

# 申請時電子データのLengthにまつわる 実務上の問題点と対処法

○片山 雅仁、平野 勝也、鈴木 ゆら  
(塩野義製薬株式会社 解析センター)

## Business Issues for Variable Length and a Solution on Electronic Submission of Study Data

Masahito Katayama, Katsunari Hirano, Yura Suzuki  
Biostatistics Dept., Shionogi & Co., Ltd.

## 要旨:

PMDAの申請時電子データ提出に向けて、変数のLengthについて適切な値であるかどうかを確認するルールが存在する。あまり本質的でないため重視されていないが、本発表ではこの問題への解決策を紹介する。

キーワード: CDISC, SDTM, PMDA, FDA, 日米同時申請, バリデーション



- Disclaimer

- 本発表の内容は検討中のアイデアであり、所属組織で運用している手順ではありません。



- 目次
  - 背景
  - Length問題とは？
  - Pinnacle21のバリデーションルールについて
  - いつ, Length問題を解決するか？
  - Length問題を解決する場合の問題点
  - 問題点への解決策
  - SASプログラム紹介

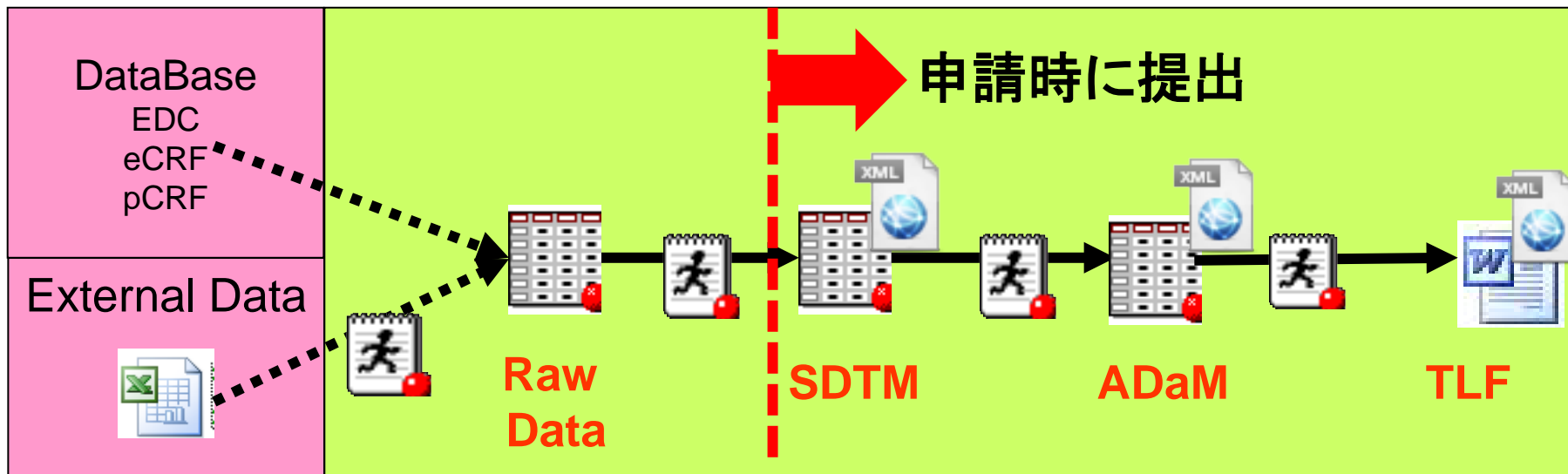


- 目次
  - 背景
  - Length問題とは？
  - Pinnacle21のバリデーションルールについて
  - いつ, Length問題を解決するか？
  - Length問題を解決する場合の問題点
  - 問題点への解決策
  - SASプログラム紹介

- 背景

- 2014年6月20日 PMDA「基本通知」発出
- 2015年4月27日 PMDA「実務通知」発出
- 2015年5月28日・6月3日:「承認申請時の電子データ提出に関する実務的事項について(通知)」等に基づく運用と技術的事項に関する説明会 (PMDA主催)
- 2015年9月28日:申請時電子データ提出にかかる実務担当者のためのワークショップ (製薬協・日本CRO協会・PMDA共催)
- 2015年11月18日 PMDAバリデーションルール一覧

## データ収集から申請までの流れ



## PMDAへ提出が必要な成果物

SDTMデータ (xpt形式) +

- Define.xml (SDTMのメタデータ)
- Annotated CRF
- SDRG (Study Data Reviewer's Guide)

ADaMデータ (xpt形式) +

- Define.xml (ADaMのメタデータ)
- ADRG (Analysis Data Reviewer's Guide)
- ADaM作成プログラム

解析資料

- (TLF: Table, Listing and Figure) +
- 帳票作成プログラム
  - Define.xml (Analysis Result Metadata)



- 目次

- 背景

- Length問題とは？

- Pinnacle21のバリデーションルールについて

- いつ, Length問題を解決するか？

- Length問題を解決する場合の問題点

- 問題点への解決策

- SASプログラム紹介



## Length問題とは？

– Pinnacle21に実装されている[あの](#)ルール

- SDTMデータセットの変数のLengthについて、事前に規定した変数のLengthに対して実際に入力されたデータのLengthに余裕がある場合に、Warningを出力されるルール

A	B	C	D	E	F	G	H	I	J
Domain	Record	Count	Variables	Values	Pinnacle 21 ID	Publisher ID	Message	Category	Severity
AE			Variable, Excess	AEHLGT, 40	SD1082		Variable length is too long for actual data	Metadata	Warning
AE			Variable, Excess	AEREFID, 36	SD1082		Variable length is too long for actual data	Metadata	Warning
AE			Variable, Excess	AESEV, 2	SD1082		Variable length is too long for actual data	Metadata	Warning
AE			Variable, Excess	USUBJID, 20	SD1082		Variable length is too long for actual data	Metadata	Warning
AE			Variable, Excess	AESOC, 33	SD1082		Variable length is too long for actual data	Metadata	Warning
AE			Variable, Excess	AETERM, 169	<a href="#">SD1082</a>		Variable length is too long for actual data	Metadata	Warning
CM			Variable, Excess	CMSTRT, 14	SD1082		Variable length is too long for actual data	Metadata	Warning
CM			Variable, Excess	CMSTRF, 14	SD1082		Variable length is too long for actual data	Metadata	Warning
CM			Variable, Excess	CMSPID, 36	SD1082		Variable length is too long for actual data	Metadata	Warning
CM			Variable, Excess	CMTRT, 148	SD1082		Variable length is too long for actual data	Metadata	Warning
CM			Variable, Excess	STUDYID, 27	SD1082		Variable length is too long for actual data	Metadata	Warning
CM			Variable, Excess	USUBJID, 20	SD1082		Variable length is too long for actual data	Metadata	Warning
CM			Variable, Excess	CMCAT, 12	SD1082		Variable length is too long for actual data	Metadata	Warning
CM			Variable, Excess	CMDOSU, 13	SD1082		Variable length is too long for actual data	Metadata	Warning
CM			Variable, Excess	CMSTRPT, 14	SD1082		Variable length is too long for actual data	Metadata	Warning
CM			Variable, Excess	CMREFID, 33	SD1082		Variable length is too long for actual data	Metadata	Warning

Pinnacle21の大量のWarningのスクリーンショット  
担当者にとって非常に厄介...



- 実際に行っていること
  - 例) AETERM (Length = \$200.)

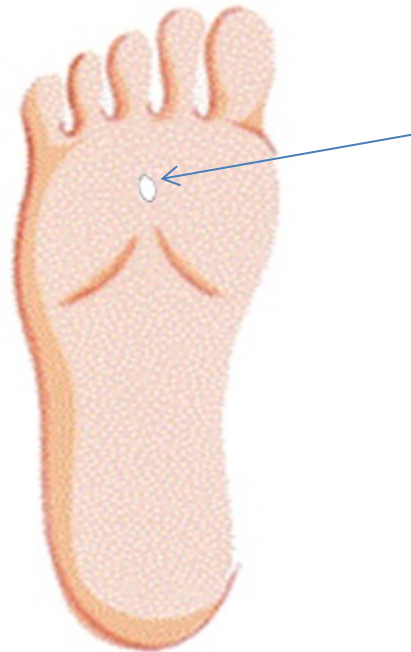
AETERM (Length = \$200.)		
Headache	<input type="text"/>	8文字
Diarrhea	<input type="text"/>	8文字
Hot Flash	<input type="text"/>	9文字

余分なスペース

各変数について、ドメイン内での最大長を変数のLengthにする必要がある!!

⇒この場合、AETERMのLengthは\$9となる

- Length問題を例えると...
  - まさに、「足の裏の米粒」...
  - 取らないと気持ち悪いけど、取っても食えない...





- 目次

- 背景
- Length問題とは？
- Pinnacle21のバリデーションルールについて
- いつ, Length問題を解決するか？
- Length問題を解決する場合の問題点
- 問題点への解決策
- SASプログラム紹介

- Pinnacle21のバリデーションルールについて
  - Pinnacle21 V2.1.1 SDTM (PMDA)

Pinnacle 21 ID	Message	Description	Severity
SD1082	Variable length is too long for actual data	Variable length should be assigned based on actual stored data to minimize file size. Datasets should be re-sized to the maximum length of actual data used prior to splitting.	Warning

PMDA 「承認申請時の電子データ提出等に関する技術的ガイド」

3.6.1 CDISC準拠データに対するバリデーション

(ア) 違反している場合には修正されるまで審査が開始されないルール

(イ) 違反している場合に、事前に説明がない場合には修正されるまで審査が開始されないルール

(ウ) 違反した場合でも必ずしもその説明を求めないルール ←本ルールはこの部分

まさに足の裏の米粒...

- ただし, FDAになると...
  - Pinnacle21 V2.1.1 SDTM (FDA)

Pinnacle 21 ID	Message	Description	Severity
SD1082	Variable length is too long for actual data	Variable length should be assigned based on actual stored data to minimize file size. Datasets should be re-sized to the maximum length of actual data used prior to splitting.	Error

なんとErrorになってしまう...

FDA Technical Conformance Guide V3.0

3.3.3 Dataset Column Length

The allotted length for each column containing character (text) data should be set to the maximum length of the variable used across all datasets in the study. This will significantly reduce file sizes. For example, if USUBJID has a maximum length of 18, the USUBJID's column size should be set to 18, not 200.

その変数については試験内の全てのドメインで使用されている変数の最大長に設定する必要がある



## (参考) OpenCDISC Validatorの時代は

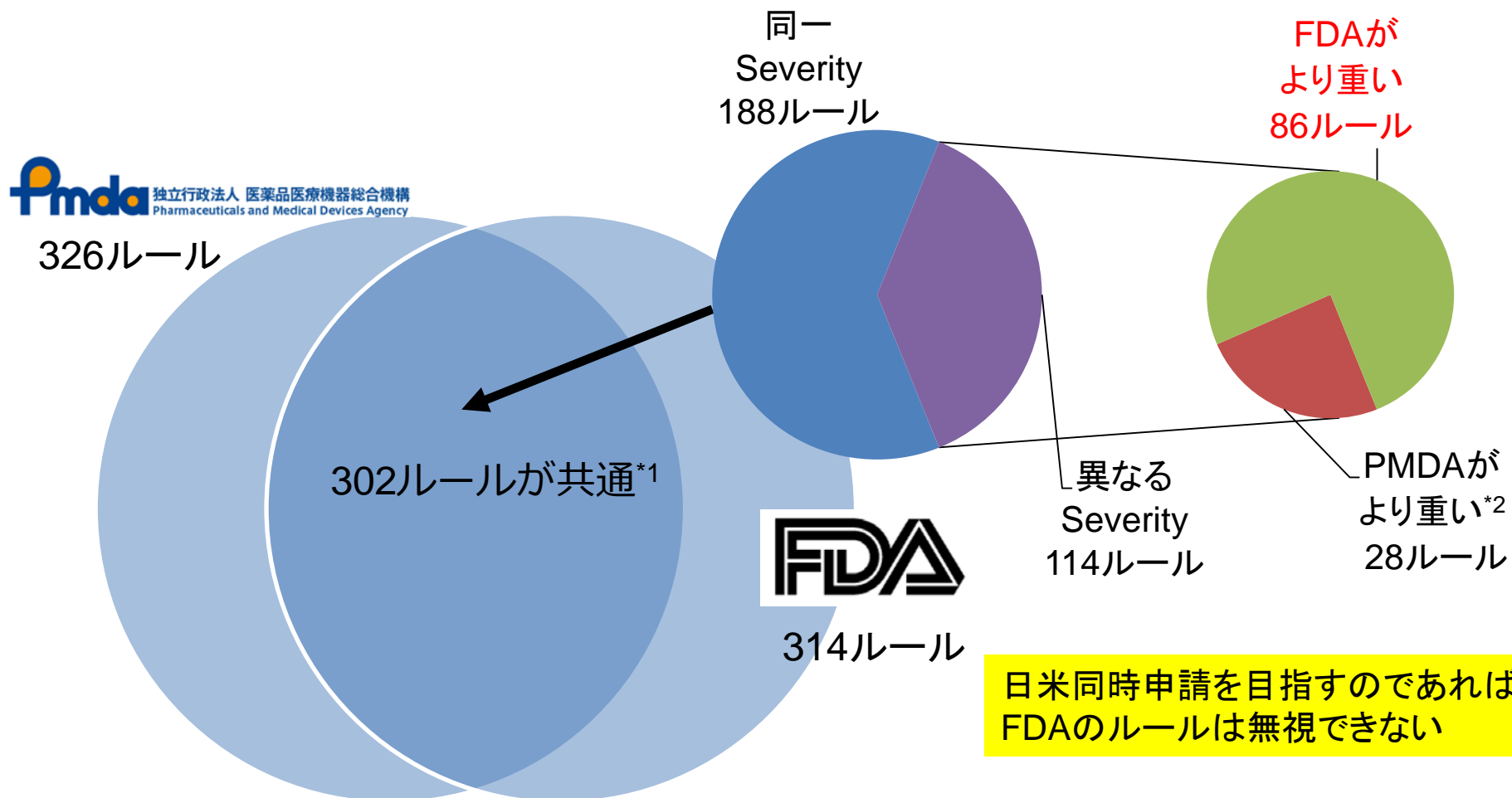
- 余分なLengthに応じて異なるSeverity

Rule ID	Message	Description	Severity
SD1080	-- variable length is too long for actual data	Variable length should be assigned based on actual stored data to avoid to minimize file size. Datasets should be resized to the maximum length used prior to splitting.	Information
SD1081	-- variable length is too long for actual data	Variable length should be assigned based on actual stored data to avoid to minimize file size. Datasets should be resized to the maximum length used prior to splitting.	Warning
SD1082	-- variable length is too long for actual data	Variable length should be assigned based on actual stored data to avoid to minimize file size. Datasets should be resized to the maximum length used prior to splitting.	Error

廃止

FDAのClinical Validation SpecificationsとOpenCDISC Community V2.0の登場とともに、SD1080とSD1081は廃止つまり、FDAの当該ルールがエラーだけになった!!

- PMDAとFDAのSDTMバリデーションルールの違いは？



\*1 : Pinnacle21のRule IDが同じだが、狭義では完全に同一でないルールもある

\*2 : より重いとする基準は、Reject > Error > Warning > Information



- 規制当局間のLengthのルール(ツール, ガイド上)の違い

	Standards	PMDA	FDA
Pinnacle21	SDTM	Warning	Error
	ADaM	ルールなし	ルールなし
	SEND	対象外	Error
技術的ガイド*1	SDTM	記載なし	記載あり
	ADaM	記載なし	記載あり
	SEND	対象外	記載あり

ADaMは  
対象外?

Pinnacle21にはADaMのLengthのルールが実装されていないが、FDAに提出する場合はTechnical Conformance Guideの記載されている!

- \*1 : PMDA. 承認申請時の電子データ提出等に関する技術的ガイド  
 FDA. Study Data Technical Conformance Guide.

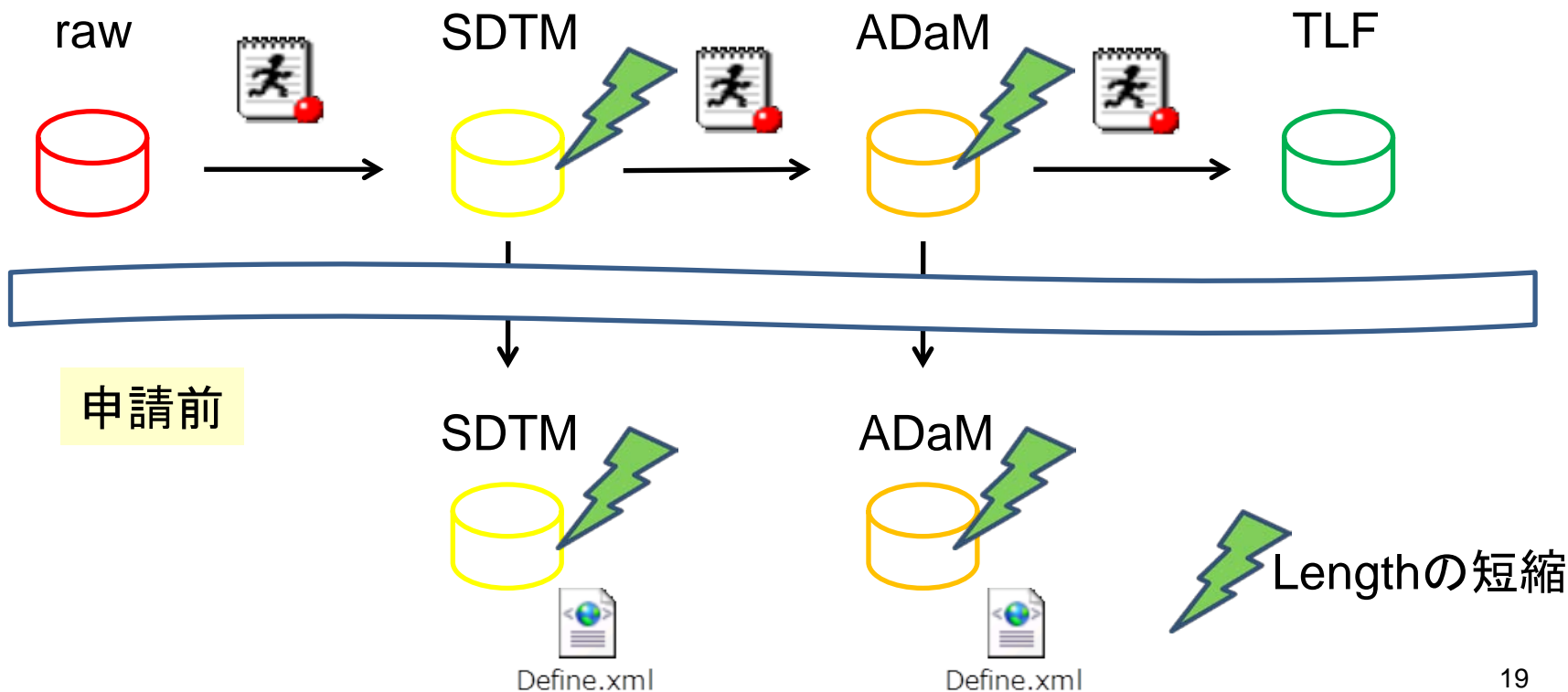


- 目次

- 背景
- Length問題とは？
- Pinnacle21のバリデーションルールについて
- **いつ, Length問題を解決するか？**
- Length問題を解決する場合の問題点
- 問題点への解決策
- SASプログラム紹介

- いつ, Length問題を解決するか?
  - CSR作成時 or 申請前

## CSR作成時





- PhUSEの事例紹介

- PhUSE Data Sizing Best Practice

- [http://www.phusewiki.org/wiki/index.php?title=Data\\_Sizing\\_Best\\_Practices\\_Recommendation](http://www.phusewiki.org/wiki/index.php?title=Data_Sizing_Best_Practices_Recommendation)

- データセットのサイズ問題

- ドメインの分割

- Length問題

- いつ, Lengthを短くすべきか？
  - PhUSE Best Practice FAQより

**Q.** At what point during data processing do variable sizes need to be reduced?

**A.** The point of variable, and ultimately, data set size reduction is to facilitate data processing upon submission. For this reason, variable re-sizing can wait until the end of the process when packaging of data for submission takes place. Prior to this point, variable sizes can be kept at the lengths necessary for operational purposes.

申請前の化粧直しで, 良いらしい!!

- 実際に解析したデータと異なるが、問題ないのか？
  - PhUSE Best Practice FAQより

**Q.** Is there an issue if the variable lengths of the datasets used to produce the analysis differ from the submitted datasets?

**A.** Depending on sponsor process, the variable lengths in the submitted files may differ from the variable lengths of the datasets used to produce analysis. For transparency the Study Data Reviewer's Guide should note the method used for processing variable lengths for submission purpose.

CSR作成時のデータと申請時のデータが異なることを、SDRGに記載することで、問題ないらしい



でも、申請前で本当に良いの？

- Length問題を申請前に対応するデメリット
  - Warning/Errorが残った状態で、CSR作成時のバリデーションを完了しなければならない
  - Define.xmlの再作成とバリデーションが必要になる
  - バリデーションツールのバージョンがCSR作成時と同一でない可能性あり(新たな問題を発見するリスク)



- 目次

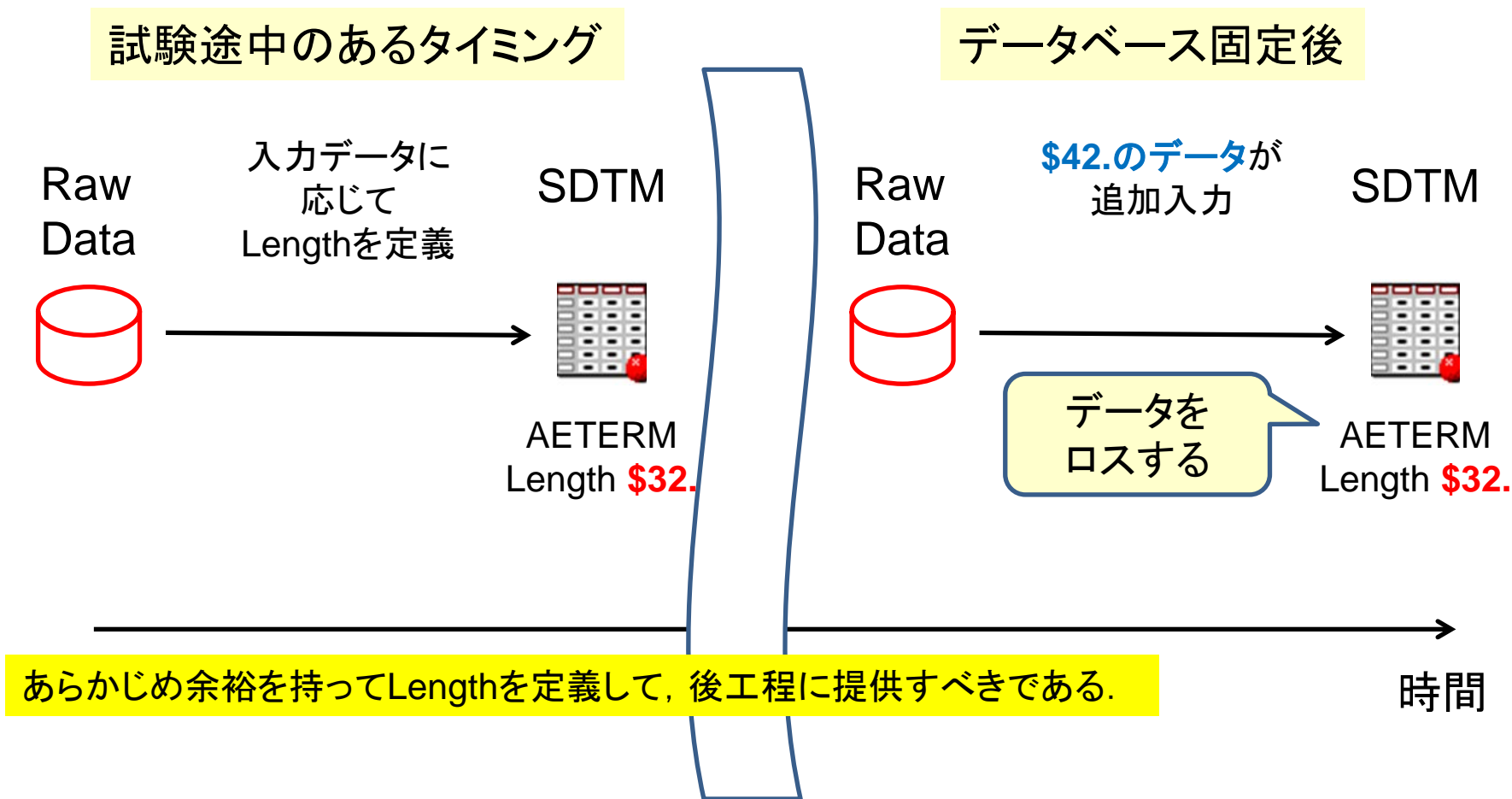
- 背景
- Length問題とは？
- Pinnacle21のバリデーションルールについて
- いつ, Length問題を解決するか？
- **Length問題を解決する場合の問題点**
- 問題点への解決策
- SASプログラム紹介



- Length問題を解決する場合の問題点
  - データをロスする可能性がある
  - 成果物の整合性確保の問題
  - 統合解析時の注意点

- Length問題を解決する場合の問題点
  - データをロスする可能性がある
  - 成果物の整合性確保の問題
  - 統合解析時の注意点

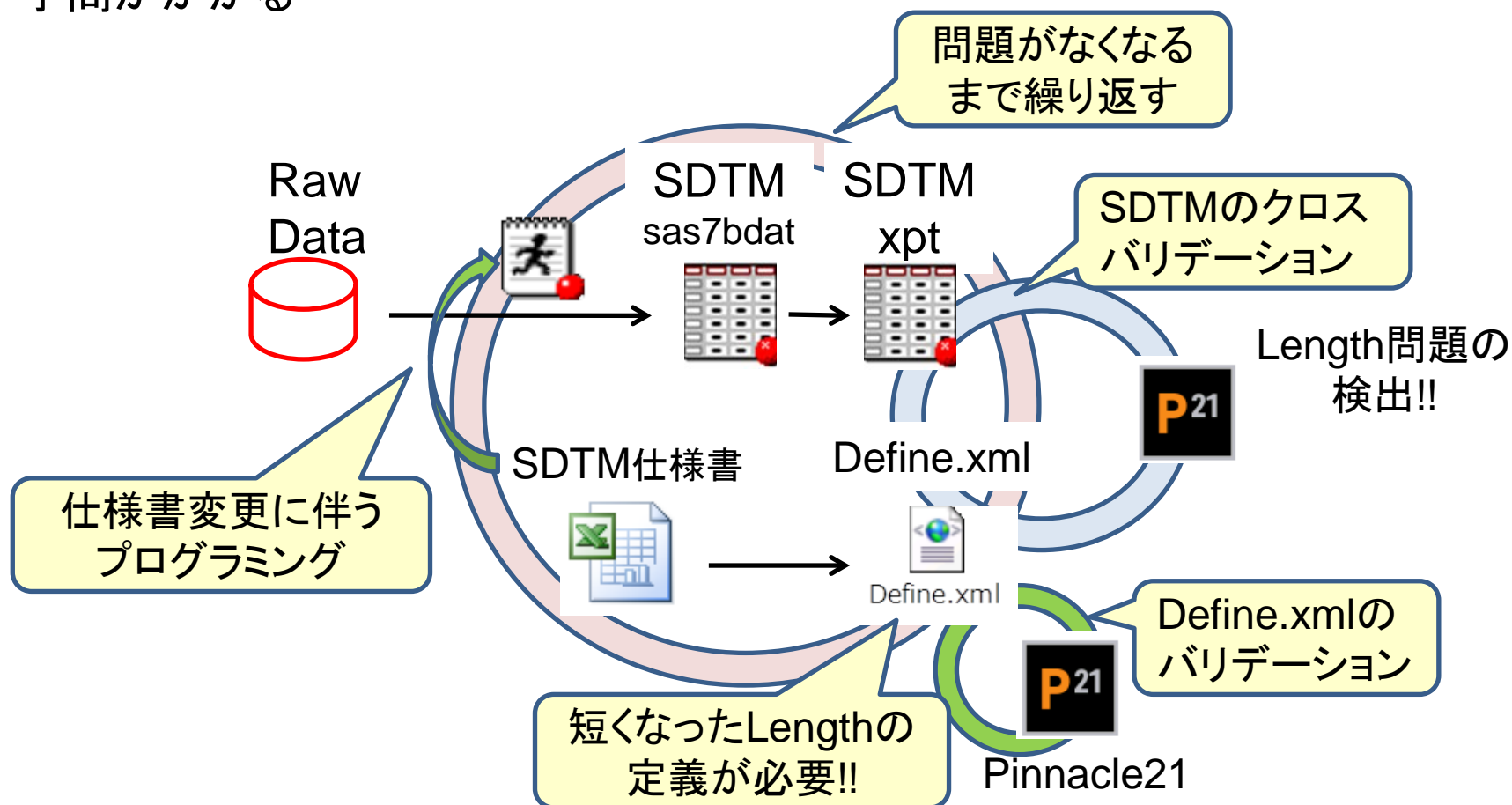
- データをロスする可能性がある



- Length問題を解決する場合の問題点
  - データをロスする可能性がある
  - 成果物の整合性確保の問題
  - 統合解析時の注意点

## ● 成果物の整合性確保の問題

仕様書, sas7bdat, xpt, Define.xmlのすべての整合性を確保しようとする  
手間がかかる



## • 仕様書とDefine.xmlの役割

	仕様書	Define.xml
用途	SASプログラマ用	規制当局用
形式	自由	xml形式
定義内容	データセット名, 変数名, ラベル, <b>Length</b> などの基本情報 プログラマ向けのルール	データセット名, 変数名, ラベル, <b>Length</b> などの基本情報 コメント, 誘導方法, 参照ドキュメントへのリンク
必要な タイミング	データセット作成時	データセット作成時 (バリデーションに利用) または申請時
Length	定義したLengthに従って sas7bdatは作成される	<b>xptの変数のLengthと整合している 必要がある</b>

仕様書とsas7bdat, Define.xmlとxptが整合するように, 作成すればよいのでは?  
(もちろん, sas7bdatとxptのLength以外の中身は同一でなければいけない)

- Length問題を解決する場合の問題点
  - データをロスする可能性
  - 成果物の整合性確保の問題
  - 統合解析時の注意点

## 統合解析時の注意点

Length調整したデータセットを解析で利用した場合

- Study 1 ADAE1
  - ADAE1.AETERM Length \$32.
  - ADAE1.AEACN \$12.
- Study 2 ADAE2
  - ADAE2.AETERM Length \$42.
  - ADAE2.AEACN \$10.
- Study 3 ADAE3
  - ADAE3.AETERM Length \$51.
  - ADAE3.AEACN \$14.



統合

データをロスする可能性。  
SASログにWarningの  
メッセージ。



ISS  
ADAE

WARNING: 入力データセットによって、変数 AETERM に複数の長さが指定されました。  
データの切り捨てが発生します。

```
DATA ADAE;  
  SET ADAE1 /* Study 1*/  
      ADAE2 /* Study 2*/  
      ADAE3 /* Study 3*/;  
RUN;
```

- ISS ADAE
  - ADAE.AETERM Length \$32.
  - ADAE.AEACN \$12.





- **目次**

- 背景
- Length問題とは？
- Pinnacle21のバリデーションルールについて
- いつ, Length問題を解決するか？
- Length問題を解決する場合の問題点
- **問題点への解決策**
- SASプログラム紹介

## 問題点への解決策

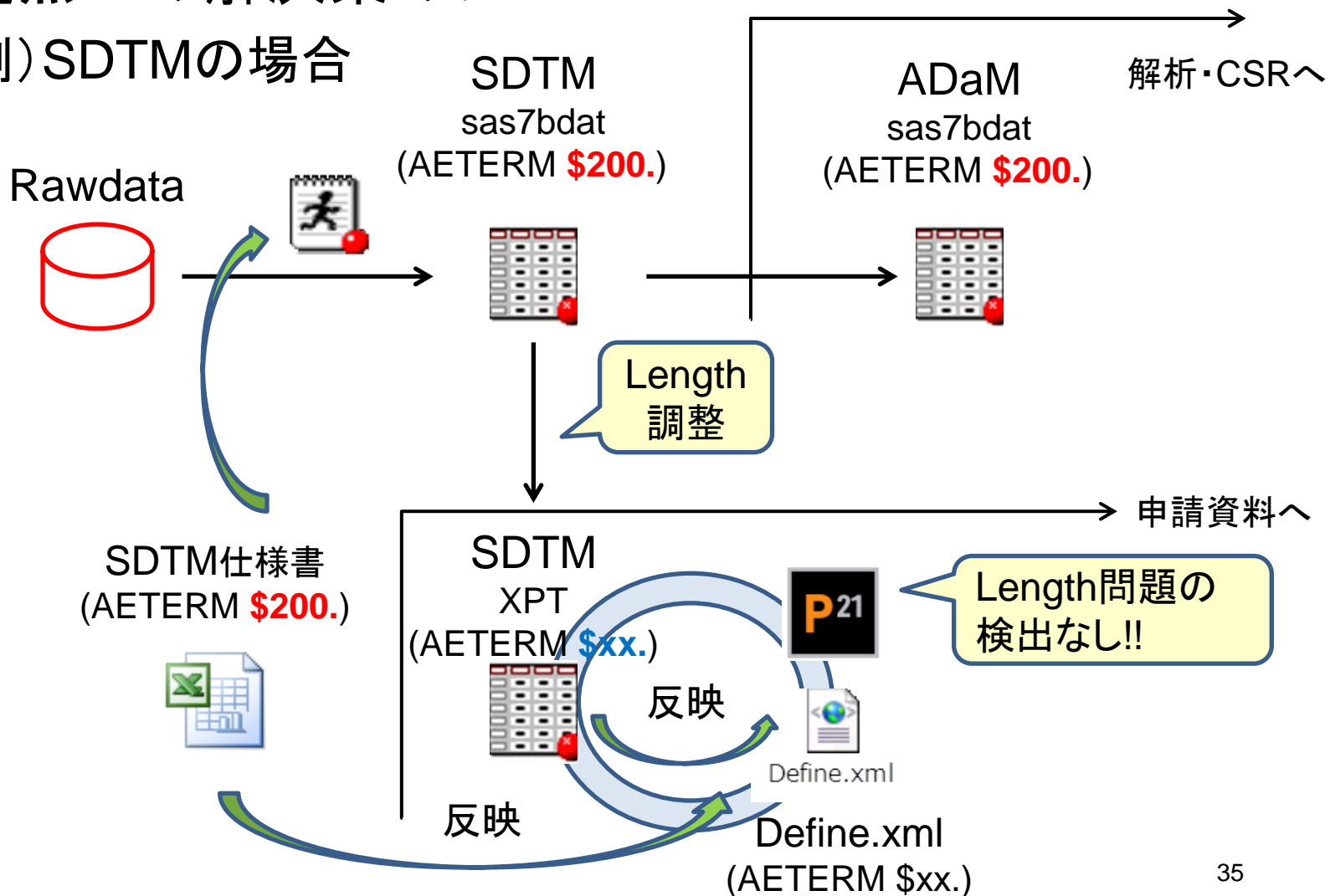
- 仕様書/sas7bdatと, Define.xml/xptを分けて考える
  - 仕様書とsas7bdatは整合させる
  - Define.xmlとxptは整合させる
- Length調整しないsas7bdatを解析で利用する
- Length調整したxptは申請資料として利用する

つまり, CSR作成時に問題なくLengthを短くすることができ, 申請資料としてのxptファイルも同時に作成できる



- 問題点への解決策のフロー

- 例) SDTMの場合



- 対応策に見る成果物ごとのLength対応

	仕様書	sas7bdat	Define.xml	xpt
用途	SASプログラマ用	CSR作成用	規制当局用	規制当局用
形式	自由	sas7bdat	xml形式	XPORT
Length	最大200 そのまま	最大200 そのまま	短くした Length	短くした Length

- この対応策で解決できること

- 仕様書のLengthはそのままで、プログラムの修正も不要！
- sas7bdat形式のデータセットのLengthをそのままにすることで、後のプロセスでのデータをロスする可能性を回避！
- xpt作成時にLengthを短くし、その情報をDefine.xmlに反映！
- Pinnacle21のバリデーションで、Warning/Errorを回避可能！！



- 目次

- 背景
- Length問題とは？
- Pinnacle21のバリデーションルールについて
- いつ, Length問題を解決するか？
- Length問題を解決する場合の問題点
- 問題点への解決策
- SASプログラム紹介

## • SASプログラム紹介

SDTM sas7bdat  
(AETERM \$200.)



この部分の  
プログラム  
を紹介

SDTM xpt  
(AETERM \$xx.)



“Actual” Lengthの算出(基本構文) - PROC SQLの利用-

\*文字型変数の場合;

```
%LET AETERM_LEN = ;
```

```
PROC SQL ;
```

```
  SELECT MAX(LENGTHN(AETERM)) INTO :AETERM_LEN FROM AE;
```

```
QUIT;
```

\* 数値型の変数はデフォルトで 8 とする;

Point

SQLプロシジャではSASの関数を使える!!

SQLのMAX関数も使える!!

コーディング量を抑えることができる。

- “Actual” Lengthの算出 (1/5) (dataset準備)

```
%macro ACTLEN(InLIB= ,FPATH = ) ;  
DATA __VARS;  
  SET SASHELP.VCOLUMN(where=(libname=upcase("&inLIB.") &  
index(MEMNAME,"_")=0)) ;  
  call symput(cats("ds" ,_n_),MEMNAME);  
  call symput(cats("var" ,_n_),NAME);  
  call symput(cats("type" ,_n_),TYPE);  
  ID=_N_;  
  KEEP LIBNAME MEMNAME NAME LENGTH TYPE FORMAT ID;  
RUN;  
PROC SQL NOPRINT;  
  CREATE TABLE __VARSX AS  
  SELECT __VARS.*,  
         VTABLE.LIBNAME, VTABLE.MEMNAME, VTABLE.MEMLABEL  
  FROM __VARS LEFT JOIN SASHELP.VTABLE  
  ON __VARS.LIBNAME=VTABLE.LIBNAME AND  
  __VARS.MEMNAME=VTABLE.MEMNAME  
  ORDER BY ID ;  
QUIT;
```

SASHELPの  
VCOLUMN,  
VTABLEより  
データセット情報を  
取得

## • “Actual” Lengthの算出 (2/5) (最大長を取得)

```

/* Counting number of variables in the datasets */
PROC SQL NOPRINT;
  SELECT COUNT (*) INTO :NVAR$ FROM __VARSX;
QUIT;
/* Computing max. length of each variable in the dataset */
%DO i = 1 %TO &NVAR$;
  %IF( %UPCASE ("&&TYPE&i..")= "CHAR") %THEN %DO;
    PROC SQL NOPRINT;
      SELECT MAX (LENGTHN(&&var&i..)) INTO :mlen&i.
    FROM %trim(&inLIB.)%trim(&&ds&i.);
    QUIT;
  %END; %ELSE %DO;
    %LET mlen&i.= 8;
  %END;
%END;

```

前述の  
基本構文を利用

数値型の変数はデフォルトで 8 とする

ループの中身

ID	&&ds&i..	&&var&i..	&&type&i..
1	AE	STUDYID	Char
2	AE	USUBJID	Char



## • “Actual” Lengthの算出 (3/5) (dataset間での調整)

```

/* Specifying maximum length into contents datasets */
DATA __VARSEX; SET __VARSEX;
  %DO i = 1 %TO &NVARX.;
    IF (ID = &i.) THEN DO; MAXLEN= &&mlen&i.; END;
  %END;
RUN;

/* Computing maximum length of the variable used across all datasets in inLIB */
PROC SORT DATA = __VARSEX; BY NAME DESCENDING MAXLEN; RUN;
DATA __VARSEX; SET __VARSEX;
  RETAIN ACTLEN;
  BY NAME DESCENDING MAXLEN;
  IF ( FIRST.NAME= 1) THEN ACTLEN= MAXLEN;
RUN;
PROC SORT DATA = __VARSEX; BY MEMNAME ID; RUN;
    
```

各変数の最大長を定義する

FDAのTechnical Conformance Guideに記載されているとおり、試験内の全てのドメインで使用されている変数の最大長に設定されるように、同じ変数内で並べ替えて調整する

MEMNAME	NAME	MAXLEN	ACTLEN
LB	VISIT	15	15
VS	VISIT	7	15

## • “Actual” Lengthの算出 (4/5) (dataset出力)

```
DATA _NULL_; SET __VARSEX;  
  BY MEMNAME ID;  
  IF (FIRST.MEMNAME=1) THEN DO;  
call execute("LIBNAME TEMPXPT XPORT "|| "&FPATH." ||strip(MEMNAME)||".xpt;");  
call execute("PROC SQL NOPRINT;");  
call execute("CREATE TABLE "|| "tempxpt."||  
strip(MEMNAME)||"(LABEL="||strip(MEMLABEL)||") AS");  
call execute("SELECT");  
  END;  
  IF (ACTLEN > 0) THEN DO;  
    IF (LAST.MEMNAME=1) THEN DO;  
      call execute( strip(NAME)||" LENGTH="||strip(put(ACTLEN,BEST.)) );  
    END; ELSE DO;  
      call execute( strip(NAME)||" LENGTH="||strip(put(ACTLEN,BEST.))||"," );  
    END;  
  END;  
END;
```

XPTファイル出力のため  
LIBNAMEを指定

1つのデータステップで、  
ドメイン内の全てのデータセット  
を最大長に修正するよう  
PROC SQLを利用して定義

変数が空白  
だったとき、  
作成されない

## • “Actual” Lengthの算出 (5/5) (dataset出力)

```
IF (LAST.MEMNAME=1) THEN DO;  
  call execute("FROM "|| strip(LIBNAME) || "."|| strip(MEMNAME) ||");  
  call execute("QUIT;");  
END;  
RUN;  
PROC DATASETS LIBRARY= WORK; DELETE __.; RUN; QUIT;  
%mend;  
%ACTLEN(InLIB=SDTM ,FPATH=%str(X:¥XXXXXX¥XXXXXX¥XXXXXX¥XXXXXX¥XXXXXX¥));
```

後半のCALL  
EXECUTEで  
生成される  
PROC SQLの構文

```
1 + PROC SQL NOPRINT;  
2 + CREATE TABLE TEMPXPT.DM(LABEL="Demographics") AS  
3 + SELECT  
4 + STUDYID LENGTH=13,  
5 + DOMAIN LENGTH=2,  
6 + USUBJID LENGTH=20,  
23 + ACTARMCD LENGTH=8,  
24 + ACTARM LENGTH=15,  
25 + COUNTRY LENGTH=3  
26 + FROM DM;
```

Actual Length!!

- まとめ

- Length問題は問題として大きくないが実務上問題になる
- PMDAとFDA間でルールに対するスタンスが異なる
- SDTMに関しては現状対応が必要(ADaMは?)
- sas7bdat形式のファイルとxptファイルの用途をすみ分け、前述のSASマクロをプロセスに挟むことによって、CSR作成時にも柔軟に対応することができる

## 参考資料

1. FDA. Study Data Technical Conformance Guide.; March 2016.  
Available from: URL : <http://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM384744.pdf>
2. PMDA. 承認申請時の電子データ提出等に関する技術的ガイド .;平成27年4月27日.  
Available from: URL : <http://www.pmda.go.jp/files/000204728.pdf>
3. PMDA. 「承認申請時の電子データ提出等に関する実務的事項について」の質疑応答集(Q&A)について .;平成27年4月27日.  
Available from: URL : <http://www.pmda.go.jp/files/000204727.pdf>
4. PhUSE Data Sizing Best Practice.; 2013.  
Available from: URL:[http://www.phusewiki.org/wiki/index.php?title=Data\\_Sizing\\_Best\\_Practices\\_Recommendation](http://www.phusewiki.org/wiki/index.php?title=Data_Sizing_Best_Practices_Recommendation)
5. 神谷亜香里, 坂井絵理, 惟高裕一, 北西由武, 角谷伸一, 小坂明子. PMDAへの承認申請時CDISC標準電子データ提出に向けた社内標準のリモデリング. SASユーザー総会. 東京. 2014.
6. 坂井絵理. もしも, とある開発品目Xを申請電子データ提出するとしたら.... ナニワデータサイエンス研究会. 大阪. 2016.
7. 高浪洋平. もしも, 日米で申請電子データ提出するとしたら.... ナニワデータサイエンス研究会. 大阪. 2016.



End of File