

OpenCDISCのバリデーションルールとともに Define-XML specification (ver. 2.0)を紐解く

坂上 拓

株式会社 中外臨床研究センター バイオメトリクス部
データサイエンスグループ

Open Define-XML Specification (ver. 2.0) with OpenCDISC Validation Rules

Taku Sakaue

Chugai Clinical Research Center CO., LTD.

Biometrics Dept. Data Science Group

要旨：

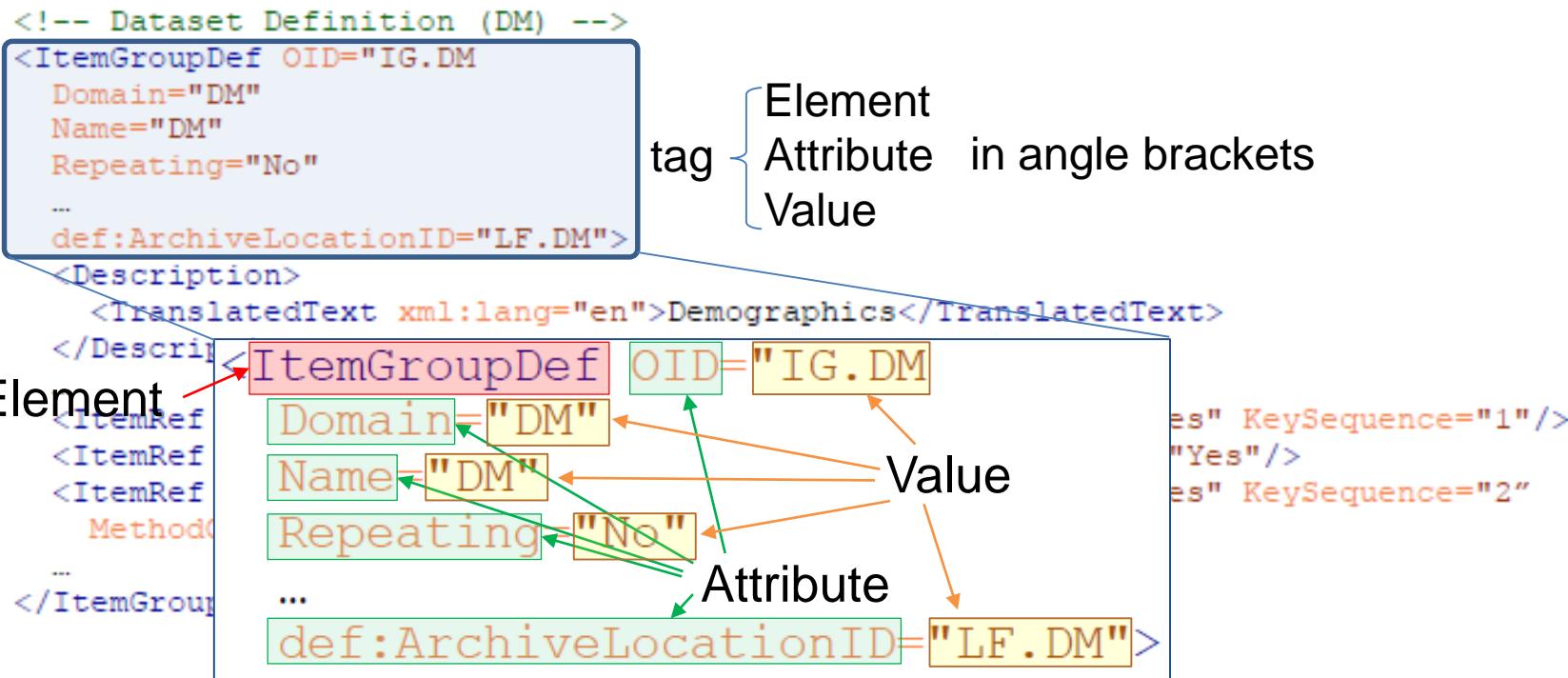
医薬品製造販売承認申請時の電子データ提出に伴い、CDASHやSDTM、ADaMといったCDISC標準の導入は進んできたものの、Define-XMLはXMLをベースとした標準という特殊性から、Define-XMLに係る学習や標準の導入に高いハードルがある。近年、Define-XMLに係る論文や、define.xml作成ツール等が公開され、担当者を選ばずdefine.xmlを作成できるようになってきたが、OpenCDISCバリデーターを用いた検証の段階で、発生したエラーの対応に、Define-XMLに係る知識の差が大きく影響してしまう。

本発表では、こういった知識の差を埋めるべく、OpenCDISCバリデーターのバリデーションルールと、Define-XMLのspecification (ver. 2.0)との関連を示しながら、Define-XMLの概要を説明するとともに、Define-XMLのspecificationの着眼点を示す。

キーワード: define.xml, OpenCDISC Validator, Define-XML Specification

Background of XML required to read Define-XML specification

- XML stands for eXtensible Markup Language.
It's composed of “tags”.



Background of XML required to read Define-XML specification

- To describe the relationship between the elements,
The elements compose parent-child relationship.

parent

```
<ItemDef OID="IT.VS.VSORRESU" Name="VSORRESU" DataType="text" Length="9"  
SASFieldName= parent / child  
<Description>  
  <TranslatedText lang="en">Standard Units</TranslatedText>  
</Description>      child  
<def:ValueListRef ValueListOID="VL.VS.VSORRESU"/>  
</ItemDef>
```

XPath showing where the element belongs in the XML.

example :

ItemDef/Description/TranslatedText
ItemDef/def:ValueListRef

Background of XML required to read Define-XML specification

- How describe these information on the Define-XML specification?

5.3.3. ODM Element

The first XML element in a file is known as the root element. For Define-XML files, the root element is the ODM element.

This element identifies the namespaces used and includes attributes that affect the processing of the document as a whole.

Element Name:	ODM
Element XPath:	/ODM
Element Textual Value:	None
Usage	Requirement: Required Cardinality: Exactly one Other Information: This is the root element for the <i>define.xml</i> document
Attributes:	xmlns, xmlns:def, xmlns:xlink, xmlns:xsi, xsi:schemalocation, ODMVersion, FileType, FileOID, CreationDateTime, AsOfDateTime, Originator, SourceSystem, SourceSystemVersion
Child Elements:	Study

Attribute
relationship between elements

Attribute	Usage	Allowable Values	Description
xmlns	Required	"http://www.cdisc.org/ns/odm/v1.3"	Identifies the default namespace for this document.
xmlns:def	Required	"http://www.cdisc.org/ns/def/v2.0"	XML namespace for Define-XML v2.0.0. While "def:" is suggested prefix for the Define-XML namespace, it should not be relied upon by the receiving application.

Background of XML required to read Define-XML specification

- How to convert define.xml to human-readable format?

```
<?xml version="1.0" encoding="UTF-8"?>
<ODM xmlns="http://www.cdisc.org/ns/odm/v1.3"
      xmlns:xlink="http://www.w3.org/1999/xlink"
      xmlns:def="http://www.cdisc.org/ns/def/v2.0"
      FileType="Snapshot" ODMVersion="1.3.2"
      FileOID="Studycdisc01-Define-XML_2.0.0"
      CreationDateTime="2012-06-21T11:07:23-05:00"
      Originator="CDISC XML Technologies Team">
  <Study OID="cdisc01">
    <GlobalVariables>
      <StudyName>CDISC01</StudyName>
      <StudyDescription>CDISC Test Study</StudyDescription>
      <ProtocolName>CDISC 01</ProtocolName>
    </GlobalVariables>
    <MetaDataVersion OID="MDV.CDISC01.SDTMIG.3.1.2.SDTM.1.2">
      Name="Study CDISC01, Data Definitions"
      Description="Study CDISC01, Data Definitions"
      def:DefineVersion="2.0.0"
      def:StandardName="CDISC SDTM"
      def:StandardVersion="3.1.2">
```

```
< Annotated Case Report Forms (def:AnnotatedCRF) >
< Supplemental Data Definitions (def:SupplementalDoc) >
< Value Level Metadata (def:ValueListDef) >
< Where Clause Definitions (def:WhereClauseDef) >
< Domain Level Metadata (ItemGroupDef) >
< Variable Level Metadata (ItemDef) >
< Controlled Terminology Metadata (CodeList) >
< Computational Algorithms (MethodDef) >
< Comments (def:CommentDef) >
< Referenced Documents (def:leaf) >
```



Analysis Datasets for Study CDISC-Sample (ADaM-IG 1.0)

Dataset	Description	Class	Structure	Purpose	Keys	Location	Documentation
ADSL	Subject-Level Analysis	SUBJECT LEVEL ANALYSIS DATASET	one record per subject	Analysis	STUDYID, USUBJID	adsl.xpt	Screen Failures are excluded since they are not needed for this study analysis. See Analysis Data Reviewer's Guide, page 6. Analysis Data Reviewer's Guide
ADQSDAS	ADAS-Cog Analysis	BASIC DATA STRUCTURE	One record per subject per parameter per analysis visit per analysis date	Analysis	STUDYID, USUBJID, PARAMCD, AVISIT, ADT	adqsdas.xpt	See referenced dataset creation program and Analysis Data Reviewer's Guide, Section 2.1 adqsdas.sas Analysis Data Reviewer's Guide
ADAE	Adverse Events Analysis Dataset	OCCURRENCE DATA STRUCTURE	one record per subject per adverse event	Analysis	STUDYID, USUBJID, AETERM, ASTDT, AESEQ	adae.xpt	See SAS program adae.sas

Go to the [top](#) of the define.xmlSubject-Level Analysis (ADSL) [Location: [adsl.xpt](#)]

Variable	Label	Key	Type	Length / Display Format	Controlled Terms or Format	Source/Derivation/Comment
STUDYID	Study Identifier	1	text	12		Predecessor: DM.STUDYID
USUBJID	Unique Subject Identifier	2	text	11		Predecessor: DM.USUBJID
SUBJID	Subject Identifier for the Study		text	4		Predecessor: DM.SUBJID
SITEID	Study Site Identifier		text	3		Predecessor: DM.SITEID
SITEGR1	Pooled Site Group 1		text	3		Derived: refer to SAP, Section 7.1 - if not pooled then SITEGR1=SITEID. If pooled, SITEGR1 will be 900
ARM	Description of Planned Arm		text	20	["Placebo", "Xanomeline Low Dose", "Xanomeline High Dose"] <Actual Treatment>	Predecessor: DM.ARM
TRTO1P	Planned Treatment for Period 01		text	20	["Placebo", "Xanomeline Low Dose", "Xanomeline High Dose"] <Actual Treatment>	Predecessor: DM.ARM

Composition of OpenCDISC Validation Check

- 2 Types of Rule ID
 - Consist of 2 types of Rule ID (ODxxxx, DDxxxx)
 - ODxxxx : ODM specific check
 - DDxxxx : Define-XML specific check
 - Target XML element and attribute have prefix “def:” (e.g. def:WhereClauseDef element, def:ExtendedValue attribute).
- 7 Categories
 - Each Rule ID are divided into 7 main categories (Consistency, Cross-Reference, Format, Metadata, Presence, Structure, Terminology).

OpenCDISC Validation Check

- Structure
 - Checking adherence with definition of ODM and Define-XML structure and data types.
 - Checking XML header
 - Checking ODM root element settings

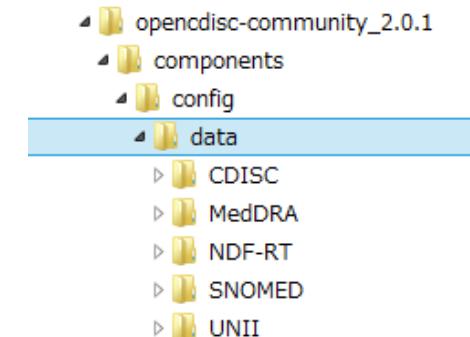
5.3.3. ODM Element

Attribute	Usage	Allowable Values	Description
xmlns	Required	"http://www.cdisc.org/ns/odm/v1.3"	Identifies the default namespace for this document.
xmlns:def	Required	"http://www.cdisc.org/ns/def/v2.0"	XML namespace for Define-XML v2.0.0. While "def:" is suggested prefix for the Define-XML namespace, it should not be relied upon by the receiving application.
xmlns:xlink	Conditional Required when xlink:href is provided.	"http://www.w3.org/1999/xlink"	XML namespace for XLink.
xmlns:xsi	Conditional Required when xsi:schemalocation is provided.	"http://www.w3.org/2001/XMLSchema-instance"	XML Schema instance namespace. Required when xsi:schemalocation is provided.

OpenCDISC Validation Check

- Metadata
 - Checking position and validity of the elements and the attributes by comparing with OpenCDISC defined metadata, and checking the elements parent-child relationship.

in OpenCDISC, compare with define.xml and own metadata
(Define Metadata #.#.xml, ODM Metadata #.#.#.xml).



OD0004, DD0005, DD0007,
DD0010, DD0011, DD0023

For ARM is unsupported, Some errors are detected when
contains ARM definitions in define.xml.

OD0005, DD0007, DD0008

OpenCDISC Validation Check

- Format
 - Checking value of the elements and the attributes.
 - Target attributes (OrderNumber, SASFormatName, CodedValue, Version*).
* Version : attribute of ExtraCodeList element

```
< <CodeList OID="CL.AESEV" Name="Severity/Intensity Scale for Adverse Events" ;" />
  DataType="text" SASFormatName="$AESEV">
```

Attribute	Usage	Allowable Value	Description
SASFormatName	Optional	Text	<p>The SASFormatName must be a legal SAS format.</p> <p><u>Business Rules:</u> The SASFormatName needs to start with a "\$" character in case the CodeList DataType is "text".</p>

OpenCDISC Validation Check

- Presence
 - Checking required elements and attributes.

Rule ID	check attributes - missing or empty		
5.3.5. MetaDataVersion Element		CreationDateTime	
Attribute	Usage	Allowable Values	Description
OID	Required	Text	Unique ID for the MetaDataVersion. See the ODM specification section 2.11 for OID considerations.
Name	Required	Text	Name for the MetaDataVersion.
Description	Optional	Text <u>Sample:</u> "Study CDISC01 Data Definitions"	Description for the MetaDataVersion.

MetadataVersion/def:leaf/@ID, @xlink:href

OpenCDISC Validation Check

- Presence
 - Checking required elements and attributes.

5.3.4.1. GlobalVariables Element

GlobalVariables is the first child of Study.

Element Name:	GlobalVariables
Element XPath:	/ODM/Study/GlobalVariables
Element Textual Value:	None
Usage	Required Cardinality: One or More Type: GlobalVariables
Attributes:	None
Child Elements:	Study
5.3.6.1. def:DocumentRef Element	
Element Name:	def:DocumentRef
Element XPath:	/ODM/Study/MetaDataVersion/def:AnnotatedCRF/def:DocumentRef /ODM/Study/MetaDataVersion/def:SupplementalDoc/def:DocumentRef /ODM/Study/MetaDataVersion/ItemDef/def:Origin/def:DocumentRef /ODM/Study/MetaDataVersion/MethodDef/def:DocumentRef /ODM/Study/MetaDataVersion/def:CommentDef/def:DocumentRef
Element Textual Value:	None
Usage	<p><u>Requirement:</u> Conditional Required for def:AnnotatedCRF and def:SupplementalDoc Optional for def:Origin, MethodDef and def:CommentDef</p> <p><u>Cardinality:</u> One or More</p> <p><u>Business Rule:</u> If multiple documents are provided there will be multiple def:DocumentRef child elements within the def:SupplementalDoc element.</p>
Attributes:	leafID
Child Elements:	def:PDFPageRef

OpenCDISC Validation Check

- Presence
 - Checking required elements and attributes with conditions.

5.3.11. ItemDef Element

Attribute	Usage	Allowable Values	Description
SignificantDigits	Conditional Required if DataType is “float”.	Integer	The number of digits following the decimal point in a floating point number. <u>Business Rule:</u> When DataType is float both Length and SignificantDigits must be provided.

OD0071

OpenCDISC Validation Check

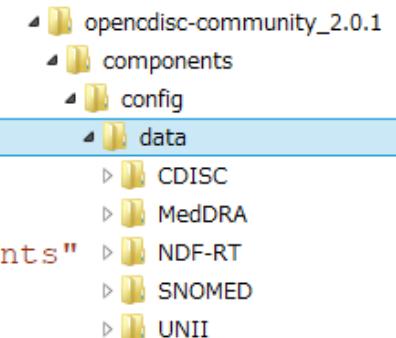
- Terminology
 - Checking allowable values defined in the Define-XML specification.

Ru	5.3.10. ItemGroupDef Element			
	Attribute	Usage	Allowable Value	
ODC	Repeating	Required	<p>Allowable Values:</p> <p>Yes</p> <p>No</p>	<ul style="list-style-type: none">Indicates whether a domain contains more than one record per subject or only one record per subject. <p>Business Rules:</p> <ul style="list-style-type: none">Set Repeating='No' when used in the reference data section
ODC				e / text
ODC				
ODC				
ODC				
DDC	IsReferenceData	Optional	<p>Allowable Values:</p> <p>Yes</p> <p>No</p>	Indicates whether the dataset contains reference data (not subject data)
DDC		<u>Default Value:</u> No	<p>Sample:</p> <p>For Trial Design Domains: Yes</p> <p>For Demographics Domain: No</p>	

OpenCDISC Validation Check

- Terminology
 - Checking values consistent with the Control Terminology.

in the OpenCDISC, compare with define.xml and own terminology (SDTM Terminology.odm.xml).



```
<CodeList OID="CL.AESEV" Name="Severity/Intensity Scale for Adverse Events" DataType="text" SASFormatName="$AESEV">
  <CodeListItem CodedValue="MILD" Rank="1">
    <Decode>
      <TranslatedText xml:lang="en">Grade 1</TranslatedText>
    </Decode>
    <Alias Name="C41338" Context="nci:ExtCodeID"/>
  </CodeListItem>
```

DD0027

DD0028

For the OpenCDISC does not refer to the ADaM Terminology, Some errors are detected when date imputation flag (C81223) or derivation type (C81224) are defined in define.xml.

OpenCDISC Validation Check

- Consistency
 - Checking appearance frequency of the elements consist with ODM and Define-XML specifications.

5.3.4.1. GlobalVariables Element

GlobalVariables is the first child of Study.

5.3.10. ItemGroupDef Element

GlobalVariables			
Attribute	Usage	Allowable Values	Description
OID	Required	Text	<p>Unique ID for the ItemGroupDef (dataset).</p> <p>See the ODM specification section 2.11 for OID considerations.</p>

same applies to **ItemDef** and **CodeList**)

DD0031

DD0032

OpenCDISC Validation Check

- Cross-Reference
 - Checking consistency of reference relationship.

5.3.11. ItemDef Element

For each Variable and Value definition an ItemDef element should be provided. Use one ItemDef element per unique variable.

Element Name:	ItemDef
Element XPath:	/ODM/Study/MetaDataVersion/ItemDef
Element Textual Value:	<i>None</i>
Usage	<p>Requirement: Required</p> <p><u>Cardinality:</u></p> <ul style="list-style-type: none"> An ItemDef element is required for each ItemOID value that appears in an ItemRef contained in a MetaDataVersion..
Attributes:	OID, Name, DataType, Length, SignificantDigits, SASFieldName, def:DisplayFormat, def:CommentOID
Child Elements :	Description , CodeListRef , def:Origin , def:ValueListRef

ents
emOID
'@CodeListOID
ntRef/@leafID
Ref/@ValueListOID

5.3.8.1. ItemRef Element

Attribute	Usage	Allowable Values	Description
ItemOID	Required	Text	Reference to the unique ID of an ItemDef element.

Conclusion

- Especially Important points in Define-XML

5.3.10. ItemGroupDef Element

The ItemGroupDef element with the set of attributes and child elements as described below is used to describe the dataset metadata in XML.

One for each dataset included within the study submitted.

Element Name:	ItemGroupDef
Element XPath:	/ODM/Study/MetaDataVersion/ItemGroupDef
Element Textual Value:	None
Usage	Requirement: Required Cardinality: One or more Business Rule: <ul style="list-style-type: none">The Implementation Guides for each relevant standard specify the datasets required for FDA submissions.
Attributes:	OID, Name, Repeating, IsReferenceData, SASDatasetName, Domain, Purpose, def:Structure, def:Class, def:ArchiveLocationID, def:CommentOID
Child Elements:	Description, ItemRef , Alias, def:leaf

Basic Rules

Parent-Child Relationship

Required Item

Attribute	Usage	Allowable Values	Description
OID	Required	Text Value Settings	Unique ID for the ItemGroupDef (dataset). See the ODM specification section 2.11 for OID considerations.
Name	Required	Text	Short description for the ItemGroup.

Conclusion

- Relationship between elements

“Defs” elements	“Refs” elements
ItemDef/@OID	ItemRef/@ItemOID
CodeList/@OID	CodeListRef/@CodeListOID
def:leaf/@ID	def:DocumentRef/@leafID
def:ValueListDef/@OID	def:ValueListRef/@ValueListOID
MethodDef/@OID	ItemRef/@MethodOID
def:WhereClauseDef/@OID	def:WhereClauseRef/@WhereClauseOID
def:CommentDef/@OID	def:WhereClauseDef/@def:CommentOID
ItemDef/@OID	def:WhereClauseDef/RangeCheck/@def:ItemOID
ItemGroupDef/@OID	arm:AnalysisDataset/@ItemGroupOID
def:CommentDef/@OID	ItemGroupDef/@def:CommentOID
def:CommentDef/@OID	ItemDef@/@def:CommentOID
ItemDef/@OID	arm:AnalysisResult/@ParameterOID
ItemDef/@OID	arm:AnalysisVariable/@ItemOID

References

- CDISC Define-XML Specification Version 2.0
<http://www.cdisc.org/define-xml>
- OpenCDISC
<http://www.opencdisc.org/>