

Handling Non-English Clinical Trials Data in SDTM Conversions for US-FDA Submissions

Ramanan D. Bathala, TAKE Solutions Limited, Chennai, INDIA

1. ABSTRACT

Converting Legacy clinical trial data in to CDISC-SDTM (Study Data Tabulations Model) for US-FDA submission purposes is challenge by itself. If these trials were conducted outside United States and that legacy data has non-English content, it poses additional challenges of translating them to English before converting to SDTM standard, and still maintaining the data integrity. This paper presents a case study on how this has been achieved for clinical trials data which had Japanese language contents. This paper also outlines the challenges faced and solutions adapted to address those challenges. Though this case study talks about how this has been done for clinical trial data in Japanese language, but the same concepts are applicable to clinical trials data collected in other languages as well. The focus of this paper is SAS options and techniques used to; view the non-English data, extract such non-English text from the data, insert the translated English text in respective observation and corresponding variables, cost effective use of translation services and finally how to maintain the data integrity during the entire process.

2. INTRODUCTION

The players involved in this translation and SDTM conversion of collected (RAW) data were a CRO, Sponsor Company and a third party, language translation service provider. The problems which were to be handled are multi-dimensional including translating the protocol, CRF, CSR and most importantly; last but not least SAS data items in the submission. (**SAS Data Items** is a collective term used in this paper to represent both SAS datasets and SAS Format catalogs). The last part, "**Translating SAS Data Items with Japanese text into SAS Data Items in English**" is core topic and focus of this paper. The following flow diagram illustrates the workflow involved in translating SAS data items.

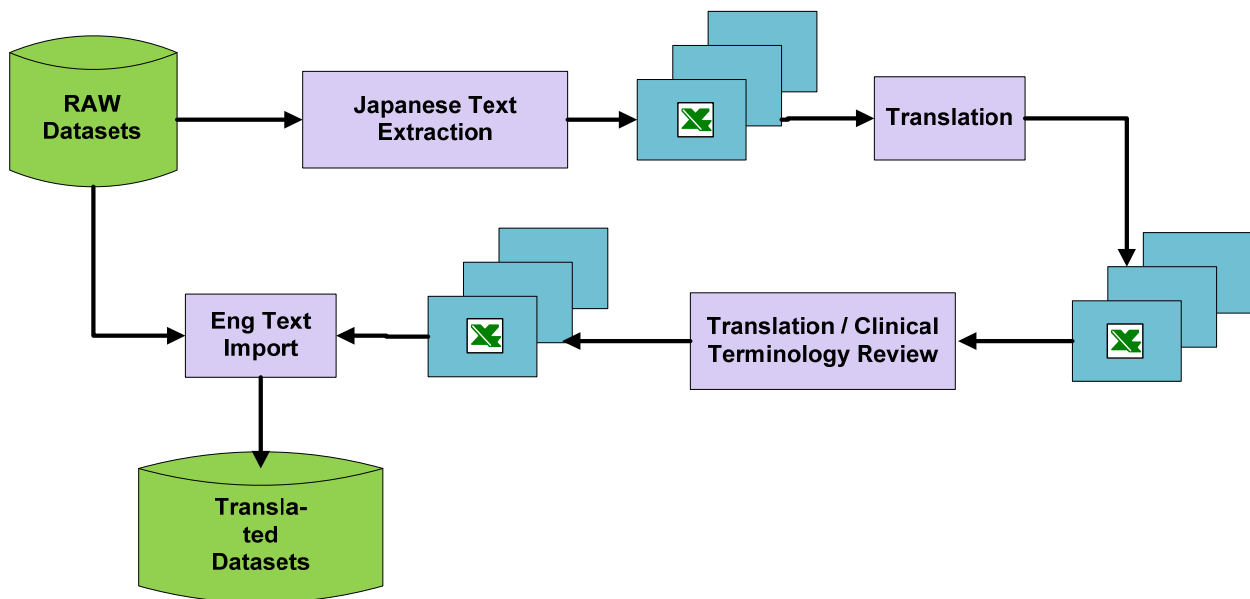


Figure 1: Translation work-flow

3. CHALLENGES AND SOLUTIONS

The following are few critical challenges we faced in translating SAS Data Items.

1. First challenge was finding out what needs to be done in SAS 9.1.3, such that one can see the Japanese contents in SAS data items. We used SAS NLS (National Language Support) options and more details on this are discussed in following sections.
2. How Japanese text data can be sent to the translation service provider, whose staff are not equipped with skills and tools to view SAS datasets, and how to insert the translated text in respective observations and corresponding variables.
3. How to extract the non-English text from SAS datasets using SAS 9.1.3. The techniques used in SAS are discussed in later sections.
4. How to minimize the use of the translation services, because typically language translation service providers charge the sponsor or CRO by the number of words/sentences/pages they translate.
5. How to maintain the data integrity during the entire process.

3.1 SAS NLS OPTIONS

We have to enable DBCS support through the following options in SAS configuration file. More information on this subject can be found in page 14 of User Installation Guide available at http://support.sas.com/documentation/installcenter/the_sas_system/9.1.3_TS1M3/win/win/user.pdf

Upon following the instructions the following NLS options should appear in the SAS configuration file.

```
-DBCS  
-DBCSTYPE PCMS  
-DBCSLANG JAPANESE
```

To view the Japanese text we should also install relevant fonts in the operating system by following the steps given below.

1. Click Start>Settings>Control Panel>Regional and Language Options.
2. Under Languages tab>Supplemental language support, select the "Install files for east Asian languages" check box.
3. Under Advanced tab>Language for Non-unicode programs, select Japanese
4. Click OK and reboot.

3.2 SENDING JAPANESE TEXT FOR TRANSLATION

The other challenge is how these Japanese text strings can be sent for translation. The translators are not likely to have enough expertise to see and work with SAS tools for viewing datasets with Japanese text data. So, we have to extract them and put them into a some commonly used tools like excel workbooks or sheets so that they can view them and translate them and return it to the CRO/Sponsor organization.

To extract the Japanese text strings from SAS Data Items a SAS program is written and executed for a given study. This SAS program reads every data item in that study and creates one excel workbook with following sheets in it.

1. Variables (All variables whose labels are in Japanese)
2. Values (All Japanese text values of all variables from all datasets)
3. Decodes (Format catalog entries whose values are in Japanese)

Variables sheet would have the following columns in it. The highlighted column (englabel) will be filled in by the translator.

dsname	varnum	varname	label	englabel
AE	6	AE_NM	有害事象名	
AE	7	AE_CD	有害事象コード	
AE	8	AE_DF	有害事象発現日	
AE	9	SEVER	重症度	
AE	10	SERIOUS	重篤性	

Figure 2: Variables Sheet

Values sheet would have the following columns in it. The highlighted column (engvalue) will be filled in by the translator.

dsname	varname	value	engvalue
AE	ACTION	減量その後増量	
AE	AE_CMT	偶発事象である。	
AE	AE_CMT	偶発的なものであり治験薬との関連なし。	
AE	AE_CMT	偶発的な事象であり、治験薬との因果関係は無い。	
AE	AE_CMT	偶発的な事象である。	
AE	AE_CMT	偶発的な事象であると判断する	
AE	AE_NM	プロラクチン上昇	
AE	AE_NM	過鎮静	
AE	AE_NM	寒気	
AE	AE_NM	感冒	
AE	AE_NM	肝機能異常	
AE	AE_NM	疑惑	
AE	AE_NM	急性上気道炎	
AE	AE_NM	筋強剛	

Figure 3: Values Sheet

Decodes sheet will have the following columns in it. The highlighted column (engdecode) will be filled in by the translator.

fmtname	code	japdecode	engdecode
AAA001S	1	なし	
AAA001S	2	あり	
AAA002S	1	なし	
AAA002S	2	あり	
AAA002S	8	不明	

Figure 4: Decodes Sheet

3.3 EXTRACTING JAPANESE TEXT

The other challenge we faced is how to identify Japanese text in SAS Data and how those can be extracted from them. Because of the fact that Japanese characters are symbols and there are many of them it is almost impossible for us to as non-Japanese professionals we could not come up with the finite number of Japanese characters to look for in the data. Rather it would be easier to identify come up with the finite list of non-Japanese list of characters in data. Hence the following lists of possible non-Japanese character strings are formed.

1. English Alphabet Characters(A to Z and a to z), which are stored in macro variable **alpchar**
2. Special Characters like (~!@#\$\$%^&*()),./' "[]-=<>?:{ }_+;),|\u\y which are stored in macro variable **sp1char**

3. Numeric Characters(0 to 9) which are stored in macro variable **numchar**
4. Miscellaneous double byte characters found in Values(. 0 1 2 3 4 5 6 7 8 9 ·×) which are stored in macro variable **mischar**

With these char list is defined, we can use a SAS function **KVERIFY** to help us identify if the value has any other character other than the ones in these four strings. The KVERIFY function returns the position of the first character in source that is not present in any of the supplied arguments. If KVERIFY finds that every character in source is in at least one excerpt, it returns a 0.

The **kverify** function used in the code as

```
kverify(value, "&alpchar", "&splchar", "&numchar", "&mischar") > 0
```

This piece of code yields a true condition if any Japanese character is present in the value.

3.4 MINIMIZING THE TRANSLATION COST

Since, typically language translation service providers charge the sponsor or CRO by the number of words/ sentences/ pages they translate, we have to minimize the translation cost by only sending the unique Japanese text strings for translation. We can easily do it for the values of variables. Since the variable values contain majority of the Japanese text strings if we can identify the unique strings here per variable per dataset, it would significantly reduce the translation cost. But, this comes with its own complexity of updating the English values for the corresponding Japanese text strings wherever they appear in a particular variable in that dataset. But this complexity was handled by writing a SAS macro code one time and the same was reused for many other studies.

3.5 MAINTAINING DATA INTEGRITY

Maintaining the data integrity is most important in this process. We have to ensure that the rows and columns are not deleted in the excel sheets that were exported for translation. For this purpose we can use excel cell locking feature. Apart from locking we have to make sure that the QC process has check list with the following.

1. Number of rows and columns submitted in each excel sheet for translation.
2. Number of rows and columns received from translation service provider in each sheet should match the counts in item 1.
3. After the successful import of the sheets and English text replacement, we need to also make sure that the observations and variable counts are matching between the original datasets and the translated datasets.
4. We should also preserve the sort order during the entire process and make sure that the observations in translated datasets are retained in sequence as they were in the original datasets.

4. CONCLUSION

Now that the datasets and format catalogs have all translated English text one can go ahead with the CDISC SDTM standardization of all the RAW datasets as per the process that CRO follows.

5. REFERENCES

http://support.sas.com/documentation/installcenter/the_sas_system/9.1.3_TS1M3/win/win/user.pdf

6. CONTACT INFORMATION

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

Name : Ramanan D. Bathala
Enterprise : TAKE Solutions Limited
Address : 80/81 MBC Towers, 6th Floor, TTK Road
City, State ZIP : Chennai, INDIA 600 118.
Work Phone : +91-44-6696 4306
Fax : +91-44-4216 9348
E-mail : ramanan.dayalan@takesolutions.com
Web : www.takesolutions.com

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.