



GCE SOLUTIONS

Derive Value From Excellence ...

To Create define.xml using .XML and OpenCDISC Validator
Murali Neela, GCE solutions, India



Agenda

- ❖ CDISC Mission Statement
- ❖ Why Define.xml is important?
- ❖ Creating the define.xml
- ❖ Difference between V_1.0 and V2.0 of define.xml
- ❖ Creating the define.xml using open CDISC Validator
- ❖ Questions and Answers



CDISC MISSION STATEMENT:

CDISC is a global, open, multidisciplinary, non-profit organization that has established standards to support the acquisition, exchange, submission and archive of clinical research data and metadata. The CDISC mission is to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare. CDISC standards are vendor-neutral, platform-independent and freely available via the CDISC website.
<http://www.cdisc.org>



❖ Why Define.xml is important?

- ❖ Standard for providing Case Report Tabulations Data Definitions in an XML format for submission to a U.S. Food and Drug Administration (FDA).
- ❖ define file as a container of metadata (and table of contents) that describes all of the data and analysis that a submission contains. While it is a machine readable file
- ❖ Describe the format and content of the submitted SDTM, SEND, and AdaM datasets
- ❖ Avoids unnecessary queries.
- ❖ To increase the level of automation and improve the efficiency of the Regulatory Review process, define.xml can be used to provide the Data Definition Document in a machine-readable format. The formal name for this is the Case Report Tabulation Data Definition (CRT DD) specification.



XML Technology?

❖ There are three different types that work together in XML Technology

❖ Schema

❖ Definition and declaration of elements and their attributes file with extension: .XSD

❖ XML file

❖ Description of data and metadata in machine-readable format using the elements defined in the schema file with extension: .XML

❖ Style sheet

❖ Definition of the layout in a browser tool for the human readable representation of the XML file, file with extension: .XSL



Specifications for Define .xml

- ❖ Dataset definitions
- ❖ Dataset variable definitions
- ❖ Controlled Terminology definitions
- ❖ Value list definitions
- ❖ Links to supporting documents
- ❖ Computational method definitions
- ❖ Comments definitions



Define.xml contains metadata and is machine readable

Define.xml opened without using XSL

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<?xml-stylesheet type="text/xsl" href="define1-0-0.xsl"?>
<ODM
  xmlns="http://www.cdisc.org/ns/odm/v1.2"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:def="http://www.cdisc.org/ns/def/v1.0"
  xsi:schemaLocation="http://www.cdisc.org/ns/odm/v1.2 define1-0-0.xsd"
  FileOID="Study1234"
  ODMVersion="1.2"
  FileType="Snapshot"
  CreationDateTime="2004-07-28T12:34:13-06:00">
<Study OID="1234">
  <GlobalVariables>
    <StudyName>1234</StudyName>
    <StudyDescription>1234 Data Definition</StudyDescription>
    <ProtocolName>1234</ProtocolName>
  </GlobalVariables>
  <MetaDataVersion OID="CDISC.SDTM.3.1.0"
    Name="Study 1234, Data Definitions"
    Description="Study 1234, Data Definitions"
    def:DefineVersion="1.0.0"
    def:StandardName="CDISC SDTM"
    def:StandardVersion="3.1.0">
    <def:AnnotatedCRF>
      <def:DocumentRef leafID="blankcrf"/>
    </def:AnnotatedCRF>
    <def:leaf ID="blankcrf" xlink:href="blankcrf.pdf">
      <def:title>Annotated Case Report Form</def:title>
    </def:leaf>
    <def:SupplementalDoc>
      <def:DocumentRef leafID="SupplementalDataDefinitions"/>
    </def:SupplementalDoc>
    <def:leaf ID="SupplementalDataDefinitions"
      xlink:href="supplementaldatadefinitions.pdf">
      <def:title>Supplemental Data Definitions Document</def:title>
    </def:leaf>
```



Define.xml Opened when XSL is present in the Same Folder.

Firefox - Study XYZ123, Data Definitions

File: W:\Stat\184\184900\src\dev\prog\define.xml

Most Visited Getting Started

SDTM-IG 3.1.2

Annotated Case Report Form
Reviewers Guide
Complex Algorithms
Tabulation Datasets
Value Level Metadata
Controlled Terminology
Computational Algorithms
Comments

Date of document generation: 2014-04-09T15:04:35
Stylesheet version: 2013-04-24

Tabulation Datasets for Study XYZ123 (SDTM-IG 3.1.2)

Dataset	Description	Class	Structure	Purpose	Keys	Location	Documentation
TA	Trial Arms	Trial Design	One record per planned Element per Arm	Tabulation	STUDYID, ARM, TAETORD	../export/ta.xpt	
TE	Trial Elements	Trial Design	One record per planned Element	Tabulation	STUDYID, ETCD	../export/te.xpt	
TI	Trial Inclusion/Exclusion Criteria	Trial Design	One record per I/E criterion	Tabulation	STUDYID, IETESTCD	../export/ti.xpt	
TS	Trial Summary	Trial Design	One record per trial summary parameter value	Tabulation	STUDYID, TSSEQ	../export/ts.xpt	
TV	Trial Visits	Trial Design	One record per planned Visit per Arm	Tabulation	STUDYID, VISITNUM, ARM	../export/tv.xpt	
DM	Demographics	Special Purpose	Special Purpose - One record per event per subject	Tabulation	STUDYID, USUBID	../export/dm.xpt	See Reviewer's Guide, Section 2.1 Demographics Reviewers Guide
AE	Adverse Events	Events	Events - One record per event per subject	Tabulation	STUDYID, USUBID, AEDECOD, AESTDTC	../export/ae.xpt	
EX	Exposure	Interventions	One record per constant dosing interval per subject	Tabulation	STUDYID, USUBID, EXTRT, EXSTDTC	../export/ex.xpt	
LB	Laboratory Tests	Findings	Findings - One record per lab test per subject	Tabulation	STUDYID, USUBID, LBCAT, LBTESTCD, VISITNUM	../export/lb.xpt	
XP	Pain Scores	Findings	One record per subject per visit	Tabulation	STUDYID, USUBID, XPTESTCD, VISITNUM	../export/xp.xpt	
SUPPDM	DM - Supplemental Qualifiers	Supplemental Qualifier	Supplemental Qualifier - One record per qualifier	Tabulation		../export/supdm.xpt	

Go to the [top](#) of the define.xml



Variable Level Metadata

Microsoft Excel - CDISC_Domain_Field_Detail									
Type a question for help									
G8									
	A	B	C	E	F	G	H	I	J
1	DOMAIN	DOMAIN LABEL	VARIABLE	TYPE	LENGTH	FORMAT	ORIGIN	ROLE	COMMENT
2	DM	Demographics	STUDYID	text	7		Sponsor Defined	Identifier	Value will be 'ABC-123'.
3	DM	Demographics	DOMAIN	text	2		Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.
4	DM	Demographics	USUBID	text	21		Sponsor Defined	Identifier	Value will be study ID, site ID and subject ID concatenated together (ABC-123-SITEXX-XXXXXX).
5	DM	Demographics	SUBID	text	10		Sponsor Defined	Topic	Value will be the 6 digit subject ID from Demographics CRF.
6	DM	Demographics	RFSTDTC	text	19	ISO 8601	CRF Page 36	Record Qualifier	Lymphoseek first date of injection.
7	DM	Demographics	RFENDTC	text	19	ISO 8601	CRF Page 17	Record Qualifier	Completion or Withdrawn date as captured on the Final Status CRF.
8	DM	Demographics	SITEID	text	6		Sponsor Defined	Record Qualifier	Value will be verbatim text from Demographics CRF.
9	DM	Demographics	BRTHDTC	text	10	ISO 8601	CRF Page 9	Record Qualifier	Value will be verbatim text from Demographics CRF.
10	DM	Demographics	AGE	integer	8		CRF Page 9	Record Qualifier	Value will be AGE from Demographics CRF.
11	DM	Demographics	AGEU	text	5		Derived	Variable Qualifier	Value will be 'YEARS'.
12	DM	Demographics	SEX	text	1	SEX	CRF Page 9	Record Qualifier	Value will be GENDER from Demographics CRF.
13	DM	Demographics	RACE	text	41		CRF Page 9	Record Qualifier	Value will be RACE from Demographics CRF.
14	DM	Demographics	ETHNIC	text	12		CRF Page 9	Record Qualifier	Value will be ETHNIC from Demographics CRF.
15	DM	Demographics	ARMCD	text	10		Derived	Record Qualifier	Value will be 'Open-Label'.
16	DM	Demographics	ARM	text	23		Derived	Synonym Qualifier	Value will be '0.2 - 0.4 mL Lymphoseek'.
17	DM	Demographics	COUNTRY	text	3		Derived	Record Qualifier	Value will be 'USA'.
18	CO	Comments	STUDYID	text	7		Sponsor Defined	Identifier	Value will be 'ABC-123'.
19	CO	Comments	DOMAIN	text	2		Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.
20	CO	Comments	USUBID	text	21		Sponsor Defined	Identifier	Value will be study ID, site ID and subject ID concatenated together (ABC-123-SITEXX-XXXXXX).
21	CO	Comments	COSEQ	integer	8		Sponsor Defined	Identifier	Value will be sequence number.
22	CO	Comments	COREF	text	53		CRF Page 5	Record Qualifier	Value will be Form Name from Comments CRF.
23	CO	Comments	COVAL	text	200		CRF Page 5	Topic	Value will be verbatim comment.
24	CO	Comments	CODTC	text	10	ISO 8601	CRF Page 5	Timing	Value will be date of event.
25	CO	Comments	COIPT	text	19		Sponsor Defined	Timing	Value will be time point of event.
26	SV	Subject Visits	STUDYID	text	7		Sponsor Defined	Identifier	Value will be 'ABC-123'.
27	SV	Subject Visits	DOMAIN	text	2		Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.



Controlled Terminology :

Microsoft Excel - CDISC_Format

File Edit View Insert Format Tools Data Window Help

Times New Roman 10

A2 1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ORDER	FORMAT	CODE	VALUE										
2	1	SEX	F	FEMALE										
3	2	SEX	M	MALE										
4	3	SEX	U	UNKNOWN										
5	1	NY	Y	YES										
6	2	NY	N	NO										
7	1	RoleCodeList	GROUPING QUALIFIER	GROUPING QUALIFIER										
8	2	RoleCodeList	IDENTIFIER	IDENTIFIER										
9	3	RoleCodeList	RECORD QUALIFIER	RECORD QUALIFIER										
10	4	RoleCodeList	RESULT QUALIFIER	RESULT QUALIFIER										
11	5	RoleCodeList	SYNONYM QUALIFIER	SYNONYM QUALIFIER										
12	6	RoleCodeList	TIMING	TIMING										
13	7	RoleCodeList	TOPIC	TOPIC										
14	8	RoleCodeList	VARIABLE QUALIFIER	VARIABLE QUALIFIER										
15														
16														
17														
18														
19														
20														
21														
22														
23														
24														
25														
26														
27														
28														
29														
30														
31														
32														
33														
34														
35														
36														
37														

Sheet1

Ready



Value Level Metadata

	A	B	C	D	E	F	G	H
1	VARIABLE	VALUE	LABEL	TYPE	CONTROL TERMINOLOGY	ORIGIN	ROLE	
2	BIOCHEM	BIOLABRES	BIOLABRES	TEXT		CRF PAGE 1		
3	BIOCHEM EXTRA	HIGHVAL	HIGHVAL	TEXT		CRF PAGE 2		
4	BIOCHEM EXTRA	VALSTRX	VALSTRX	TEXT		CRF PAGE 2		
5	CREATCLR	ORJORES	ORJORES	TEXT		CRF PAGE 4		
6	DIAG	BLASTS	BLASTS	TEXT		CRF PAGE 5		
7	HAEM	BANDSRES	BANDSRES	TEXT		CRF PAGE 6		
8	HAEM	ORJORES	ORJORES	TEXT		CRF PAGE 6		
9	HAEM	POLYSRES	POLYSRES	TEXT		CRF PAGE 6		
10	HAEM EXTRA	CNTSTRX	CNTSTRX	TEXT		CRF PAGE 9		
11	HAEM EXTRA	LOWCNT	LOWCNT	TEXT		CRF PAGE 9		
12	HAEM HCT	HAEMPLABRES	HAEMPLABRES	TEXT		CRF PAGE 11		
13	LIVER	ORJORES	ORJORES	TEXT		CRF PAGE 12		
14	MTXHPLC	DAMPRES	DAMPRES	TEXT		CRF PAGE 13		
15	MTXHPLC	MTXRESC	MTXRESC	TEXT		CRF PAGE 13		
16	NON_HPLC	MTXMEASC	MTXMEASC	TEXT		CRF PAGE 13		
17	SRCREATIN	ORJORES	ORJORES	TEXT		CRF PAGE 16		
18	TOXGRADE	TOXGRADE	TOXGRADE	TEXT		CRF PAGE 17		
19	VITALS	DIASTOLIC	DIASTOLIC	TEXT		CRF PAGE 18		
20	VITALS	PULSE	PULSE	TEXT		CRF PAGE 18		
21	VITALS	RESPIRATORY	RESPIRATORY	TEXT		CRF PAGE 18		
22	VITALS	SYSTOLIC	SYSTOLIC	TEXT		CRF PAGE 18		
23	VITALS	TEMPCEL	TEMPCEL	TEXT		CRF PAGE 18		
24								
25								
26								
27								
28								
29								
30								
31								
32								
33								
34								
35								
36								
37								



Define-XML Document Structure

```
<?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="Studydisc01-Define-XML_2.0.0"
  CreationDateTime="2012-06-21T11:07:23-05:00"
  Originator="CDISC XML Technologies Team">
  <Study OID="disc01">
    <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) >
```



❖ Difference between V_1.0 and V2.0 of define.xml

Version : 1.0

- ❖ CDISC Study Data Tabulation Model Implementation Guide (SDTM-IG)
- ❖ Non- complaint structure.
- ❖ Missing or incomplete value level metadata(CRF, controlled terminology and derivations)
- ❖ Many deficiencies are due to limitations in Define.xml version 1.0

Version : 2.0

- ❖ Version 3.1.2 of the CDISC Study Data Tabulation Model Implementation Guide (SDTM-IG)
- ❖ Version 2.1 of the CDISC Analysis Dataset Model (ADaM)



Open CDISC Validator - Pinnacle 21

In this example, we are now ready to start the validation of the AE SDTM file. Click the start button.

When the validation is complete, you will receive an information window providing you with how long it took for the Validator to run, the number of records read, the number of datasets validated, the number of messages generated and the number of checks performed:

At this point, you may choose to view the report or start a new session. Choose 'View Report'.

The report

will consist of 4 tabs within the Excel document:

- a) Dataset Summary
- b) Issue Summary
- c) Details and Rules.

The Dataset Summary tab provides a brief overview of what was encountered by the



Open CDISC Validator - Pinnacle 21

continued...

❖ VALIDATING SDTM and ADAM FILES

- ❖ Open the 'opendisc-validator' folder
- ❖ Double click on the client.bat file. This will bring up the OpenCDISC Validator window:
- ❖ Browse and select Validate Data.
- ❖ Choose the Format the default is XPORT. Note that the SDTM files must be in SAS Transport. or a delimited file.
- ❖ Choose the source data by clicking on the Browse button on the right hand side. The following window will appear once you change the directory to a location that contains SAS XPT files
- ❖ Highlight the SDTM file or others files you want to validate.
- ❖ Click to Open, The OpenCDISC Validator window will appear with the files or files you have selected in the Source Data field
- ❖ Choose the configuration. The default for SDTM files is sdtm-3.1.2.xml.
- ❖ Choose the report format. The default is Excel



Open CDISC Validator - Pinnacle 21

continued...

- ❖ Validation against SDTM 1.4 model and SDTMIG v3.2
- ❖ Validation of Define-XML v2.0
- ❖ Support for Dataset-XML format
- ❖ Updated SEND config for FDA compliance
- ❖ Various bug fixes and minor engine enhancements.
- ❖ Various updates to SDTM 3.1.3, 3.1.2, 3.1.1, and SEND 3.



