ABSTRACT
Case Report Tabulation Data Definition Specifications (CRT-DDS) commonly known as Define-xml is an integral part of FDA’s electronic submissions. The most recent addition to define.xml package is Analysis Result Metadata. Define.xml facilitates the review process by providing information about the structure and content of standard clinical data (SDTM and ADaM) in a machine readable format whereas Analysis Result Metadata provides information about the purpose of the analysis, ADaM datasets and analysis methods used for analysis. This paper presents a data-driven approach for generating define.xml v2.0 where a Define ready ADaM specification plus ADaM datasets are passed into a macro that automatically generates define.xml v2.0. For Analysis Result metadata, a standard TFL annotation spreadsheet is also passed into a macro that automatically generates analysis result metadata, providing information about the relationship between analysis results, analysis datasets and documentation of the analysis.

INTRODUCTION
Sponsor generate ADaM specifications which serves as a programming guidance document for generating ADaM datasets. These ADaM specifications are also submitted to FDA in the Case Report Tabulation Data Definition Specifications format or Define.xml prepared by CDISC define.xml team. Define.xml provide the structure and content of ADaM datasets in a machine readable format.

Analysis Results Metadata(ARM.XML) provides traceability for a given analysis result to the specific ADaM data used as input. It also provides standard format for identification of key analyses, information about analysis method used and reason for analysis which can assists regulatory reviewers by linking results, documentation and datasets.

The process of generating a define.xml and ARM.xml can be very time consuming if done manually. This paper depicts a process of generating the documents using ADaM specifications, ADaM datasets and ARM.xml with least manual input.

A Define.xml provides information about the following metadata’s:
- Dataset Metadata:
  - Name, Description, Class, Structure, Purpose, Key Variables, Location, Documentation
- Variable Metadata:
  - Name, Label, Type, Length/Display Format, Controlled Terminology, Source/Derivation/Comment
- Parameter Value Level Metadata
  - Describes method used to derive a particular variable for different PARAMCDs in an BDS ADaM dataset
- Controlled Terminology
- Analysis Derivation
- Comments
- External Documentation (ADRG, SAS codes)

PRE-REQUIRMENTS
Find enclosed the list of the items that are required to generate define.xml using this macro:
- Define Stylesheet(define2-0-0.xls)
- ADaM Datasets
PhUSE US Connect 2018

- Trimmed ADaM Datasets where character variable lengths are adjusted to the maximum length of the variable used across all the datasets in the study.
- SAS Transport file using Trimmed ADaM Datasets

- ADaM ADS specifications
- External Documentation
  - ADRG
  - SAS Codes
  - Other external documentation that is being referred in ADaM ADS specifications

ADaM Specifications

A standard ADaM specification template is created to generate define.xml. A separate worksheet is created for each metadata that is being presented in define.xml.

Study Information

In this section, we populate Protocol Number, Study Description, Protocol Name, ADaM IG version and language used for writing ADaM specifications and generating ADaM analysis datasets.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>StudyName</td>
<td>111-P-222</td>
</tr>
<tr>
<td>StudyDescription</td>
<td>Define for Study XYZ</td>
</tr>
<tr>
<td>ProtocolName</td>
<td>Define for Study XYZ</td>
</tr>
<tr>
<td>StandardName</td>
<td>ADaM-IG</td>
</tr>
<tr>
<td>StandardVersion</td>
<td>1.0</td>
</tr>
<tr>
<td>Language</td>
<td>en</td>
</tr>
</tbody>
</table>

Figure 1. Study Information

Dataset Metadata

Dataset metadata serves as a TOC that lists out all the analysis datasets. The following information is collected in dataset metadata:

- Dataset Name
- Dataset Description or Dataset label
- Dataset Structure
- Dataset Purpose
  - Analysis for ADaM Datasets
- Dataset Class (ADSL, BDS, OCCD and ADaM OTHER)
- Key Variables
- Repeating
  - Yes: For ADS that have more than one record for each USUBJID
  - No: For ADS that have one record for each USUBJID
- Documentation
- Document
  - A unique short name given to each external document. Each short name has to start with a prefix “LR.”
Variable Metadata

All the variables that are included in each analysis dataset are described in variable metadata.

- Dataset Name
- Variable Name
- Variable Label
- Variable Type as per data types give in CDISC ODM
  - text, integer, float, date, time, datetime, partialDate, partialTime, partialDatetime, incompleteDatetime
- Variable length
- Sigdigit: Significant digits. Only populated where variable type is float
- Key: Integer value assigned to all the key variables
- Format: date and time associated with date/time numeric variables
- Codelist: A unique name assigned to standard or sponsor defined valid values and decodes
- Origin:
  - Derived, Assigned and Predecessor
- Source/Derivation
- Valuelist
  - Populated only for BDS type of ADaM datasets
  - Populated with a value of “Y” for those variables whose derivation changes for each parameter
  - Value Level Metadata will be populated
- Mandatory
  - Populated with values of Yes/No depending upon whether a variable is required or not.
- Document
  - A unique short name given to each external document. Each short name has to start with a prefix “LF.”

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Variable</th>
<th>Label</th>
<th>Type</th>
<th>Length</th>
<th>Sigdigit</th>
<th>Key</th>
<th>Format</th>
<th>CodeList</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADSL</td>
<td>STUDYID</td>
<td>Study Identifier</td>
<td>text</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL</td>
<td>SITEID</td>
<td>Study Site Identifier</td>
<td>text</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL</td>
<td>INVID</td>
<td>Investigator Identifier</td>
<td>text</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL</td>
<td>INVNAM</td>
<td>Investigator Name</td>
<td>text</td>
<td>200</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Value Level Metadata

Value level metadata will be populated for all the variables for whom ValueList column is marked as “Y” in the variable metadata.

- **Dataset Name**
  - **Variable attributes for which value level metadata will be created**
    - Variable Name
    - Variable Label
    - Variable Type
    - Variable Length
    - Sigdigit: Significant digits. Only populated when variable type is float
    - Origin: Derived/Assigned

- **PARAMCD:** List of PARAMCD that are being generated in a particular analysis dataset
- **Source:** Describes the derivation of a variable for each PARAMCD. If the derivation is more than 1000 characters, this derivation will be saved in an external document. A short name for this document will be listed in Document column
- **Document**
  - A unique short name given to each external document. Each short name has to start with a prefix “LF.”

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Variable</th>
<th>Label</th>
<th>Type</th>
<th>Length</th>
<th>Sigdigit</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADVS</td>
<td>PARAM</td>
<td>Parameter</td>
<td>text</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>ADVS</td>
<td>PARAM</td>
<td>Parameter</td>
<td>text</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>ADVS</td>
<td>PARAM</td>
<td>Parameter</td>
<td>text</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>ADVS</td>
<td>PARAM</td>
<td>Parameter</td>
<td>text</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>ADVS</td>
<td>PARAM</td>
<td>Parameter</td>
<td>text</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>ADVS</td>
<td>PARAM</td>
<td>Parameter</td>
<td>text</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>ADVS</td>
<td>PARAM</td>
<td>Parameter</td>
<td>text</td>
<td>200</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Origin</th>
<th>PARAMCD</th>
<th>Source</th>
<th>Document</th>
</tr>
</thead>
<tbody>
<tr>
<td>Derived</td>
<td>VSALL</td>
<td>PARAM= trim(left(VS.VSTEST))</td>
<td></td>
</tr>
<tr>
<td>Derived</td>
<td>BMI</td>
<td>PARAM= trim(left(VS.VSTEST))</td>
<td></td>
</tr>
<tr>
<td>Derived</td>
<td>DIABP</td>
<td>PARAM= trim(left(VS.VSTEST))</td>
<td></td>
</tr>
<tr>
<td>Derived</td>
<td>HEIGHT</td>
<td>PARAM= trim(left(VS.VSTEST))</td>
<td></td>
</tr>
<tr>
<td>Derived</td>
<td>PULSE</td>
<td>PARAM= trim(left(VS.VSTEST))</td>
<td></td>
</tr>
</tbody>
</table>
Derived SYSBP PARAM= trim(left(VS.VSTEST)) || " (" || trim(left(VS.VSSTRESU)) || ")"

Derived WEIGHT PARAM= trim(left(VS.VSTEST)) || " (" || trim(left(VS.VSSTRESU)) || ")"

Figure 4. Value Level Metadata

Controlled Terminology

Controlled terminology collects standard or sponsor defined valid values and decodes.

- Codelist Name: A unique short name assigned to a codelist
- Codelist Label: A short description of codelist
- Codelist Type: text/integer/float
- Code: Standard or sponsor defined valid values
- Decode: Coded value for each standard or sponsor defined value
- Rank: Populated only for codelists that have a numeric significance. Rank should not be used to define a display order
- OrderNumber: OrderNumber is used to define a display order for the items in a particular codelist
- CodeListNum: Populate with C-codes provided in National Cancer Institute’s Enterprise Vocabulary System if unique short name of a codelist matches with standard codelist names provided in National Cancer Institute’s Enterprise Vocabulary System.
- CodeNum: Populate with C-codes provided in National Cancer Institute’s Enterprise Vocabulary System if any of standard or sponsor defined valid values matches with standard names provided in National Cancer Institute’s Enterprise Vocabulary System.

<table>
<thead>
<tr>
<th>Codelist Name</th>
<th>CodeList Label</th>
<th>CodeList Type</th>
<th>Code</th>
<th>Decode</th>
<th>Rank</th>
<th>OrderNumber</th>
<th>CodeListNum</th>
<th>CodeNum</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGEU</td>
<td>Age Unit</td>
<td>text</td>
<td>YEARS</td>
<td>Screening</td>
<td>1</td>
<td>C66781</td>
<td>C29848</td>
<td></td>
</tr>
<tr>
<td>SEX</td>
<td>Sex</td>
<td>text</td>
<td>M</td>
<td>1</td>
<td>1</td>
<td>C66731</td>
<td>C20197</td>
<td></td>
</tr>
<tr>
<td>SEX</td>
<td>Sex</td>
<td>text</td>
<td>F</td>
<td>2</td>
<td>2</td>
<td>C66731</td>
<td>C16576</td>
<td></td>
</tr>
<tr>
<td>AVISITN</td>
<td>AVISITN</td>
<td>integer</td>
<td>1</td>
<td>Screening</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AVISITN</td>
<td>AVISITN</td>
<td>integer</td>
<td>2</td>
<td>Day -2 to -1</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AVISITN</td>
<td>AVISITN</td>
<td>integer</td>
<td>3</td>
<td>Day 1 to 3</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AVISITN</td>
<td>AVISITN</td>
<td>integer</td>
<td>4</td>
<td>Day 8</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RACE</td>
<td>Race</td>
<td>text</td>
<td>WHITE</td>
<td>5</td>
<td></td>
<td>C74457</td>
<td>C41261</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5. Controlled Terminology

Documents

This section collects the following information about each external documentation that is being referred in ADaM ADS specifications:

- Document ID: A unique short name with a prefix of “LF:”, given to each external document.
- Title: Text that will be displayed in define.xml
- HRef: Physical location of the file with the actual file name and file extension. If define.xml and external documentation is located in the same folder, populate actual file name and file extension only.
<table>
<thead>
<tr>
<th>Document ID</th>
<th>Title</th>
<th>HRef</th>
</tr>
</thead>
<tbody>
<tr>
<td>LF.ADSL.PGM</td>
<td>adsl.sas</td>
<td>d0_adsl.sas</td>
</tr>
<tr>
<td>LF.ADV.SPGM</td>
<td>advs.sas</td>
<td>d1_advs.sas</td>
</tr>
</tbody>
</table>

**Figure 6. Documents**

**ARM.XML**

For generating ARM.xml, a standard template will be populated TLF mocks annotation (similar to ADaM specs template) with below tabs.

- Analysis
- WhereClause
- Documents

**Analysis**

- Table Number: Table number defined in mock
- Table Title: Table title defined in mock
- CSR Ref: Any reference
- Analysis: Type of analysis need to be performed
- Merge: Any merge variable
- Reason: Reference documents e.g SAP/Protocol
- Dataset: Dataset name
- Analysis Variable: Analysis variables used in the analysis
- PARAMCD: Parameter Code
- Description: Description to support the analysis
- SAP: Reference to SAP section
- Code: Statistical procedure

**Figure 7. ARM.XMl Analysis**

**WhereClause**

- Table Number:
- Analysis: Type of analysis to be performed
- Dataset: Dataset name
- Variable: Variable names
- Value: Value for each variable
### Table Number

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Dataset</th>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference between treatment by week</td>
<td>ADKNEE</td>
<td>PARAMCD</td>
<td>MIKPW1</td>
</tr>
<tr>
<td>Difference between treatment by week</td>
<td>ADKNEE</td>
<td>FASFL</td>
<td>Y</td>
</tr>
<tr>
<td>Difference between treatment by week</td>
<td>ADKNEE</td>
<td>AVISITN</td>
<td>3</td>
</tr>
<tr>
<td>Difference between treatment by week</td>
<td>ADKNEE</td>
<td>AVISITN</td>
<td>4</td>
</tr>
<tr>
<td>Difference between treatment by week</td>
<td>ADKNEE</td>
<td>AVISITN</td>
<td>5</td>
</tr>
<tr>
<td>Difference between treatment by over 4 weeks</td>
<td>ADKNEE</td>
<td>PARAMCD</td>
<td>MIKPW1</td>
</tr>
<tr>
<td>Difference between treatment by over 4 weeks</td>
<td>ADKNEE</td>
<td>FASFL</td>
<td>Y</td>
</tr>
<tr>
<td>Difference between treatment by over 4 weeks</td>
<td>ADKNEE</td>
<td>AVISITN</td>
<td>5</td>
</tr>
<tr>
<td>Difference between treatment by over 4 weeks</td>
<td>ADKNEE</td>
<td>AVISITN</td>
<td>6</td>
</tr>
<tr>
<td>Difference between treatment by over 4 weeks</td>
<td>ADKNEE</td>
<td>AVISITN</td>
<td>7</td>
</tr>
<tr>
<td>Incidence of Treatment Emergent Serious Adverse Events by Treatment Group</td>
<td>ADSL</td>
<td>SAFFL</td>
<td>Y</td>
</tr>
<tr>
<td>Incidence of Treatment Emergent Serious Adverse Events by Treatment Group</td>
<td>ADAE</td>
<td>TEAE</td>
<td>Y</td>
</tr>
<tr>
<td>Incidence of Treatment Emergent Serious Adverse Events by Treatment Group</td>
<td>ADAE</td>
<td>AESER</td>
<td>Y</td>
</tr>
</tbody>
</table>

**Figure 8. ARM.XMl WhereClause**

**Documents**

This section collects the following information about each external documentation that is being referred in ARM.XML specifications:

- **Document ID**: A unique short name with a prefix of “LF.”, given to each external document.
- **Title**: Table Number that will be displayed in ARM.xml
- **HRef**: Physical location of the file with the actual file name and file extension.

<table>
<thead>
<tr>
<th>ID</th>
<th>Title</th>
<th>Href</th>
</tr>
</thead>
<tbody>
<tr>
<td>LF.SAP-SEC-10.4</td>
<td>SAP Section 10.4</td>
<td>sap.pdf</td>
</tr>
<tr>
<td>LF.Table 14.2.1.1.2.1</td>
<td>Table 14.2.1.1.2.1</td>
<td>csr.pdf</td>
</tr>
<tr>
<td>LF.Table 14-5.02.sas</td>
<td>Table 14-5.02.sas</td>
<td>at14-5-02-sas.txt</td>
</tr>
</tbody>
</table>

**Figure 9. ARM.XMl Documents**
The location of trimmed ADaM analysis datasets, final ADaM ADS specifications and TFL Annotation are passed into define.sas macro which generates define.xml v2.0 and ARM.xml. Define.sas macro refers to trimmed ADaM analysis datasets for variable attributes (Name, label, type, length), order of the variables in a particular analysis dataset and controlled terminology for derived variables.

Figure 10. Define.xml v2.0 viewed via Microsoft Internet Explorer
This data driven approach using ADaM Specifications, ADaM Datasets and TFL Annotation is a time-saving, straightforward and easy way of generating Define.xml v2.0 and ARM.xml. Programmers don’t need to have XML knowledge for generating Define.xml or ARM.xml via this method.

REFERENCES


RECOMMENDED READING

CDISC Define-XML Specification, Version 2.0
Study Data Technical Conformance Guide, March 2018

CONTACT INFORMATION

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

Mark Crangle
ICON Clinical Research
Email: mark.crangle@iconplc.com

Hima Bhatia
ICON Clinical Research
Email: hima.bhatia@iconplc.com

Brand and product names are trademarks of their respective companies.