the changes in analysis derivation rules - 1

DerivationRule Seq	Identifier	DerivationRuleName	DerivationRule
1	*DEFAULT*	Baseline Flag	Y when the record is the last non-missing value when date <= ADSL.TRT01SDT
1	*ADEG*	Baseline Flag	Y when AVISITN = 0
2	*ALL*	Planned Treatment - Record level	ADSL.TRT01P when TR01SDT<=date = TR01EDT; ADSL.TRT02P when TR02SDT<=date <= TR02EDT
3	*ALL*	Actual Treatment - Record level	ADSL.TRT01A when TR01SDT<=date = TR01EDT; ADSL.TRT02A when TR02SDT<=date <= TR02EDT

The derivation rules for determining baseline and record-level planned treatment are listed in an analysis derivation rule metadata. The derivation rule names and rules are listed in the third and fourth column. The column IDENTIFIER is used to indicate which endpoint will apply this rule. In this case, all endpoints except ADEG will share the same rule. The rule of determining record level planned treatment in this example will be applied to all BDS datasets.

A derivation rule can be copied to a BDS variable metadata using computer programs. Or the rule metadata can stand alone, as well. In the second approach, a sequence number can be created to connect the rule to the metadata. In BDS variable metadata itself, the derivation rule sequence is specified such that a link between variable metadata and derivation rule can be established.

Table 2.1.4.2 Example of Metadata to embrace the changes in analysis derivation rules - 2

Dataset Name	Parameter Identifier	Variable Name	Variable Label	Variable Type	Length	Display Format	Codelist / Controlled Terms	Core	Source	Derivation
ADLB	*ALL*	BASE	Baseline Value	integer/float	8			Cond		DerivationRule.Seq=1

2.2 DETERMINE VARIABLE METADATA SEGMENTS

The examples above show the possible variable metadata segments according to the varying source. Besides the varying sources from SDTM, ADSL for non-ADSL dataset, there are changes from analysis derivations. How to determine the segments related to analysis? One may group segments based on the following principles.

2.2.1 Group analysis rules or derivations that will be applied to multiple datasets

If a rule is only for one dataset, it can be directly inserted into variable metadata. This principle can be applied to variables related to baseline flag (ABLFL), analysis flag (ANLxxFL), treatment at record level (TRTP, TRTA), ...

2.2.2 Group analysis rules or derivations that generate results needed for code list

The variables could be grouped into this group are AVISIT, AVISITN, APERIOD, APHASE..etc. Those variables will have code list.

2.2.3 Group analysis rules or derivations that can be easily implemented using computer programs

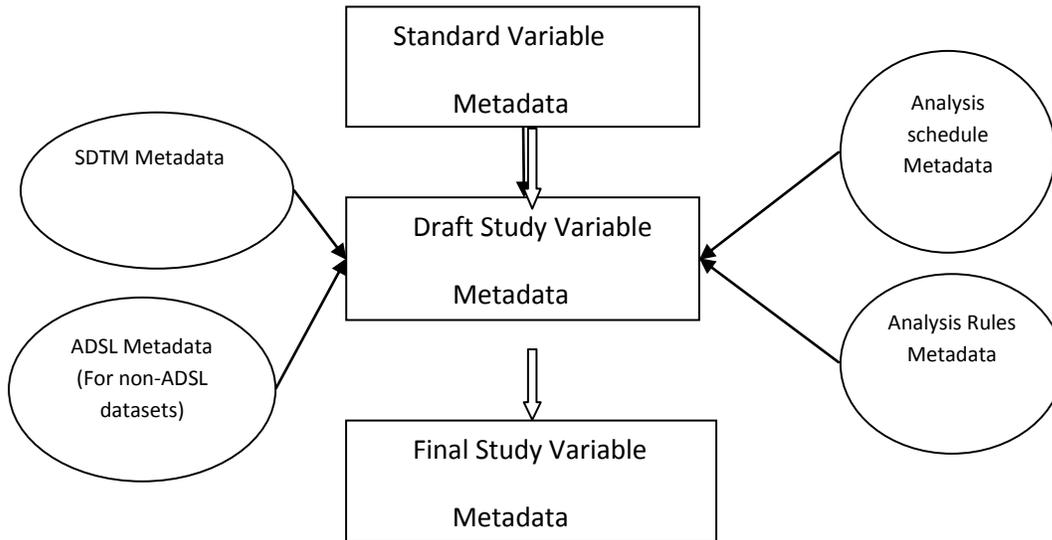
Grouping analysis rules which can be easily implemented using computer program would pave the road for automating the process of creating metadata and analysis datasets. The variables could be grouped into this group are CHG, PCHG, --DUR (simple calculated duration), --DY (simple relative day), R2BASE (Ratio to Baseline), R2A1LO(Ratio to Analysis Range 1 Lower Limit), R2A1HI(Ratio to Analysis Range 1 High Limit), etc.

2.3 STAGE 2: INTEGRATE METADATA SEGMENTS DYNAMICALLY

The process of building the flexible metadata starts from standard variable metadata. Based on that, with the dynamical link using programs to SDTM metadata, ADSL metadata (for non-ADSL datasets), analysis schedule metadata, and analysis rule metadata. Study level variable metadata can be built very flexible to accommodate the

changes from the sources and keep core parts consistent across studies . The flow of the process can be as follows shown in Figure 2.2.

Figure 2.2 The flow of the building flexible variable metadata process



3 CONCLUSIONS

Methods of building ADaM analysis variable metadata with flexibility that can accommodate the changes in variables from SDTM data, changes in variables from ADSL, and changes in analysis rules have been presented. The metadata built using these methods give the developers the flexibility needed to handle the changes in various sources dynamically. Computer programmers for updating the metadata built this way can be easily developed. Further, principles for dividing metadata into segments have also been proposed.

REFERENCES

- [1] CDISC Analysis Data Model Version 2.1.

[2] CDISC SDTM Implementation Guide (Version 3.1.2) CDISC Submission Data Standards Team.

[3] CDISC ADaM Implementation Guide Version 1.0 CDISC Analysis Data Model Team.

[4] Updating SDTM Metadata Excel File (Define.xls) with SAS, Lin Yan, CC07, PharmaSUG 2010, Orlando, FL.

ACKNOWLEDGEMENT

Songhui Zhu thanks the K & L Consulting Services and Lin Yan thanks Celgene Corp. for the support they received.

CONTACT INFORMATION

Your comments and questions are valuable and appreciated. Authors can be reached at

Songhui Zhu, PhD

K&L Consulting Service

1300 Virginia Dr., #408

Fort Washington, PA 19034 - 3266

Lin Yan

Biostatistics and Statistical Programming

Celgene Corporation

Summit, NJ 07920

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.