

# SDTMデータセットに「日本語データ」を含む場合のDefine-XML作成方法

CDISC技術チーム(エーザイ株式会社)

高島 秀也

# Disclaimer

- 本資料の内容は、CDISC技術チームの見解に基づくものであり、チームメンバーが所属する会社の公式見解を示すものではありません。
- 個々の申請で発生する特殊ケースについては、規制当局からの公式通知、及びPMDAとの相談にてご確認ください。

# 本発表の内容

○「承認申請時の電子データ提出等に関する技術的ガイドについて（※）」の別紙3に対応するDefine-XMLの作成方法を紹介

例1: 日本語の医師記載用語を英訳せずそのまま格納する場合

例2: 原文の日本語を英訳することが困難であり、使用する英語文字列を相互に区別する必要がある場合

※ 以降は技術的ガイドとする

## ○前提

・英数字データセットに対応するDefine-XMLを対象とする

・例示に使用したDefine-XMLのバージョンは「v2.0」とする

スタイルシート、例示のXMLコードは、CDISCで公開されているDefine-XMLの仕様に基づく

例1: 日本語の医師記載用語を英訳せず  
そのまま格納する場合

# 技術的ガイドに記載されている要件

## ・対象のSDTMデータセット

AE ドメイン (英数字データセット)

- 日本語でデータ収集したテキスト項目については英語文字列 (「JAPANESE TEXT IN SOURCE DATABASE」等) を格納する (AETERM 等を日本語で収集)。

Row	STUDYID	DOMAIN	USUBJID	AESEQ	AETERM	AESTDTC	AEENDTC	AEMODIFY	AEDECOD
1	ABC123	AE	123101	1	JAPANESE TEXT IN SOURCE DATABASE	2005-10-12	2005-10-12		Headache
2	ABC123	AE	123101	2	JAPANESE TEXT IN SOURCE DATABASE	2005-10-13T13:05	2005-10-13T19:00		Back pain
3	ABC123	AE	123101	3	JAPANESE TEXT IN SOURCE DATABASE	2005-10-21			Pulmonary embolism

Row	AEBODSYS	AEEV	AESER	AEACN	AEREL
1 (cont)	Nervous system disorders	SEVERE	N	NOT APPLICABLE	DEFINITELY NOT RELATED
2 (cont)	Musculoskeletal and connective tissue disorders	MODERATE	N	DOSE REDUCED	PROBABLY RELATED
3 (cont)	Vascular disorders	MODERATE	Y	DOSE REDUCED	PROBABLY NOT RELATED

Row	AEOUT	AESCONG	AESDISAB	AESDTH	AESHOSP	AESLIFE	AESMIE	AESTDY	AEENDY	AENRF
1 (cont)	RECOVERED/RESOLVED							-1	-1	
2 (cont)	RECOVERED/RESOLVED							1	1	
3 (cont)	RECOVERING/RESOLVING				Y	Y		9		AFTER

(技術的ガイド「別紙3. 日本語データ格納方法の例 例1」より抜粋)

## ・考慮すべき要件

英語文字列 (「JAPANESE TEXT IN SOURCE DATABASE」等) は1試験では一貫して同一のものを扱い、データガイド又はデータセットの定義書 (define.xml) に明示すること。

(技術的ガイド「4.1.5 日本語データの取り扱いについて」より抜粋)

# Define-XML作成のアプローチ

- 「*JAPANESE TEXT IN SOURCE DATABASE*」は、フリーテキストとして入力された日本語文字列を、一律同じ英語文字列に変換したもの
  - 英語文字列をデータセットの定義書 (Define-XML) に明示する場合、Define-XMLのcomment definitionsを用いて明示することができる
- 英語文字列に対する説明を明示する方法
  - 方法1: ItemGroupDefに定義する
    - Datasets ListのDocumentationに反映される
  - 方法2: ItemDefに定義する
    - Variables ListのDerivation/Commentに反映される

# Define-XMLの例示

## - 方法1: ItemGroupDefに定義する

### Tabulation Datasets for Study CDISC01 (SDTM-IG 3.1.2)

Dataset	Description	Class	Structure	Purpose	Keys	Location	Documentation
AE	<a href="#">Adverse Events</a>	EVENTS	One record per adverse event per subject	Tabulation	STUDYID, USUBJID, AEDECOD, AESTDTC	<a href="#">ae.xpt</a>	Original text of AETERM was converted to 'JAPANESE TEXT IN SOURCE DATABASE'.
DS	<a href="#">Disposition</a>	EVENTS	One record per disposition status or protocol milestone per subject	Tabulation	STUDYID, USUBJID, DSSTDY, DSSTDTC, DSCAT, DSDECOD	<a href="#">ds.xpt</a>	

※日本語の挿入も可能

<XMLコード>

```
<ItemGroupDef OID="IG.AE"
```

```
⋮
```

```
  def:Class="EVENTS
```

```
  def:CommentOID="COM.DOMAIN.AE"
```

```
⋮
```

```
</ItemGroupDef>
```

```
⋮
```

```
<def:CommentDef OID="COM.DOMAIN.AE">
```

```
  <Description>
```

```
    <TranslatedText xml:lang="en">Original text of AETERM was converted to 'JAPANESE TEXT  
    IN SOURCE DATABASE'. </TranslatedText>
```

```
  </Description>
```

```
</def:CommentDef>
```

ItemGroupDef Definitions  
Section (OID = "IG.AE")

Comments Definitions  
Section

# Define-XMLの例示

## - 方法2: ItemDefに定義する

### Adverse Events (AE) [Location: [ae.xpt](#)]

Variable	Label	Key	Type	Length	Controlled Terms or Format	Origin	Derivation/Comment
STUDYID	Study Identifier	1	text	7		Protocol	
DOMAIN	Domain Abbreviation		text	2	["AE" = "Adverse Events"] <Domain Abbreviation (AE)>	Assigned	
USUBJID	Unique Subject Identifier	2	text	14		Derived	Concatenation of STUDYID and SUBJID
AESEQ	Sequence Number		integer	1		Derived	Sequential number identifying records within each USUBJID in the domain.
AESPID	Sponsor-Defined Identifier		text	4		CRF Page <a href="#">21</a>	
AETERM	Reported Term for the Adverse Event		text	32		CRF Page <a href="#">21</a>	Original text was converted to 'JAPANESE TEXT IN SOURCE DATABASE'.

※日本語の挿入も可能

### <XMLコード>

```
⋮
<ItemDef OID="IT.AE.AETERM" Name="AETERM" DataType="text" Length="25" SASFieldName="AETERM"
  def:CommentOID="COM.AETERM"
⋮
</ItemDef>
⋮
<def:CommentDef OID="COM.AETERM">
  <Description>
    <TranslatedText xml:lang="en">Original text was converted to 'JAPANESE TEXT IN SOURCE DATABASE'.
  </TranslatedText>
  </Description>
</def:CommentDef>
```

ItemDef Definitions Section

Comments Definitions Section



## 例1のまとめ:

日本語の医師記載用語を英訳せずそのまま格納する場合

- Define-XMLで日本語データについて説明する場合は、Define-XMLのcomment definitionsを用いて定義する
  - ItemGroupDef:
    - データセットレベルで日本語データの有無がわかる
      - ※変数レベルで日本語データの有無については特定されないため、ItemGroupDefを使用する場合は、対象の変数を記載する
  - ItemDef:
    - 変数レベルで日本語データの有無がわかる
- データガイドで英語文字列を明示することも可能

例2: 原文の日本語を英訳することが困難であるが、使用する英語文字列を相互に区別する必要がある場合

# 技術的ガイドに記載されている要件

## ・対象のSDTMデータセット

Row	STUDYID	DOMAIN	USUBJID	QSSEQ	QSTESTCD	QSTEST
1	ABC123	QS	123101	1	QSTEST1	JAPANESE TEXT IN SOURCE DATABASE01
2	ABC123	QS	123101	2	QSTEST2	JAPANESE TEXT IN SOURCE DATABASE02
3	ABC123	QS	123101	3	QSTEST3	JAPANESE TEXT IN SOURCE DATABASE03

(技術的ガイド「別紙3. 日本語データ格納方法の例 例2」より抜粋)

## ・考慮すべき要件

日本語を含む質問表やコードリスト等に対応する箇所には適切な英訳又は英語文字列を格納すること。なお、英語文字列を格納する場合、その英語文字列を相互に区別する必要があるならば、末尾に番号をつけるなど個別に適切な対応を取り(例:「JAPANESE TEXT IN SOURCE DATABASE 01」、「JAPANESE TEXT IN SOURCE DATABASE 02」)、元の日本語との対応付けをデータガイド等に示すこと。

(技術的ガイド「4.1.5 日本語データの取り扱いについて」より抜粋)

# Define-XML作成のアプローチ

QS ドメイン (英数字データセット)

Row	STUDYID	DOMAIN	USUBJID	QSSEQ	QSTESTCD	QSTEST
1	ABC123	QS	123101	1	QSTEST1	JAPANESE TEXT IN SOURCE DATABASE01
2	ABC123	QS	123101	2	QSTEST2	JAPANESE TEXT IN SOURCE DATABASE02
3	ABC123	QS	123101	3	QSTEST3	JAPANESE TEXT IN SOURCE DATABASE03

英数字文字列に変換された対象は「Codelist」  
→Define-XMLに反映



ケース1: Value Level Metadataを定義しない場合

→Sponsor DefinedのCodelistを作成する

ケース2: Value Level Metadataを定義する場合

(例えば、レコードによってデータの属性が異なる等)

→Value Level Metadataを作成し、Sponsor DefinedのCodelistを作成する

※英語文字列(変換後)のCodelistを作成しない場合、Define-XMLに不備が生じてしまう

# Define-XMLの例示

- ケース1: Value Level Metadataを定義しない場合

方法: Sponsor DefinedのCodelistを作成する

## QS Tests [CL.QSTEST]

### Permitted Value (Code)

JAPANESE TEXT IN SOURCE DATABASE01

JAPANESE TEXT IN SOURCE DATABASE02

JAPANESE TEXT IN SOURCE DATABASE03

## QS Test Codes [CL.QSTESTCD]

### Permitted Value (Code)

### Display Value (Decode)

QSTEST1

JAPANESE TEXT IN SOURCE DATABASE01

QSTEST2

JAPANESE TEXT IN SOURCE DATABASE02

QSTEST3

JAPANESE TEXT IN SOURCE DATABASE03

# Define-XMLの例示 (XMLコード)

## - ケース1: Value Level Metadataを定義しない場合

<XMLコード>

ItemGroupDef、ItemDef Definitions Sectionは省略

⋮

```
<CodeList OID="CL.QSTEST" Name="Question Name" DataType="text">
  <EnumeratedItem CodedValue="JAPANESE TEXT in SOURCE DAYABASE01"/>
  <EnumeratedItem CodedValue="JAPANESE TEXT in SOURCE DAYABASE02"/>
  <EnumeratedItem CodedValue="JAPANESE TEXT in SOURCE DAYABASE03"/>
</CodeList>
```

```
<CodeList OID="CL.QSTESTCD" Name=" Question Short Name" DataType="text">
  <CodeListItem CodedValue="QSTEST1">
    <Decode>
      <TranslatedText xml:lang="en"> JAPANESE TEXT in SOURCE DAYABASE01</TranslatedText>
    </Decode>
  </CodeListItem>
  <CodeListItem CodedValue="QSTEST2">
    <Decode>
      <TranslatedText xml:lang="en"> JAPANESE TEXT in SOURCE DAYABASE02</TranslatedText>
    </Decode>
  </CodeListItem>
  <CodeListItem CodedValue="QSTEST3">
    <Decode>
      <TranslatedText xml:lang="en"> JAPANESE TEXT in SOURCE DAYABASE03</TranslatedText>
    </Decode>
  </CodeListItem>
</CodeList>
```

⋮

CodeList Definitions Section

# Define-XML作成のアプローチ

## - ケース2: Value Level Metadataを定義する場合

Row	STUDYID	DOMAIN	USUBJID	QSSEQ	QSTESTCD	QSTEST
1	ABC123	QS	123101	1	QSTEST1	JAPANESE TEXT IN SOURCE DATABASE01
2	ABC123	QS	123101	2	QSTEST2	JAPANESE TEXT IN SOURCE DATABASE02
3	ABC123	QS	123101	3	QSTEST3	JAPANESE TEXT IN SOURCE DATABASE03

### 追加の条件を仮定:

- QSTEST1、QSTEST2の場合、QSORRESはCodelist <No Yes Response> に従う
- QSTEST3の場合、QSORRESは1桁の数値が格納される  
⇒ QSTESTCDの内容に応じて、データの属性が異なる



Value Level Metadataを定義



方法: Value Level Metadataを作成し、Sponsor DefinedのCodelistを作成する

# Define-XMLの例示

## - ケース2: Value Level Metadataを定義する場合

方法: Value Level Metadataを作成し、Sponsor DefinedのCodelistを作成する

### Value Level Metadata - QS [QSORRES]

Variable	Where	Type	Length / Display Format	Controlled Terms or Format	Origin	Derivation/Comment
QSORRES	<a href="#">QSTESTCD</a> EQ QSTEST1 (JAPANESE TEXT IN SOURCE DATABASE01)	text	2	["N" = "No", "NA" = "Not Applicable", "Y" = "Yes"] <No Yes Response>	CRF Page <a href="#">17</a>	
QSORRES	<a href="#">QSTESTCD</a> EQ QSTEST2 (JAPANESE TEXT IN SOURCE DATABASE02)	text	2	["N" = "No", "NA" = "Not Applicable", "Y" = "Yes"] <No Yes Response>	CRF Page <a href="#">17</a>	
QSORRES	<a href="#">QSTESTCD</a> EQ QSTEST3 (JAPANESE TEXT IN SOURCE DATABASE03)	integer	1		CRF Page <a href="#">17</a>	

### QS Tests [CL.QSTEST]

Permitted Value (Code)
JAPANESE TEXT IN SOURCE DATABASE01
JAPANESE TEXT IN SOURCE DATABASE02
JAPANESE TEXT IN SOURCE DATABASE03

### QS Test Codes [CL.QSTESTCD]

Permitted Value (Code)	Display Value (Decode)
QSTEST1	JAPANESE TEXT IN SOURCE DATABASE01
QSTEST2	JAPANESE TEXT IN SOURCE DATABASE02
QSTEST3	JAPANESE TEXT IN SOURCE DATABASE03



# Define-XMLの例示 (XMLコード)

## - ケース2: Value Level Metadataを定義する場合

### <XMLコード①>

```
<def:ValueListDef OID="VL.QS.QSORRES">
  <ItemRef ItemOID="IT.QS.QSORRES.QSTEST1" OrderNumber="1" Mandatory="Yes">
    <def:WhereClauseRef WhereClauseOID="WC.QS.QSTESTCD.QSTEST1"/>
  </ItemRef>
  <ItemRef ItemOID="IT.QS.QSORRES.QSTEST2" OrderNumber="2" Mandatory="Yes">
    <def:WhereClauseRef WhereClauseOID="WC.QS.QSTESTCD.QSTEST2"/>
  </ItemRef>
</def:ValueListDef>
<ItemRef ItemOID="IT.QS.QSORRES.QSTEST3" OrderNumber="3" Mandatory="Yes">
  <def:WhereClauseRef WhereClauseOID="WC.QS.QSTESTCD.QSTEST3"/>
</ItemRef>
</def:ValueListDef>
  ⋮
<def:WhereClauseDef OID="WC.QS.QSTESTCD.QSTEST1">
  <RangeCheck SoftHard="Soft" def:ItemOID="IT.QS.QSTESTCD" Comparator="EQ">
    <CheckValue>QSTEST1</CheckValue>
  </RangeCheck>
</def:WhereClauseDef>
<def:WhereClauseDef OID="WC.QS.QSTESTCD.QSTEST2">
  <RangeCheck SoftHard="Soft" def:ItemOID="IT.QS.QSTESTCD" Comparator="EQ">
    <CheckValue>QSTEST2</CheckValue>
  </RangeCheck>
</def:WhereClauseDef>
<def:WhereClauseDef OID="WC.QS.QSTESTCD.QSTEST3">
  <RangeCheck SoftHard="Soft" def:ItemOID="IT.QS.QSTESTCD" Comparator="EQ">
    <CheckValue>QSTEST3</CheckValue>
  </RangeCheck>
</def:WhereClauseDef>
```

ValueListDef Definitions Section

WhereClauseDef Definitions Section

# Define-XMLの例示 (XMLコード)

## - ケース2: Value Level Metadataを定義する場合

<XMLコード②>

ItemGroupDef Definitions Sectionは省略

⋮

```
<ItemDef OID="IT.QS.QSORRES" Name="QSORRES" DataType="text" Length="14" SASFieldName="QSORRES">
  <Description>
    <TranslatedText xml:lang="en">Finding in Original Units</TranslatedText>
  </Description>
  <def:Origin Type="CRF">
    <def:DocumentRef leafID="LF.blankcrf">
      <def:PDFPageRef PageRefs="17" Type="PhysicalRef"/>
    </def:DocumentRef>
  </def:Origin>
  <def:ValueListRef ValueListOID="VL.QS.QSORRES"/>
</ItemDef>
<ItemDef OID="IT.QS.QSORRES.QSTEST1" Name="QSTEST1" DataType="text" Length="2" SASFieldName="QSTEST1">
  <Description>
    <TranslatedText xml:lang="en">JAPANESE TEXT in SOURCE DATABASE01</TranslatedText>
  </Description>
  <CodeListRef CodeListOID="CL.NYNA"/>
  <def:Origin Type="CRF">
    <def:DocumentRef leafID="LF.blankcrf">
      <def:PDFPageRef PageRefs="17" Type="PhysicalRef"/>
    </def:DocumentRef>
  </def:Origin>
</ItemDef>
```

⋮

ItemDef Definitions Section

# Define-XMLの例示 (XMLコード)

## - ケース2: Value Level Metadataを定義する場合

<XMLコード③>

⋮

```
<CodeList OID="CL.QSTEST" Name="Question Name" DataType="text">
  <EnumeratedItem CodedValue="JAPANESE TEXT in SOURCE DAYABASE01"/>
  <EnumeratedItem CodedValue="JAPANESE TEXT in SOURCE DAYABASE02"/>
  <EnumeratedItem CodedValue="JAPANESE TEXT in SOURCE DAYABASE03"/>
</CodeList>


<CodeList OID="CL.QSTESTCD" Name=" Question Short Name" DataType="text">
  <CodeListItem CodedValue="QSTEST1">
    <Decode>
      <TranslatedText xml:lang="en"> JAPANESE TEXT in SOURCE DAYABASE01</TranslatedText>
    </Decode>
  </CodeListItem>
  <CodeListItem CodedValue="QSTEST2">
    <Decode>
      <TranslatedText xml:lang="en"> JAPANESE TEXT in SOURCE DAYABASE02</TranslatedText>
    </Decode>
  </CodeListItem>
  <CodeListItem CodedValue="QSTEST3">
    <Decode>
      <TranslatedText xml:lang="en"> JAPANESE TEXT in SOURCE DAYABASE03</TranslatedText>
    </Decode>
  </CodeListItem>
</CodeList>
```

⋮

CodeList Definitions Section

# 変換前/変換後の文字列の対応付け

日本語を含む質問表やコードリスト等に対応する箇所には適切な英訳又は英語文字列を格納すること。なお、英語文字列を格納する場合、その英語文字列を相互に区別する必要があるならば、末尾に番号をつけるなど個別に適切な対応を取り(例:「JAPANESE TEXT IN SOURCE DATABASE 01」、「JAPANESE TEXT IN SOURCE DATABASE 02」)、元の日本語との対応付けをデータガイド等に示すこと。



(技術的ガイド「4.1.5 日本語データの取り扱いについて」より抜粋)

QSTEST(変換前の日本語文字列)	QSTEST(変換後の英語文字列)
自分で布団を敷けますか？	JAPANESE TEXT IN SOURCE DATABASE01
ぞうきんがけはできますか？	JAPANESE TEXT IN SOURCE DATABASE02
ラジオ体操をしても平気ですか？	JAPANESE TEXT IN SOURCE DATABASE03

Define-XMLを用いて、上記のような対応表を作成することは困難  
→データガイドによる説明が簡易

## 例2のまとめ:

原文の日本語を英訳することが困難であるが、使用する英語文字列を相互に区別する必要がある場合

- ・変換後の英語文字列に対応したCodelistを作成する  
→作成しない場合、Define-XMLに不備が発生する  
※必要に応じて、Value Level Metadataを作成する
- ・変換前の日本語文字列と変換後の英数字文字列の対応付け  
→データガイドによる説明が簡易

# Acknowledgement

本資料を作成するにあたり、多くの意見や示唆を頂いたCDISC技術チームの皆様に感謝致します

- 三沢 秀敏 (ファイザー株式会社) ※CDISC技術チームリーダー 以下五十音順
- 青沼 秀樹 (シミック株式会社)
- 浅見 由美子 (第一三共株式会社)
- 内田 浩伸 (バイエル薬品株式会社)
- 小田 拓馬 (ヤンセンファーマ株式会社)
- 坂上 拓 (株式会社中外臨床研究センター)
- 高浪 洋平 (武田薬品工業株式会社)
- 中嶋 優一 (ノバルティスファーマ株式会社)
- 濱田 聖子 (株式会社CACエクシケア)
- 牧野 奈緒 (イーピーエス株式会社)
- 三沢 喬 (エイツーヘルスケア株式会社)
- 水野 義久 (アステラス製薬株式会社)
- 山口 孝一 (日本イーライリリー株式会社)