

Paper DS07

Generating Define.xml and Analysis Result Metadata using Specifications, Datasets and TFL Annotation

Mark Crangle, ICON Clinical Research, Marlow, United Kingdom

ABSTRACT

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

Sponsors generate ADaM specifications which serve as a programming guidance document for generating ADaM datasets. Much of the information in ADaM specifications is replicated in Define.xml which are generated as part of a submission package to provide the structure and content of ADaM datasets in a machine readable format.

The process of generating a define.xml can be very time consuming if done manually and requires an in depth knowledge of the xml language and the define.xml structure. This paper depicts a process of generating define.xml v2.0 using ADaM specifications, ADaM datasets and user friendly input templates for Analysis Results Metadata (ARM) with manual input. This means that users with less knowledge of the xml programming language can create accurate and compliant define.xml files, with support of subject matter experts who can assist with any queries.

ICONS macros utilize a standard excel templates for the ADaM Specifications and to collect the Analysis Results Metadata (if required) and combine these with the datasets themselves to create the define.xml.

This paper will demonstrate how the examples from the Analysis Results Metadata (ARM) v1.0 for Define-XML v2.0 release package are transcribed in the ICON ADaM specifications and ARM templates and show how this information is populated in the define.xml file by ICON standard macros.

ADAM DATASETS

The first step in the process of creating the define.xml is to prepare the ADaM datasets and convert into the xpt dataset format required for eCTD submission.

At the time in the process that datasets are prepared for submission, they will already have been validated for CDISC compliance by Pinnacle 21 and other tools so minimal processing of the datasets is required at this stage. The only processing that does occur is that character variable lengths are "trimmed" to the maximum length of any value within that variable. This ensures that variable lengths are consistent across the dataset package, and to minimise the size of the datasets being submitted for efficiency.

For example, the RACE variable could be defined in ADSL with length 200 with values:

- AMERICAN INDIAN OR ALASKA NATIVE
- ASIAN
- BLACK OR AFRICAN AMERICAN
- WHITE

The longest of these values has length 32 so the variable length is updated in the dataset to have length 32, and this is also applied to any other ADaM datasets that also contain this variable to ensure that the length is consistent across the package.

The macro also examines the size of the datasets themselves to ensure that each dataset being submitted is below the 5 gigabyte limit set by the FDA in their Technical Conformance Guide and outputs a message to users if any datasets are over this size limit.

DATASET SPECIFICATIONS

The primary function of the ADS specifications document is for the programming team to use in the creation and validation of the datasets but the standard ADaM ADS specification template is designed to also allow the generation of the define.xml. This re-using of the ADS specifications reduces the amount of work needed to create the define.xml and ensures consistency between the datasets themselves and the define.xml

A Dataset Metadata tab defines dataset level metadata for all datasets and then a worksheet tab is created for each dataset to define the variable level metadata. Where a dataset uses Value Level Metadata, this is also defined in a separate tab, and there are also tabs to define Controlled Terminology and the external document links that are needed for the define.xml.

DATASET METADATA

The dataset metadata tab lists all analysis datasets included in the package and describes dataset level metadata. The following information is collected in dataset metadata:

Column Label	Explanation	Use in define.xml
Dataset Name	Name of the dataset	Used for the Name and SASDatasetName attribute of the ItemGroupDef element <pre><ItemGroupDef OID="IG.ADSL" Name="ADSL" SASDatasetName="ADSL" Repeating="No" IsReferenceData="No" Purpose="Analysis" def:Structure="One record per subject" def:Class="SUBJECT LEVEL ANALYSIS DATASET" def:CommentOID="COM.ADSL" def:ArchiveLocationID="LF.ADSL"></pre>
Dataset Description	Dataset label. Should match with the actual label of the dataset	Used in the Description element within each ItemGroupDef element <pre><Description> <TranslatedText xml:lang="en">Subject-Level Analysis Dataset</TranslatedText> </Description></pre>
Dataset Structure	Brief description of dataset structure	Used for the def:Structure attribute of the ItemGroupDef element <pre><ItemGroupDef OID="IG.ADSL" Name="ADSL" SASDatasetName="ADSL" Repeating="No" IsReferenceData="No" Purpose="Analysis" def:Structure="One record per subject" def:Class="SUBJECT LEVEL ANALYSIS DATASET" def:CommentOID="COM.ADSL" def:ArchiveLocationID="LF.ADSL"></pre>
Class of Dataset	Dataset class. Permitted values for ADaM define.xml files are: - SUBJECT LEVEL ANALYSIS DATASET - ADVERSE EVENTS ANALYSIS DATASET - BASIC DATA STRUCTURE - ADAM OTHER	Used for the def:Class attribute of the ItemGroupDef element <pre><ItemGroupDef OID="IG.ADSL" Name="ADSL" SASDatasetName="ADSL" Repeating="No" IsReferenceData="No" Purpose="Analysis" def:Structure="One record per subject" def:Class="SUBJECT LEVEL ANALYSIS DATASET" def:CommentOID="COM.ADSL" def:ArchiveLocationID="LF.ADSL"></pre>
Documentation	Comment about the content of the dataset or to explain where to find further information	If populated then a def:CommentOID attribute is added to the ItemGroupDef and the comment is used to create a def:CommentDef element <pre><ItemGroupDef OID="IG.ADSL" Name="ADSL" SASDatasetName="ADSL" Repeating="No" IsReferenceData="No" Purpose="Analysis" def:Structure="One record per subject" def:Class="SUBJECT LEVEL ANALYSIS DATASET" def:CommentOID="COM.ADSL" def:ArchiveLocationID="LF.ADSL"></pre>
Purpose	Purpose of the dataset. For ADaM define.xml files the expected value is "Analysis"	Used for the Purpose attribute of the ItemGroupDef element <pre><ItemGroupDef OID="IG.ADSL" Name="ADSL" SASDatasetName="ADSL" Repeating="No" IsReferenceData="No" Purpose="Analysis" def:Structure="One record per subject" def:Class="SUBJECT LEVEL ANALYSIS DATASET" def:CommentOID="COM.ADSL" def:ArchiveLocationID="LF.ADSL"></pre>

Column Label	Explanation	Use in define.xml
Repeating	Indicates whether a dataset contains more than one row per subject.	Used for the Repeating attribute of the ItemGroupDef element <pre><ItemGroupDef OID="IG.ADSL" Name="ADSL" SASDatasetName="ADSL" Repeating="No" IsReferenceData="No" Purpose="Analysis" def:Structure="One record per subject" def:Class="SUBJECT LEVEL ANALYSIS DATASET" def:CommentOID="COM.ADSL" def:ArchiveLocationID="LF.ADSL"></pre>
Document	External document references that will be included in the Analysis Datasets section of the define.xml. The document ID should match what has been specified in the Documents tab where filenames and titles are defined.	If populated then a def:DocumentRef element is added to the def:CommentDef element for this dataset. <pre><def:CommentDef OID="COM.ADSL"> <Description> <TranslatedText>Screen Failures are excluded since they are not needed for this study analysis. See Analysis Data Reviewer's Guide, page 6.</TranslatedText> </Description> <def:DocumentRef leafID="LF.ADRG"/> </def:CommentDef></pre>

The following table shows how the information from the Analysis Results Metadata release package is translated to the Dataset Metadata tab in the specifications template:

Dataset Name	Dataset Description	Dataset Structure	Class of Dataset	Documentation	Purpose	Repeating	Document
ADSL	Subject-Level Analysis Dataset	One record per subject	SUBJECT LEVEL ANALYSIS DATASET	Screen Failures are excluded since they are not needed for this study analysis. See Analysis Data Reviewer's Guide, page 6.	Analysis	No	LF.ADRG;
ADQSADAS	ADAS-Cog Analysis Dataset	One record per subject per parameter per analysis visit per analysis date	BASIC DATA STRUCTURE	See referenced dataset creation program and Analysis Data Reviewer's Guide, Section 2.1	Analysis	Yes	LF.ADRG; LF.ADQSDAS.PGM;
ADAE	Adverse Events Analysis Dataset	One record per subject per adverse event	OCCURRENCE DATA STRUCTURE	See SAS program	Analysis	Yes	LF.ADAE.PGM;

VARIABLE METADATA

All the variables that are included in each analysis dataset are described in variable metadata. This information is combined with the ADaM datasets and any discrepancies between the information in the dataset specifications document and the datasets are highlighted to the user. The variable length is also taken directly from the trimmed datasets that have been prepared for submission.

Variable information is used to create both ItemRef elements giving the order of variables in the dataset and ItemDef elements giving details of the variable attributes. The information entered in the Variable Metadata tab is used to create both elements.

The excel template contains one tab for each dataset included in the package with the following information:

Column Label	Explanation	Use in define.xml
Variable Name	Variable name	Used for the Name and SASFieldName attributes of the ItemDef element <pre><ItemDef OID="IT.ADSL.STUDYID" Name="STUDYID" SASFieldName="STUDYID" DataType="text" Length="1"></pre>
Variable Type	Variable type as per the permitted data types given in the CDISC define-xml standard.	Used for the DataType attribute of the ItemDef element <pre><ItemDef OID="IT.ADSL.STUDYID" Name="STUDYID" SASFieldName="STUDYID" DataType="text" Length="1"></pre>
Format	SAS Format associated with a date/time numeric variables	Used for the def:DisplayFormat attribute of the ItemDef element <pre><ItemDef OID="IT.ADSL.TRTSDT" Name="TRTSDT" SASFieldName="TRTSDT" DataType="integer" Length="8" def:DisplayFormat="date9."></pre>
Sigdigit	Number of significant digits for float variables	Used for the SignificantDigits attribute of the ItemDef element <pre><ItemDef OID="IT.ADSL.AVGDD" Name="AVGDD" SASFieldName="AVGDD" DataType="float" SignificantDigits="1" Length="8"></pre>
Codelist	A unique name assigned to standard or sponsor defined valid values and decodes. Matches with an entry on the controlled terminology tab	If populated then a CodeListRef element is added to the ItemDef element <pre><ItemDef OID="IT.ADSL.TRT01PN" Name="TRT01PN" SASFieldName="TRT01PN" DataType="integer" Length="8" def:CommentOID="COM.ADSL.TRT01PN"> <CodeListRef CodeListOID="CL.ARMN"/> </ItemDef></pre>
Origin	Categorise how the variable is populated. Permitted values are Derived, Assigned or Predecessor	Used in the def:Origin element within the ItemDef element <pre><def:Origin Type="Predecessor"> <Description> <TranslatedText xml:lang="en">DM.ARM</TranslatedText> </Description> </def:Origin> OR <def:Origin Type="Assigned"/> OR <def:Origin Type="Derived"></pre>
Source/Derivation	Combined with the origin column, this describes how the variable is populated. If the variable source is derived, the value in this column is used in the Analysis Derivations section of the define.xml. If the variable source is Assigned, this value is used in the Comments section of the define.xml	If Origin is Predecessor then the source is added as a Description element within the def:Origin element <pre><def:Origin Type="Predecessor"> <Description> <TranslatedText xml:lang="en">DM.ARM</TranslatedText> </Description> </def:Origin></pre> If Origin is Assigned then a def:CommentDef element is added and referenced in the def:CommentOID attribute of the ItemDef element If Origin is Derived then the derivation is used to create a MethodDef element which is referenced in the MethodOID attribute of the ItemRef element <pre><ItemRef ItemOID="IT.ADSL.SITEGR1" OrderNumber="5" Mandatory="No" MethodOID="MT.ADSL.SITEGR1"/></pre>
Key	Integer value assigned to identify key variables for the dataset	Used to populate the KeySequence attribute of the ItemRef element <pre><ItemRef ItemOID="IT.ADSL.STUDYID" OrderNumber="1" Mandatory="No" KeySequence="1"/></pre>
Mandatory	Populated with values of Yes/No depending on whether the variables is required or not	Used in the Mandatory attribute of the ItemRef element <pre><ItemRef ItemOID="IT.ADSL.STUDYID" OrderNumber="1" Mandatory="No" KeySequence="1"/></pre>
Valuelist	Populated with a value of "Y" for those variables whose derivations changes for each parameter. Value Level Metadata tab will be populated for this dataset and variable.	Used to add a def:ValueListRef element to the ItemDef element <pre><ItemDef OID="IT.ADQSADAS.AVAL" Name="AVAL" SASFieldName="AVAL" DataType="integer" Length="8"> <def:ValueListRef ValueListOID="VL.ADQSADAS.AVAL"/> </ItemDef></pre>

The following table shows how select information from the release package is translated into the ICON template.

Variable Name	Variable Type	Format	Codelist	Source/Derivation	Key	Origin	Mandatory
STUDYID	text			DM.STUDYID	1	Predecessor	No
USUBJID	text			DM.USUBJID	2	Predecessor	No
SUBJID	text			DM.SUBJID		Predecessor	No
SITEID	text			DM.SITEID		Predecessor	No
SITEGR1	text			refer to SAP, Section 7.1 - if not pooled then SITEGR1=SITEID. If pooled, SITEGR1 will be 900		Derived	No
ARM	text		ARM	DM.ARM		Predecessor	No
TRT01P	text		ARM	DM.ARM		Predecessor	No
TRT01PN	integer		ARMN	Numeric code for TRT01P which corresponds to the randomized dose		Assigned	No
TRT01A	text		ARM	TRT01A=TRT01P, i.e., no difference between actual and randomized treatment in this study.		Assigned	No
TRT01AN	integer		ARMN	Numeric code for TRT01A which corresponds to the randomized dose		Assigned	No

VALUE LEVEL METADATA

Value level metadata will be populated for all the variables where derivations are described by value level metadata. Each dataset that has a variable where derivations are described by value level metadata will have a separate tab containing the information outlined below. Each variable that has value level metadata defined will have a def:ValueListDef element defined and each unique combination of variable, PARAMCD value and Source/Derivation will have ItemRef and ItemDef elements defined similarly to the Variable Metadata discussed above.

Column Label	Explanation	Use in define.xml
Variable	Variable name	Used for the Name and SASFieldName attributes of the ItemDef element and to define the OID attribute for this element <pre><ItemDef OID="IT.ADQSADAS.QSSEQ.ACITM01-ACITM14" Name="QSSEQ" SASFieldName="QSSEQ" DataType="text" Length="7"></pre>
Type	Value type as per the permitted data types given in the CDISC define-xml standard.	Used for the Datatype attribute of the ItemDef element <pre><ItemDef OID="IT.ADQSADAS.QSSEQ.ACITM01-ACITM14" Name="QSSEQ" SASFieldName="QSSEQ" DataType="text" Length="7"></pre>
Format	SAS Format associated with a date/time numeric variables	Used for the def:DisplayFormat attribute of the ItemDef element
Sigdigit	Number of significant digits for float variables	Used for the SignificantDigits attribute of the ItemDef element
Origin	Categorise how the variable is populated. Permitted values are Derived, Assigned or Predecessor	Used in the def:Origin element within the ItemDef element <pre><def:Origin Type="Predecessor"> <Description> <TranslatedText xml:lang="en">QS.QSSEQ </TranslatedText> </Description> </def:Origin> OR <def:Origin Type="Assigned"/> OR <def:Origin Type="Derived"></pre>

Column Label	Explanation	Use in define.xml
Source/Derivation	Combined with the origin column, this describes how variable is populated. If the variable source is derived, the value in this column is used in the Analysis Derivations section of the define.xml. If the variable source is Assigned, this value is used in the Comments section of the define.xml	If Origin is Predecessor then the source is added as a Description element within the def:Origin element <pre><def:Origin Type="Predecessor"> <Description> <TranslatedText xml:lang="en">QS.QSSEQ</TranslatedText> </Description> </def:Origin></pre> If Origin is Assigned then a def:CommentDef element is added and referenced in the def:CommentOID attribute of the ItemDef element If Origin is Derived then the derivation is used to create a MethodDef element which is referenced in the MethodOID attribute of the ItemRef element <pre><ItemRef ItemOID="IT.ADQSADAS.QSSEQ.ACTOT" Mandatory="No" MethodOID="MT.ADQSADAS.QSSEQ.ACTOT"></pre>
Document	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	Used to add a def:DocumentRef element to the MethodDef element <pre><MethodDef OID="MT.ADQSADAS.AVAL.ACTOT" Name="CM.ADQSADAS.AVAL.ACTOT" Type="Computation"> <Description> <TranslatedText xml:lang="en">Sum of ADAS scores for items 1, 2, 4, 5, 6, 7, 8, 11, 12, 13, and 14, see Analysis Data Reviewers Guide (Page 3) for details on adjusting for missing values.</TranslatedText> </Description> <def:DocumentRef leafID="LF.ADRG"/> </MethodDef></pre>

The following table shows how the value level metadata for the ADQSADAS dataset from the release package is populated in the excel template.

Variable	Type	PARAMCD	Source/Derivation	Origin	Document
AVAL	integer	ACITM01	QS.QSSTRESN where QSTESTCD=PARAMCD	Derived	
AVAL	integer	ACITM02	QS.QSSTRESN where QSTESTCD=PARAMCD	Derived	
AVAL	integer	ACITM03	QS.QSSTRESN where QSTESTCD=PARAMCD	Derived	
AVAL	integer	ACTOT	Sum of ADAS scores for items 1, 2, 4, 5, 6, 7, 8, 11, 12, 13, and 14, see Analysis Data Reviewers Guide (Page 3) for details on adjusting for missing values.	Derived	LF.ADRG
QSSEQ	text	ACITM01	QS.QSSEQ	Predecessor	
QSSEQ	text	ACITM02	QS.QSSEQ	Predecessor	
QSSEQ	text	ACITM03	QS.QSSEQ	Predecessor	
QSSEQ	Text	ACTOT	Set QSSEQ to missing for post baseline records. Set to QS.QSSEQ where QS.VISIT=BASELINE and QS.QSTESTCD=ACTOT	Assigned	

CONTROLLED TERMINOLOGY

Controlled terminology collects standard or sponsor defined valid values and decodes. The following information is entered into the template:

Column Label	Explanation	Use in define.xml
Codelist	Unique short name assigned to the codelist	Used to create the OID attribute of the CodeList element
Codelist Label	Short description of the codelist	Used to create the Name attribute of the CodeList element
Codelist Type	Type of values included in the codelist. Permitted values are given in the CDISC define-xml standard	Used to create the Name attribute of the CodeList element

Column Label	Explanation	Use in define.xml
Code Value	Valid values covered by the codelist	If Decode Value is not populated then the Code Value is used in an EnumeratedItem element within the CodeList element <pre><CodeList OID="CL.AGEGR1" Name="Age Group" DataType="text"> <EnumeratedItem CodedValue="<65" OrderNumber="1" Rank="1"/> </CodeList></pre> If Decode Value is populated then the Code Value is used to define a CodeListItem element <pre><CodeList OID="CL.AGEGR1N" Name="Age Group (N)" DataType="integer"> <CodeListItem CodedValue="1" OrderNumber="1" Rank="1"></pre>
Decode Value	Optional column to specify decode values for the coded values	Used to define a Decode element within a CodeListItem element <pre><CodeListItem CodedValue="1" OrderNumber="1" Rank="1"> <Decode> <TranslatedText xml:lang="en"><65</TranslatedText> </Decode></pre>
Rank	Populated only for codelists that have a numeric significance. Rank should not be used to define a display order	Used to populate the Rank attribute in either a EnumeratedItem or CodeListItem element <pre><CodeListItem CodedValue="1" OrderNumber="1" Rank="1"></pre>
OrderNumber	OrderNumber is used to define a display order for the items in a particular codelist	Used to populate the OrderNumber attribute in either an EnumeratedItem or CodeListItem element <pre><EnumeratedItem CodedValue="1" OrderNumber="1" Rank="1"></pre>
Code	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.	Used to create an Alias element within the CodeList element <pre><CodeList OID="CL.AGEU" Name="Age Unit" DataType="text"> <Alias Name="C66781" Context="nci:ExtCodeID"/> </CodeList></pre>
Codelist Code	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.	Used to create an Alias element within either an EnumeratedItem or CodeListItem element <pre><EnumeratedItem CodedValue="YEARS" OrderNumber="1" Rank="1"> <Alias Name="C29848" Context="nci:ExtCodeID"/> </EnumeratedItem></pre>

The following table shows how selected entries from the ARM release package are handled in the excel template.

Codelist	Codelist Name	Codelist Type	Code Value	Decode Value	Rank	OrderNumber	Code	Codelist Code
AGEGR1	Age Group	text	<65		1	1		
AGEGR1	Age Group	text	65-80		2	2		
AGEGR1	Age Group	text	>80		3	3		
AGEGR1N	Age Group (N)	integer	1	<65	1	1		
AGEGR1N	Age Group (N)	integer	2	65-80	2	2		
AGEGR1N	Age Group (N)	integer	3	>80	3	3		
AGEU	Age Unit	text	YEARS		1	1	C66781	C29848
ARM	Actual Treatment	text	Placebo		1	1		

Codelist	Codelist Name	Codelist Type	Code Value	Decode Value	Rank	OrderNumber	Code	Codelist Code
ARM	Actual Treatment	text	Xanomeline Low Dose		2	2		
ARM	Actual Treatment	text	Xanomeline High Dose		3	3		

DOCUMENTS

This tab collects the following information about each external documentation that is being referred in define.xml:

Column Label	Explanation	Use in define.xml
Document ID	A unique short name with a prefix of "LF.", given to each external document.	Populates the ID attribute of the def:leaf element <pre><def:leaf ID="LF.ADRG" xlink:href="analysis-data-reviewers-guide.pdf"> <def:title>Analysis Data Reviewer Guide</def:title> </def:leaf></pre>
Title	Text label that will be displayed in the define.xml	Used for the def:title element of the def:leaf element <pre><def:leaf ID="LF.ADRG" xlink:href="analysis-data-reviewers-guide.pdf"> <def:title>Analysis Data Reviewer Guide</def:title> </def:leaf></pre>
Href	Physical location of the file with the actual filename and file extension	Populates the xlink:href attribute of the def:leaf element <pre><def:leaf ID="LF.ADRG" xlink:href="analysis-data-reviewers-guide.pdf"> <def:title>Analysis Data Reviewer Guide</def:title> </def:leaf></pre>

The following table shows how the documents referenced in the dataset sections of the release package are mapped to the ICON template.

ID	Title	Href
LF.ADRG	Analysis Data Reviewer Guide	analysis-data-reviewers-guide.pdf
LF.ADQSADAS.PGM	adqsadas.sas	../programs/adqsadas-sas.txt
LF.ADAE.PGM	adae.sas	../programs/adae-sas.txt

ANALYSIS RESULTS METADATA

The information for the Analysis Results Metadata are collected in a separate spreadsheet to the dataset specifications and are then incorporated with the dataset information by the standard macros. Reflecting the way that the ARM information is represented in the xml file, there are separate tabs to describe the analysis that is being performed, the where clauses used to select data for each analysis and the documents referenced in the Analysis Results Metadata.

ANALYSIS

The analysis tab defines the analysis being described in the Analysis Results Metadata. There is one record for each table, analysis within the table and source dataset being used. This allows for multiple analyses within one table to be described, and for multiple source datasets to be referenced with information about how the source datasets are merged together.

Column Label	Explanation	Use in xml file
Table Number	Number of the table or figure being described	Used to populate the arm:ResultDisplay element in the xml file <pre><arm:ResultDisplay OID="RD.Table 14-3.01" Name="Table 14-3.01"></pre>
Table Title	Title of the table or figure being described	Used for the Description element of the arm:ResultDisplay element <pre><arm:ResultDisplay OID="RD.Table 14-3.01" Name="Table 14-3.01"> <Description> <TranslatedText xml:lang="en">Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population)</TranslatedText> </Description></pre>

Column Label	Explanation	Use in xml file
CSR Ref	Reference to the CSR section and page where the table is used. This information is in two parts, separated by "/", with the first part being the document reference and the second part being the page number within the document where the table is used	The first part is used for the leafID attribute of the def:DocumentRef The second part is used for the def:PDFPageRef element to define a link to the specific page of the CSR. <pre><def:DocumentRef leafID="LF.Table-14-3.01"> <def:PDFPageRef PageRefs="2" Type="PhysicalRef" /></pre>
Analysis	Description of the analysis performed	Used for the Description element of the arm:AnalysisResult element <pre><arm:AnalysisResult OID="AR.Table 14-3.01.R.1" ParameterOID="IT.ADQSADAS.PARAMCD" AnalysisReason="SPECIFIED IN SAP" AnalysisPurpose="PRIMARY OUTCOME MEASURE"> <Description> <TranslatedText xml:lang="en">Dose response analysis for ADAS-Cog changes from baseline </TranslatedText> </Description></pre>
Merge	Text description of the way that multiple datasets are merged together. This is an optional column and does not need to be populated if a single source dataset supports the analysis	Used to add def:CommentDef element which is then referenced by the Def:CommentOID attribute of the arm:AnalysisDatasets element. <pre><arm:AnalysisDatasets def:CommentOID="COM.Table 14- 5.02.R.1.MERGE"></pre>
Reason	Reason for the analysis. Suggested values are "SPECIFIED IN PROTOCOL", "SPECIFIED IN SAP", "DATA DRIVEN", "REQUESTED BY REGULATORY AGENCY" but the codelist governing this attribute is extensible so additional values are permitted.	Used to populate the AnalysisReason attribute of the AnalysisResult element <pre><arm:AnalysisResult OID="AR.Table 14-3.01.R.1" ParameterOID="IT.ADQSADAS.PARAMCD" AnalysisReason="SPECIFIED IN SAP" AnalysisPurpose="PRIMARY OUTCOME MEASURE"></pre>
Purpose	Purpose of the analysis. Suggested values are "PRIMARY OUTCOME MEASURE", "SECONDARY OUTCOME MEASURE", "EXPLORATORY OUTCOME MEASURE" but the codelist governing this attribute is extensible so additional values are permitted.	Used to populate the AnalysisPurpose attribute of the AnalysisResult element <pre><arm:AnalysisResult OID="AR.Table 14-3.01.R.1" ParameterOID="IT.ADQSADAS.PARAMCD" AnalysisReason="SPECIFIED IN SAP" AnalysisPurpose="PRIMARY OUTCOME MEASURE"></pre>
Dataset	Dataset name. In the xml file this is a link to the data definition in the analysis datasets section.	Populates the ItemGroupOID attribute of the arm:AnalysisDataset element <pre><arm:AnalysisDataset ItemGroupOID="IG.ADQSADAS"></pre>
Analysis Variable	Analysis variables used in the analysis. If more than one variable is used then these should be specified together, separated by a space. In the xml file this is a link to the variable definition in the variable metadata section.	Populates the ItemOID attribute of the arm:AnalysisVariable element. <pre><arm:AnalysisVariable ItemOID="IT.ADQSADAS.CHG"/></pre>
PARAMCD	Y/N flag to identify whether a single parameter in a BDS-type dataset is used to support this analysis	Used to populate the ParameterOID attribute of the AnalysisResult element <pre><arm:AnalysisResult OID="AR.Table 14-3.01.R.1" ParameterOID="IT.ADQSADAS.PARAMCD" AnalysisReason="SPECIFIED IN SAP" AnalysisPurpose="PRIMARY OUTCOME MEASURE"></pre>

Column Label	Explanation	Use in xml file
Description	Text description of the statistical methods	Used in the arm:Documentation element <pre><Description> <TranslatedText xml:lang="en">ANCOVA analysis of CHG performed to provide pairwise comparisons among treatment groups and adjusted means; using randomized treatment as class variable and site group as class variable in model and the baseline value as a covariate. </TranslatedText> </Description></pre>
SAP	Reference to the SAP section and page where this analysis is specified. Again this is done in two parts, separated by "/" with the first part being a reference to the document and the second part being a reference to the page number.	Used within the arm:AnalysisResult element. The first part is used for the leafID attribute of the def:DocumentRef The second part is used for the def:PDFPageRef element to define a link to the specific page of the SAP. <pre><def:DocumentRef leafID="LF.SAP-SEC-10.1.1"> <def:PDFPageRef PageRefs="4" Type="PhysicalRef"/> </def:DocumentRef></pre>
Code	Details of the statistical procedure which can either be code itself or a reference to a separate SAS program file which is included in the submission package.	Used for the arm:ProgrammingCode element. If code is used then an arm:Code element is added, or if a reference to separate SAS file is used then a def:DocumentRef is added. <pre><arm:ProgrammingCode Context="SAS Version 9.2"> <arm:Code> proc glm data = ADQSADAS; where EFFFL=&apos;Y&apos; and ANL01FL=&apos;Y&apos; and AVISIT=&apos;Week 24&apos; and PARAMCD=&quot;ACTOT&quot;; class SITEGR1; model CHG = TRTPN SITEGR1; run; </arm:Code> </arm:ProgrammingCode></pre>

The following table shows how the examples from the Analysis Results Metadata release package are translated into the analysis tab. The merge column is not displayed because this information is not relevant to this example.

Table Number	Table Title	CSR Ref	Analysis	Reason	Purpose
Table 14-3.01	Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population)	LF.Table-14-3.01 / 2	Dose response analysis for ADAS-Cog changes from baseline	SPECIFIED IN SAP	PRIMARY OUTCOME MEASURE
Table 14-3.01	Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population)	LF.Table-14-3.01 / 2	Pairwise comparisons to placebo for ADAS-Cog changes from baseline	SPECIFIED IN SAP	PRIMARY OUTCOME MEASURE
Dataset	Analysis Variable	PARAMCD	Description	SAP	Code
ADQSADAS	CHG	Y	Linear model analysis of CHG for dose response...	LF.SAP-SEC-10.1.1 / 4	<pre>proc glm data = ADQSADAS; where EFFFL='Y' and ANL01FL='Y' and AVISIT='Week 24' and PARAMCD='ACTOT'; class SITEGR1; model CHG = TRTPN SITEGR1; run;</pre>
ADQSADAS	CHG	Y	ANCOVA analysis of CHG...	LF.SAP-SEC-10.1.1 / 4	<pre>proc glm data = ADQSADAS; where EFFFL='Y' and ANL01FL='Y' and AVISIT='Week 24' and PARAMCD='ACTOT'; class TRTPN SITEGR1; model CHG = TRTPN SITEGR1 BASE; means TRTPN; lsmeans TRTPN / OM STDERR PDIFF CL; run;</pre>

WHERECLAUSE

The WhereClause tab is used to specify the subset conditions that are used to generate the analysis being described. There is one record in this tab for each Table, Analysis, Dataset, Variable and Variable Value used in the subset condition.

Column Label	Explanation	Use in define.xml
Table Number	Number of the table or figure being described.	Used to match with the Table Number defined on the analysis tab
Analysis	Description of the analysis performed.	Used to match with the Analysis defined on the Analysis tab
Dataset Name	Dataset used for the where clause subset criteria	Matches with the Dataset Name defined on the analysis tab. Also used to populate the OID attribute of the def:WhereClauseDef element and the def:ItemOID attribute of the RangeCheck element which provides a link to the dataset and variable definition sections of the define.xml <pre><def:WhereClauseDef OID="WC.ADQSADAS.AVAL.ACITM01-ACITM14"> <RangeCheck Comparator="IN" SoftHard="Soft" def:ItemOID="IT.ADQSADAS.PARAMCD"></pre>
Variable	Variable used for the where clause subset criteria	Used to populate the OID attribute of the def:WhereClauseDef element and the def:ItemOID attribute of the RangeCheck element which provides a link to the dataset and variable definition sections of the define.xml <pre><def:WhereClauseDef OID="WC.ADQSADAS.AVAL.ACITM01-ACITM14"> <RangeCheck Comparator="IN" SoftHard="Soft" def:ItemOID="IT.ADQSADAS.PARAMCD"></pre>
Value	Value used in the subset criteria. The Where Clause tab should contain one row for each value that the variable in question is subset on. If more than one row per variable exists, this is converted to an "IN" condition for the define.xml, otherwise it is handled as an "EQ" condition	Used to generate CheckValue elements within the def:WhereClauseDef element. <pre><RangeCheck Comparator="EQ" SoftHard="Soft" def:ItemOID="IT.ADQSADAS.PARAMCD"> <CheckValue>ACTOT</CheckValue> OR <RangeCheck Comparator="IN" SoftHard="Soft" def:ItemOID="IT.ADQSADAS.PARAMCD"> <CheckValue>ACITM01</CheckValue> <CheckValue>ACITM02</CheckValue> </RangeCheck></pre>

The following table shows how the examples from the Analysis Results Metadata release package are translated into the WhereClause tab:

Table Number	Analysis	Dataset	Variable	Value
Table 14-3.01	Dose response analysis for ADAS-Cog changes from baseline	ADQSADAS	PARAMCD	ACTOT
Table 14-3.01	Dose response analysis for ADAS-Cog changes from baseline	ADQSADAS	AVISIT	Week 24
Table 14-3.01	Dose response analysis for ADAS-Cog changes from baseline	ADQSADAS	EFFFL	Y
Table 14-3.01	Dose response analysis for ADAS-Cog changes from baseline	ADQSADAS	ANL01FL	Y
Table 14-3.01	Pairwise comparisons to placebo for ADAS-Cog changes from baseline	ADQSADAS	PARAMCD	ACTOT
Table 14-3.01	Pairwise comparisons to placebo for ADAS-Cog changes from baseline	ADQSADAS	AVISIT	Week 24
Table 14-3.01	Pairwise comparisons to placebo for ADAS-Cog changes from baseline	ADQSADAS	EFFFL	Y
Table 14-3.01	Pairwise comparisons to placebo for ADAS-Cog changes from baseline	ADQSADAS	ANL01FL	Y

DOCUMENTS

The documents tab in the Analysis Results Metadata template is populated in the same way as it is in the Dataset Specifications template.

The following table shows how the documents referenced in the example in the Analysis Results Metadata release package would be specified in the ICON excel template:

ID	Title	href
LF.Table-14-3.01	Table 14-3.01	../dummy-csr/dummy-csr.pdf
LF.Table-14-5.02	Table 14-5.02	../dummy-csr/dummy-csr.pdf
LF.SAP-SEC-10.1.1	SAP Section 10.1.1	../dummy-csr/dummy-csr.pdf

ID	Title	href
LF.SAP-SEC-11.2	SAP Section 11.2	../dummy-csr/dummy-csr.pdf
LF.at14-5-02.PGM	at14-5-02.sas	../programs/at14-5-02-sas.txt

PROCESS FLOW OF GENREATING DEFINE.XML

The location of trimmed ADaM analysis datasets and final ADaM ADS specifications are passed into the define.sas macro which generates define.xml v2.0. 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.

In addition to the information that has been entered into the excel spreadsheets, some information used in the define.xml is passed to the macro in the form of macro variables in define.sas. The following macro variables are used:

Macro Variable	Explanation	Use in define.xml
StudyName	Name of the study	Used to populate the StudyName element <code><StudyName>CDISC-Sample</StudyName></code>
StudyDescription	Description of the study taken from the protocol	Used to populate the StudyDescription element <code><StudyDescription>CDISC-Sample Data Definition</StudyDescription></code>
ProtocolName	Protocol Number	Used in the ProtocolName element <code><ProtocolName>CDISC-Sample</ProtocolName></code>
StandardName	Name of the Implementation Guide used (ie. ADaM-IG)	Used in the def:StandardName element <code>def:StandardName="ADaM-IG"</code>
StandardVersion	Version of the Implementation Guide	Used in the def:StandardVersion element <code>def:StandardVersion="1.0"</code>
Language	Language used for the description	Used throughout define.xml wherever a TranslatedText element is used <code><TranslatedText xml:lang="en">Subject-Level Analysis Dataset</TranslatedText></code> <code><TranslatedText xml:lang="en">Unique Subject Identifier</TranslatedText></code>
SASVersion	Optional variable if Analysis Results Metadata is included in the define.xml. Specifies the SAS version used to write the code included in the document	Used to define the Context attribute of the arm:ProgrammingCode element <code><arm:ProgrammingCode Context="SAS Version 9.2"></code>

CONCLUSION

This data driven approach of generating define.xml using ADaM analysis datasets and ADaM ADS specifications is a time-saving, straightforward and easy way of generating define.xml v2.0. It ensures consistency with the ADaM datasets and study programming teams don't need a detailed knowledge of the xml language in order to create define.xml files that are compliant with the CDISC standard for this document.

REFERENCES

CDISC Define-XML Specification, Version 2.0, March 5, 2013 (<http://www.cdisc.org/define-xml>)
 Analysis Results Metadata (ARM) v1.0 for Define-XML v2.0, Version 1.0, Jan 30, 2015
<https://www.cdisc.org/standards/foundational/analysis-data-model-adam/analysis-results-metadata-arm-v10-define-xml-v20>)

ACKNOWLEDGMENTS

Many thanks to Jaskaran Gill for his work on development of the macros and templates and to Hima Bhatia for her help creating this paper.

CONTACT INFORMATION

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

Mark Crangle
ICON Clinical Research
Marlow International
Marlow
Buckinghamshire
United Kingdom
SL7 1YL
Email: Mark.Crangle@iconplc.com

Brand and product names are trademarks of their respective companies.