

ADaM Specifications Automation from Company Standards to Study

Zhi Xu, Zhiping Yan, Jianyong Tong, Xinyun Gao and Jiarui He, Dizal Pharma

ABSTRACT

ADaM specifications are the bridge from ADaM designs to ADaM developments. It is also an important component of the submission-ready Analysis Package preparation since a well-organized specification will facilitate ADaM define.xml generation for regulatory submission. In addition, a clear define specifications can be served as documents for both the developer and validator as well as the statistician to work on the ADaM datasets efficiently. Practically speaking, company-level standard ADaM specification in good shape should consist of 'global standard' part which should strictly follow industry standards such as ADaM IG and company standards like certain variable names, labels, and controlled terminology as well as a non-standard part which could be adjusted depending on the different compounds and study design. Moreover, the above-mentioned non-standard part can further be classified as 'compound-level standard' and 'compound-level non-standard' or say study-specific. Ideally, studies under the same compound (or compound/indication) should follow the same company standards as well as compound level standards. However, in most cases, the ADaM specifications were copied from one study to another study, and it is more likely to run into various problems like 1) not following the most up-to-date industry standard, 2) derivations which supposed to be identical across studies under one compound were different, 3) spend extra efforts on the already existed derivation rule... all these will cause low-quality and inefficiency. Attempting to resolve this, the appropriate process from company ADaM standards to study level is constructed, and a smart App was developed to consolidate the standard part (both global standard and compound-level standard) including controlled terminology management and bring a user-friendly interface to handle study-specific requirements.

INTRODUCTION

The Overview of the Standardization Process Workflow

First, we prepare the company-level ADaM specification which includes as much as ADaM that could be used in a study. As of now, we have ADSL, ADAE, ADBASE, ADLB, ADDILI, ADEX, ADEXSUM, ADCM, ADMH, ADCV, ADVS, ADEG, ADECOG, ADRE, ADDV, ADPC, ADPP, ADTL, ADRS, ADDATES, and ADTTE. Below is an example of ADSL, as you can see, column J is the derivation rule in English while column K is in Chinese. The SME will be responsible for the maintenance of this specification and collecting user feedback and releasing a new version to the team if major revision.

	A	B	C	D	E	F	G	H	I	J	K
1	Dataset	Order	Variable	Label	Data Type	Length	Format	Codelist	Origin	Comment	Comment_ch
2	ADSL	1	STUDYID	Study Identifier	Char	20			Predecessor	DM.STUDYID	DM.STUDYID
3	ADSL	2	USUBJID	Unique Subject Identifier	Char	40			Predecessor	DM.USUBJID	DM.USUBJID
4	ADSL	3	SUBJID	Subject Identifier for the Study	Char	20			Predecessor	DM.SUBJID	DM.SUBJID
5	ADSL	4	SITEID	Study Site Identifier	Char	8			Predecessor	DM.SITEID	DM.SITEID
6	ADSL	5	INVTNAM	Investigator Name	Char	50			Predecessor	DM.INVTNAM	DM.INVTNAM
7	ADSL	6	AGE	Age	Integer	8			Predecessor	DM.AGE	DM.AGE
8	ADSL	7	AGEU	Age Units	Char	10		AGEU	Predecessor	DM.AGEU	DM.AGEU
9	ADSL	8	AGEGR1	Pooled Age Group 1	Char	20		AGEGR1	Derived	if <ADSL.AGE<65 then ADSL.AGEGR1=<65', else if ADSL.AGE=>=65 then ADSL.AGEGR1=>=65'	如果<ADSL.AGE<65则ADSL.AGEGR1=<65', 否则如果ADSL.AGE=>=65则ADSL.AGEGR1=>=65'
10	ADSL	9	AGEGR1N	Pooled Age Group 1 (N)	Integer	8		AGEGR1N	Assigned	the numeric value for ADSL.AGEGR1 based on Codelist	基于ADSL.AGEGR1受控术语列表的数值型赋值
11	ADSL	10	SEX	Sex	Char	2		SEX	Predecessor	DM.SEX	DM.SEX
12	ADSL	11	SEXN	Sex (N)	Integer	8		SEXN	Assigned	the numeric value for ADSL.SEX based on	基于ADSL.SEX受控术语列表的数值型赋值

Figure 1 Global level specification

Second, certain specifications will be copied to a compound level folder, and compound level variables as well as a codelist which could be used in each study of the same compound will be managed centralized at the compound level. When the same variable has a different derivation rule across studies, the programmer can tailor the logic by adding a new row in the compound level specification. Like below, the

compound level specification may have multiple rows for one variable and the rule used in the certain study may flag as Y as shown in the last columns.

	A	B	C	D	E	F	G	H	I	J	K	R	S	T
1	Dataset	Order	Variable	Label	Data Type	Length	Format	Codelist	Origin	Comment	Comment_zh	Study1	Study2	Study3
2	ADSL	86	DCUTDT	Data Cutoff Date	Integer	8	yyymmdd10.		Assigned	assign to "2021-01-01"	赋值为"2021-01-01"转换成数值型日期格式	Y		
3	ADSL	86	DCUTDT	Data Cutoff Date	Integer	8	yyymmdd10.		Assigned	assign to "2022-03-07"	赋值为"2022-03-07"转换成数值型日期格式		Y	
4	ADSL	86	DCUTDT	Data Cutoff Date	Integer	8	yyymmdd10.		Assigned	assign to "2022-05-07"	赋值为"2022-05-07"转换成数值型日期格式			Y
5														

Figure 2 Compound level spec

Third, to generate study level specifications and saved it in the specific study folder which could be read and used to generate the ADaM dataset directly. The study level spec will have only one variable with the derivation rule specifically for the study.

	A	B	C	D	E	F	G	H	I	J	K
1	Dataset	Order	Variable	Label	Data Type	Length	Format	Codelist	Origin	Comment	Comment_zh
2	ADSL	86	DCUTDT	Data Cutoff Date	Integer	8	yyymmdd10.		Assigned	assign to "2021-01-01"	赋值为"2021-01-01"转换成数值型日期格式

Figure 3 Study level spec

All the above functions can easily be manipulated using a user-friendly interface called 'ADaM Spec Generator' created by Python without editing excel, plus the access control design, the whole process increase the efficiency by reducing redundant work and decreasing the possibility of overwritten others' work.

'ADAM SPEC GENERATOR' INTERFACE

This interface allows the user to create a new study/task under a certain compound, then under each study/task user can work on the detail of each specification. All the possible modifications like selecting variables, setting core variables, revising derivation rules, set codelist for certain variables can be easily accomplished by simply 'clicking'. After the compound level specification is done, the user can create a perfect study level specification also just by 'clicking'.

The interface can also generate a format dataset that could be used when coding.

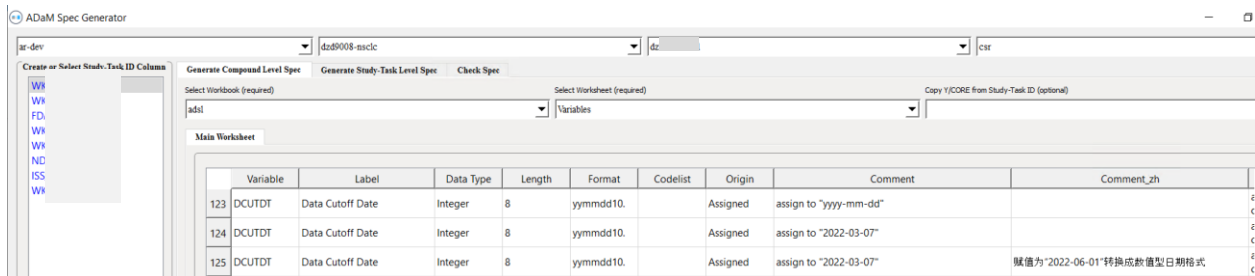


Figure 4 ADaM spec generator interface

CENTRALIZED COMPANY WISE ADAM SPECIFICATION

The centralized company-wise ADaM specification was maintained by ADaM specification expert who was responsible for 1) generating a global ADaM specification template, if certain is requested and - not in current ADaM specification library then it is the ADaM specification expert's responsibility to create a global meaning company wise ADaM specification for study to begin with. By doing so, it could ensure the content of each specification used in different compounds and different studies to follow the same format and standards like clearly stated in ADaM IG.

On the other hand, the centralized company-wise ADaM specification consist of both derivation rule in the Chinese version and the English version, after loading in the 'ADaM Spec Generator' Interface, the user can see both columns.

The design of the company-wise ADaM specification also takes creating define.xml efficiently into consideration, for example, by specifying 'Method Type' for each variable whose origin is 'Derived'.

CENTRALIZED COMPOUND-WISE ADAM SPECIFICATION

The compound level ADaM specification was all generated and maintained in the 'ADaM Spec Generator' Interface.

First, the user could create a new study/task in the interface, then the tool will automatically add a new column at the end of the certain specification say ADSL for the programmer to work specifically on his/her own study/task.

Second, when the programmer opens one dataset specification, he/she could modify it directly in the interface without opening an excel file. He/she could select the variables from the global template specification that he/she would like to be included in his/her study/task by simply clicking the cell in the 'study/task' column, 'Y' means to keep this variable, 'CORE' means to set this variable as a core variable, while null means to drop this variable. If your study needs new variables which were not in the global specification, you can simply insert a new row and edit per your needs. Here we also designed a recycle function which is a wonderful and smart idea to make the whole process clever. The recycle function could collect variables that were newly created by the user but not in the global specification and report a summary. The summary will be reviewed periodically and cases like different study teams may add-similar variables that could be standardized happen, then the global ADaM specification maintenance team will consider adding this commonly used variable into the global ADaM specification and release a new version for later use. As of now, then you have all the variables you need, then you may work on the detail like derivation rule and set codelist. Now you will realize you only need to take extra effort on the derivation rules that your study was never been defined before in your compound, since for simple ones like ADSL.AGE or complicated ones but share the same derivation rule across the studies within one compound like ADSL.TRTEDT, all you need to do was simply select this variable by the method stated above. In addition, we designed two derivation rule columns, one for English and one for Chinese and you could edit the two versions simultaneously. The benefits of this design are obvious. First, it allows the user to edit both the English and Chinese versions together which ensuring the consistency between two versions. Second, it allows the user to export either Chinese study level specification or English study level specification freely which further speeds up the process of generating define.xml of both language versions. The value-level tab is no difference from the variable tab.

Third, now we have desired ADaM datasets and desired variables, but certain variables still need controlled terminology. This could also be handled together with specifications in the same interface by simply clicking 'set codelist'. This action will direct the user to a specific code list from a global maintained codelist library. All you need was just select by clicking like what you do as the above steps. Likewise, the codelist feature also had recycle function and bilingual function. Plus, from the interface, you could generate a format dataset also just by clicking.

Generate Compound Level Spec Generate Study-Task Level Spec Check Spec

Select Workbook (required) Select Worksheet (required) Copy Y/CORE from Study-Task ID (optional)

advs Variables

Main Worksheet Codelist

	ADaMName	Ion_Extensibl	ID	Name	Name_zh	NCI Codelist Code	Data Type	Term	Term_zh	√CI Term Code	Decoded Value	Decoded Value_zh
1	ADVS		ADVS_PARA...	Parameter ...	生命体征参...		Text	BMI	BMI		Body Mass Index (kg/m2)	身体质量指数(kg/m2)
2	ADVS		ADVS_PARA...	Parameter ...	生命体征参...		Text	DIABP	DIABP		Diastolic Blood Pressure ...	舒张压(mmHg)
3	ADVS		ADVS_PARA...	Parameter ...	生命体征参...		Text	HEIGHT	HEIGHT		Height (cm)	身高(cm)
4	ADVS		ADVS_PARA...	Parameter ...	生命体征参...		Text	HIPCIR	HIPCIR		Hip Circumference (cm)	臀围(cm)
5	ADVS		ADVS_PARA...	Parameter ...	生命体征参...		Text	PULSE	PULSE		Pulse Rate (beats/min)	脉搏(beats/min)
6	ADVS		ADVS_PARA...	Parameter ...	生命体征参...		Text	SYSBP	SYSBP		Systolic Blood Pressure ...	收缩压(mmHg)
7	ADVS		ADVS_PARA...	Parameter ...	生命体征参...		Text	TEMP	TEMP		Temperature (°C)	体温(°C)
8	ADVS		ADVS_PARA...	Parameter ...	生命体征参...		Text	WCHCRTO	WCHCRTO		Waist to Hip Circumference...	腰臀比
9	ADVS		ADVS_PARA...	Parameter ...	生命体征参...		Text	WEIGHT	WEIGHT		Weight (kg)	体重(kg)
10	ADVS		ADVS_PARA...	Parameter ...	生命体征参...		Text	WSTCIR	WSTCIR		Waist Circumference (cm)	腰围(cm)

Figure 5 Set Codelist Interface

What if the shared variables edit by one study and then affect other studies? No worry! We designed an access control function and only the designated person usually the compound lead was able to edit shared variables.

STUDY-LEVEL ADAM SPECIFICATION

Study level ADaM specification was more like a harvest from all the previous efforts. The study level spec will keep only selected variables and selected language version and drop unneeded columns in the compound level specification. The study level ADaM specification will be saved under the study folder and used in the programming.

This page also embedded the function of 'generate format' in the study level and automatically save the format in the study folder and then in the code, coder can use the format directly like format PARAM or PARAMN using the ads format.

It can also generate the define spec from this APP.

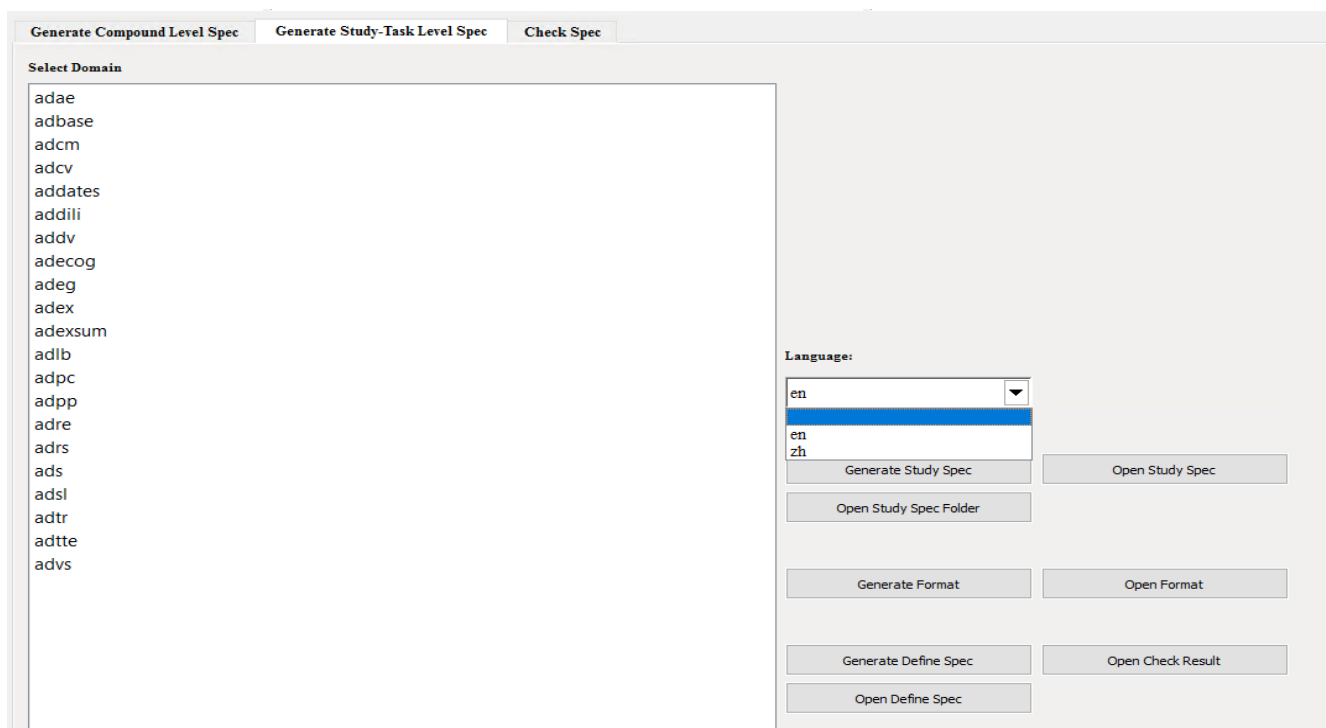


Figure 6 Generate Study-Level Spec Interface

HIGHLIGHT FEATURES

FREE SWITCH OF ENGLISH/CHINESE VERSION

To cope with the challenge of meeting the submission requirement for both US FDA and China CDE, we developed two columns in one single compound level specifications for both the English version derivation rule as well as the Chinese version. Both versions will be displayed in the ADaM spec Generator interface, the user could select one version when generating study-level specifications.

ACCESS CONTROL

Ensure consistency and avoid redundant work.

EMBEDDED CODELIST FUNCTION

Maintain codelist global wise and easy edit in the interface. Generate format datasets also in the interface.

CONCLUSION

Through this creative ADaM specification management APP, the study team can improve efficacy while ensure the accuracy when developing study specification and generating define.xml for preparing the submission package.

ACKNOWLEDGMENTS

I would like to thank Zhiping Yan for making this APP interface and Jianyong Tong, Xinyun Gao and Jiarui He at Dizal for their advice and feedback so that the process can be improved continuously.

CONTACT INFORMATION

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

Zhi Xu
Dizal Pharma
+86-15910602154
Zhi.Xu@dizalpharma.com