

Converting the define.xml to a Relational Database to enable Printing and Validation

Lex Jansen

Octagon Research Solutions, Inc.

Leading the Electronic Transformation of Clinical R&D

**PhUSE 2009,
Basel, Switzerland**

- Regulatory landscape
- Data Definition Tables
- define.pdf -- define.xml
- define.xml in SDTM / ADaM CDISC pilot (printing issue)
- define.xml as a relational data model
- Validating the define.xml
- Example define.pdf from FDA/CDISC Pilot II



Regulatory landscape



- July 2004 – FDA adds Study Data Specifications v1.0 to draft eCTD Guidance. This specification references the CDISC SDTM for data tabulation datasets

Electronic Common Technical Document (eCTD)

- Draft Guidance for Industry on Providing Regulatory Submissions in Electronic Format--Human Pharmaceutical Applications and Related Submissions. (Posted 8/28/2003)
 - *Federal Register Notice* [\[TXT\]](#) [\[PDF\]](#)
 - [The Draft Guidance](#) 

Specifications

- [eCTD Backbone Files Specification for Module 1](#) 
- [eCTD Backbone File Specification for Modules 2 through 5](#) 
- [eCTD Backbone File Specification for Study Tagging Files](#) 
- [FDA eCTD Table of Contents Headings and Hierarchy](#) 
- [Study Data Specifications](#)  **New!!** (Posted 7/21/2004)



- March 2005 – Study Data Specifications v1.1: Updates Specifications for Data Set Documentation, consisting of:
 - data definitions
 - annotated case report forms (CRFs)
- *“The specification for the data definitions for datasets provided using the CDISC SDTM is included in the Case Report Tabulation Data Definition Specification (**define.xml**) developed by the CDISC define.xml Team”*
- Data Definition for other data sets follows:
*Providing Regulatory Submissions in Electronic Format – NDA (1999), which is the **define.pdf***



- 2006 – CDISC SDTM / ADaM Pilot Project:
Collaborative Pilot project with FDA and industry to test how well the submission of CDISC compliant data sets and associated metadata meets the needs of both medical and statistical FDA reviewers
- Generation of ICH E3/eCTD clinical study report (CSR) using the CDISC data models
- Data Definition Tables were provided in XML format (CRT-DDS, **define.xml**)



- April 2006 – FDA issues final “Guidance for Industry: Providing Regulatory Submissions in Electronic Format — Human Pharmaceutical Product Applications and Related Submissions Using the eCTD Specifications”
- Application Table of Contents: XML instead of PDF
- This guidance now has the following reference for datasets:
*See the associated document "Study Data Specifications" for details on providing datasets and related files (e.g., data definition file (**define.xml**), program files)*



- After December 31, 2007 eCTD is the preferred format for electronic submissions going to CDER:
 - consistent with FDA's technical capabilities
 - more efficient than other choices



To Summarize:

- Sponsors are submitting clinical study data in electronic format to the FDA
- Currently version 5 SAS transport files (XML in the future)
- Data Definition file helps reviewers to understand data
- Format for **Data Definition** file has changed from PDF to XML



Data Definition Tables in PDF



- 1999 Guidance: sponsor has to document submitted data by including data definition tables ([define.pdf](#)) and annotated case report forms ([blankcrf.pdf](#))

Datasets For Study 1234		
Dataset	Description of dataset	Location
DEMO	Demographics	crt/datasets/1234/demo.xpt
INCLUDE	Inclusion criteria	crt/datasets/1234/include.xpt



- 1999 Guidance: sponsor has to document submitted data by including data definition tables ([define.pdf](#)) and annotated case report forms ([blankcrf.pdf](#))

Study 1234 - Demographics				
Variable	Label	Type	Codes	Comments
PATID	Patient identification	char		
STUDY	Study number	char		
CENTER	Study center	char		
TRT	Assigned treatment group	num	0= placebo 5= 5mg/day	
SEX	Sex of subject	char	f = female m = male	
BDATE	Birth date	date		Demographics page 1
WEIGHT	Weight in kg	num		



Data Definition Tables in XML



- Example from the CDISC SDS Metadata Team (2005)

Datasets for Study 1234					
Dataset	Description	Structure	Purpose	Keys	Location
DM	Demographics	Special Purpose - One record per event per subject	Tabulation	STUDYID, USUBJID	crt/datasets/1234/dm.xpt
TE	Trial Elements	Trial Design - One Record Per Element	Tabulation	STUDYID, ELEMENT	crt/datasets/1234/te.xpt
TA	Trial Arms	Trial Design - One Record per Element for each Arm	Tabulation	STUDYID, ARM	crt/datasets/1234/ta.xpt
TV	Trial Visits	Trial Design - One Record per Visit per Arm	Tabulation	STUDYID, VISIT	crt/datasets/1234/tv.xpt
SE	Subject Elements	Study Design - One Record Per Subject Element	Tabulation	STUDYID, ELEMENT	crt/datasets/1234/se.xpt
SV	Subject Visits	Study Design - One Record Per Subject Visit	Tabulation	STUDYID, VISIT	crt/datasets/1234/sv.xpt
EX	Exposure	Interventions - One record per constant dosing interval per subject	Tabulation	USUBJID, EXTRT, EXSEQ	crt/datasets/1234/ex.xpt
CM	Concomitant Medications	Interventions - One record per event per subject	Tabulation	USUBJID, CMTRT, CMSEQ	crt/datasets/1234/cm.xpt
SU	Substance Use	Interventions - One record per substance use type per subject	Tabulation	USUBJID, SUTRT, SUSEQ	crt/datasets/1234/su.xpt
AE	Adverse Events	Events - One record per event per subject	Tabulation	USUBJID, AETERM, AESEQ	crt/datasets/1234/ae.xpt
DS	Disposition	Events - One record per disposition status or protocol milestone per subject	Tabulation	USUBJID, DSTERM, DSSEQ	crt/datasets/1234/ds.xpt
MH	Medical History	Events - One record per event per subject	Tabulation	USUBJID, MHTERM, MHSEQ	crt/datasets/1234/mh.xpt
EG	ECG Test Results	Findings - One record per event per subject	Tabulation	USUBJID, EGTESTCD, EGSEQ	crt/datasets/1234/eg.xpt
IE	Inclusion/Exclusion Exceptions	Findings - One record per event per subject	Tabulation	USUBJID, IETEST, IESEQ	crt/datasets/1234/ie.xpt
LB	Laboratory Tests	Findings - One record per lab test per subject	Tabulation	USUBJID, LBTESTCD, VISITNUM, TPTNUM, LBSEQ	crt/datasets/1234/lb.xpt

- Example from the CDISC SDS Metadata Team (2005)

Adverse Events Dataset (AE)

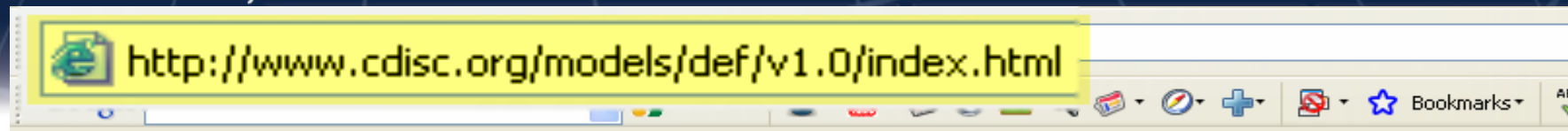
Variable	Label	Type	Controlled Terms or Format	Origin	Role	Comment
STUDYID	STUDY IDENTIFIER	text		CRF Page	Identifier	Demographics CRF Page 4
DOMAIN	DOMAIN ABBREVIATION	text		CRF Page	Identifier	DOMAIN ABBREVIATION
USUBJID	UNIQUE SUBJECT IDENTIFIER	text		CRF Page	Identifier	Demographics CRF Page 4
AESEQ	SEQUENCE NUMBER	integer		CRF Page	Identifier	SEQUENCE NUMBER
AESPID	SPONSOR ID	text		DERIVED	Identifier	SPONSOR ID
AETERM	REPORTED TERM FOR THE ADVERSE EVENT	text		CRF Page	Topic	Adverse Events CRF Page 16
AEDECOD	DICTIONARY-DERIVED TERM	text		DERIVED	Synonym Qualifier	Preferred text (COSTART dictionary)
AEBODSYS	BODY SYSTEM OR ORGAN CLASS	text		DERIVED	Result Qualifier	Body system text
AESEV	SEVERITY/INTENSITY	text		CRF Page	Record Qualifier	Adverse Events CRF Page 16
AESER	SERIOUS EVENT	text		CRF Page	Record Qualifier	Adverse Events CRF Page 16
AEACN	ACTION TAKEN WITH STUDY TREATMENT	text		CRF Page 19	Record Qualifier	Adverse Events CRF Page 16
AEACNOTH	OTHER ACTION TAKEN	text		CRF Page	Record Qualifier	Adverse Events CRF Page 16
AEREL	CAUSALITY	text		CRF Page	Result Qualifier	Adverse Events CRF Page 16
AERELNST	RELATIONSHIP TO NON-STUDY TREATMENT	text		CRF Page	Result Qualifier	Adverse Events CRF Page 16

What is the define.xml



- Case Report Tabulation Data Specification (CRT-DDS, define.xml)
- Production version: **1.0.0**
- Based on version ODM version **1.2.1**
- Maintained by CDISC's **XML Technologies Team** (formerly known as the ODM team)
- New version of define.xml expected in 2009 with additional metadata for SDTM and supporting SDTM V3.1.2





CLINICAL DATA INTERCHANGE STANDARDS CONSORTIUM

search:

go

[standards](#) | [education/training](#) | [members-only](#) | [registered solutions providers](#)

Standards

FDA Adds CDISC ODM Define.xml to Study Data Specifications

The FDA has now included the CDISC Case Report Tabulation Data Definition Specification (define.xml), which is based on the CDISC ODM, as part of the eCTD Study Data Specifications for the eCTD for submissions using the SDTM. The revised specifications are available at

<http://www.fda.gov/cder/regulatory/ersr/ectd.htm>.

(posted 7/13/2005)

Case Report Tabulation Data Definition Specification (CRT-DDS, also called define.xml)

Final Version 1.0

CRT-DDS Released for Implementation February 10, 2005.

The CDISC define.xml Team has published the Case Report Tabulation Data Definition Specification (define.xml) Version 1.0 for implementation, which supersedes all prior versions. Version 1.0 reflects changes from a comment period through the Health Level 7 (HL7) Regulated Clinical Research Information Management Technical Committee (RCRIM) in December 2003 (www.hl7.org) and CDISC's website in September 2004 as well as the work done by the define.xml team in conjunction with the CDISC ODM team to add functionality, features, and additional documentation.

Case Report Tabulation Data Definition Specification (define.xml)

Prepared by the
CDISC define.xml Team

Principal Editor: William Qubeck
Principal Contributors: Sally Cassells, Anthony Friebel, and the define.xml team

Notes to Readers

This version of the Case Report Tabulation Data Definition Specification supersedes all prior versions. Version 1.0.0 reflects changes from a comment period through the Health Level 7 (HL7) Regulated Clinical Research Information Management Technical Committee (RCRIM) in December 2003 (www.hl7.org) and CDISC's website in September 2004 as well as the work done by the define.xml team to add functionality, features, and additional documentation.

Version 1.0.0 incorporated the applicable comments, suggestions, and corrections received from the two comment periods specified above and is the official implementation version.

Revision History

Date	Version	Summary of Changes	Primary Authors
2005-02-03	1.0.0	This is the official implementation version of the Case Report Tabulation Data Definition specification.	The define.xml team
2005-02-09	1.0.0	Administrative update.	Anthony Friebel, William Qubeck, Sally Cassells

Case Report Tabulation Data Definition Specification (define.xml)
Standards Version 1.0.0

Page 1 of 45



Clinical Data Interchange Standards Consortium

Specification for the Operational Data Model (ODM)

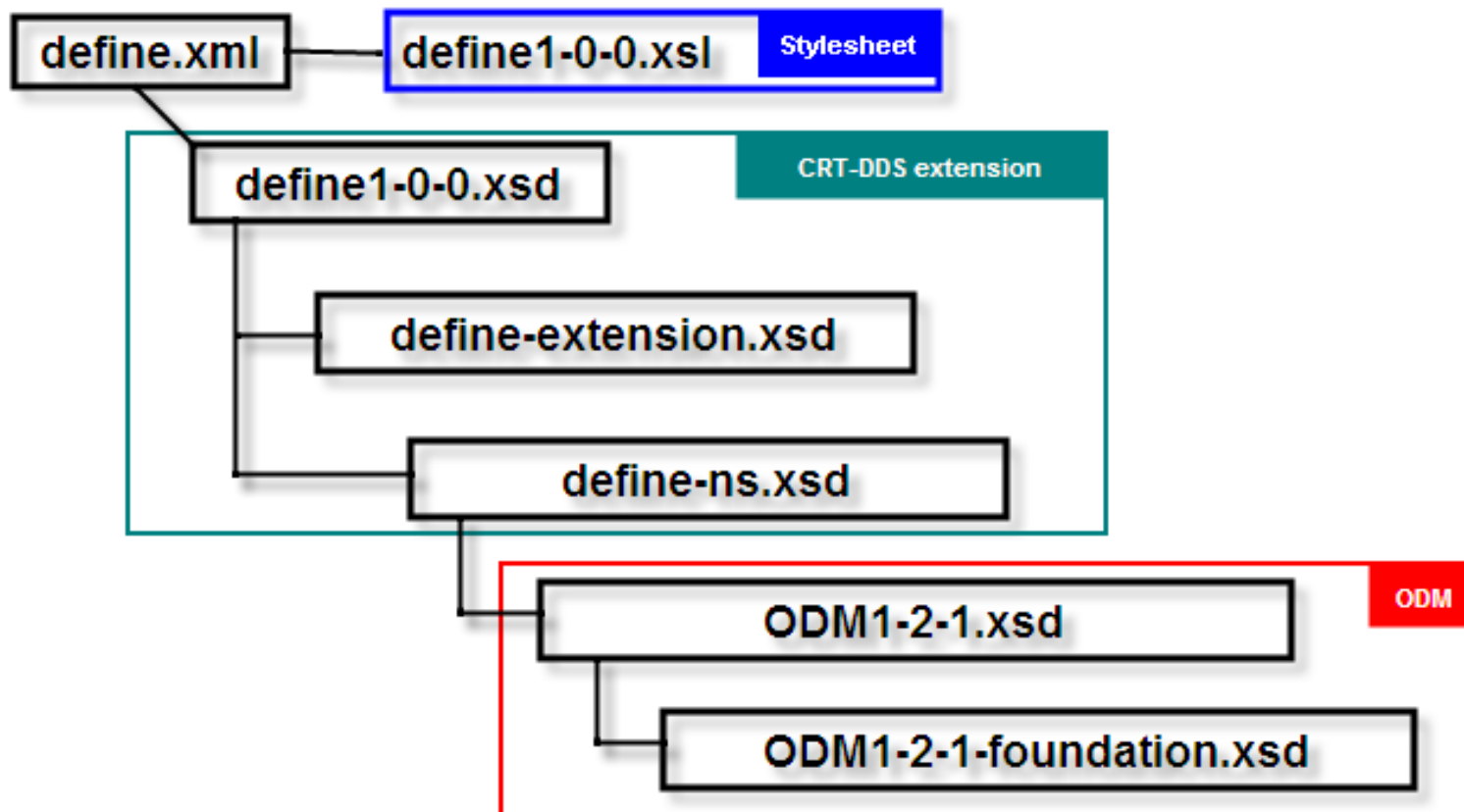
Version 1.2
Source File: ODM-1.2.5.adtd
Last Update: 19 Dec 2003 10:10 AM

Copyright © CDISC 2003 This document is the property of CDISC Inc. This document can be freely used and reproduced without limitation as long as (1) it is not modified, and (2) the entire copyright statement is included in the copy. Modifications to this document can only be made with written consent of CDISC Inc.

An official copy of this document is available at <http://www.cdisc.org/models/odm/v1.2/ODM1-2-0.html>

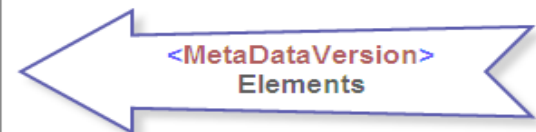
Table of Contents

- [1 Introduction \(non-normative\)](#)
- [2 General Issues](#)
 - [2.1 Content of the Standard](#)
 - [2.2 File Conformity](#)
 - [2.3 System Conformity](#)
 - [2.4 Vendor Extensions](#)
 - [2.5 Changes from Previous Versions \(non-normative\)](#)
 - [2.6 Entities and Elements](#)
 - [2.7 Clinical Data Keys](#)
 - [2.8 Single Files and Collections](#)
 - [2.9 Transactions](#)
 - [2.10 Element Ordering](#)
 - [2.11 Element Identifiers and References](#)
 - [2.12 Syntax Notation](#)
 - [2.13 Data Formats](#)
- [3 General Elements](#)
 - [3.1 ODM](#)
 - [3.2 Study](#)
 - [3.3 Global Variables](#)
 - [3.4 StudyName](#)
 - [3.5 StudyDescription](#)
 - [3.6 ProtocolName](#)
 - [3.7 Basic Definitions](#)
 - [3.8 MeasurementUnit](#)
 - [3.9 Symbol](#)
 - [3.10 TranslatedText](#)
- [4 Metadata Elements](#)



① <?xml version="1.0" encoding="ISO-8859-1" ?>
<?xml-stylesheet type="text/xsl" href="define1-0-0.xsl"?>

② <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="Studydisc01"
ODMVersion="1.2" FileType="Snapshot"
CreationDateTime="2007-04-09T12:24:09">
<Study OID="cdisc01">
③ <GlobalVariables>
④ <StudyName>CDISC01</StudyName>
<StudyDescription>CDISC01 Test Study.</StudyDescription>
<ProtocolName>CDISC01</ProtocolName>
</GlobalVariables>
<MetaDataVersion OID="CDISC.SDTM.3.1.1"
⑤ Name="Study CDISC01, Data Definitions"
Description="Study CDISC01, Data Definitions"
def:DefineVersion="1.0.0"
def:StandardName="CDISC SDTM"
def:StandardVersion="3.1.1">
⑥
< def:AnnotatedCRF >
< def:SupplementalDoc >
< def:leaf >
< def:ComputationMethod >
< def:ValueListDef >
< ItemGroupDef >
< ItemDef >
< CodeList >
</MetaDataVersion>
</Study>
</ODM>



Displaying the define.xml



- From the CDISC SDS Metadata Team (2007):
define.xml + XSL style sheet = html

Study CDISC01, Data Definitions

[Annotated Case Report Form](#)
[Datasets](#)
[Value Level Metadata](#)
[Computational Algorithms](#)
[Controlled Terminology](#)

Datasets for Study CDISC01						
Dataset	Description	Class	Structure	Purpose	Keys	Location
SV	Subject Visits	TRIAL DESIGN	One record per subject per actual visit	Tabulation	STUDYID, USUBJID, SVSTDTC, SVENDTC, VISITNUM, SVUPDES	sv.xpt
TA	Trial Arms	TRIAL DESIGN	One record per element per arm	Tabulation	STUDYID, ARMCD, TAETORD	ta.xpt
TE	Trial Elements	TRIAL DESIGN	One record per element	Tabulation	STUDYID, ETCD	te.xpt
TI	Trial Inclusion	TRIAL DESIGN	One record per I/E criterion	Tabulation	STUDYID, IETESTCD	ti.xpt
TS	Trial Summary	TRIAL DESIGN	One record per parameter value	Tabulation	STUDYID, TSPARMCD, TSVAL	ts.xpt
TV	Trial Visits	TRIAL DESIGN	One record per planned visit per arm	Tabulation	STUDYID, VISITNUM, ARMCD	tv.xpt
CO	Comments	SPECIAL PURPOSE	One record per comment per subject	Tabulation	STUDYID, USUBJID, COSEQ, IDVAR, IDVARVAL	co.xpt
DM	Demographics	SPECIAL PURPOSE	One record per subject	Tabulation	STUDYID, USUBJID	dm.xpt
CM	Concomitant Medications	INTERVENTIONS	One record per medication intervention episode	Tabulation	STUDYID, USUBJID, CMCAT, CMTRT, CMSTDTC, CMENDTC,	cm.xpt

- From the CDISC SDS Metadata Team (2007):
define.xml + XSL style sheet = html

Study CDISC01, Data Definitions

Home | RSS | Print | Page | Tools

Annotated Case Report Form

Datasets

- Subject Visits (SV)
- Trial Arms (TA)
- Trial Elements (TE)
- Trial Inclusion (TI)
- Trial Summary (TS)
- Trial Visits (TV)
- Comments (CO)
- Demographics (DM)
- Concomitant Medications (CM)
- Exposure (EX)
- Adverse Events (AE)
- Disposition (DS)
- Medical History (MH)
- Drug Accountability (DA)
- ECG (EG)
- Inclusion/Exclusion Exceptions (IE)
- Laboratory Tests (LB)
- Physical Exam (PE)
- Questionnaires (QS)
- Subject Characteristics (SC)
- Vital Signs (VS)
- Supplemental Qualifier (SUPPAE)
- Supplemental Qualifier (SUPPCM)
- Supplemental Qualifier

Concomitant Medications Dataset (CM)						cm.xpt
Variable	Label	Type	Controlled Terminology	Origin	Role	Comment
STUDYID	Study Identifier	text		CRF Page 3	IDENTIFIER	
DOMAIN	Domain Abbreviation	text		DERIVED	IDENTIFIER	
USUBJID	Unique Subject Identifier	text		SPONSOR DEFINED	IDENTIFIER	Concatenation of STUDYID.SUBJID
CMSEQ	Sequence Number	integer		DERIVED	IDENTIFIER	Sequential number uniquely identifying the records within the domain by USUBJID and generated using the key sequence from the domain level metadata
CMSPID	Sponsor-Defined Identifier	text		SPONSOR DEFINED	IDENTIFIER	ID of original SAS dataset
CMTRT	Reported Name of Drug, Med, or Therapy	text		CRF Pages 6 , 35	TOPIC	
CMMODIFY	Modified Reported Name	text		SPONSOR DEFINED	SYNONYM QUALIFIER	
CMDECOD	Standardized Medication Name	text	DRUGDICT_F	DERIVED	SYNONYM QUALIFIER	WHODRUG Version 2002/04
CMCAT	Category for Medication	text	CMCAT_F	CRF Pages 6 , 35	GROUPING QUALIFIER	Sponsor controlled terminology

- From the CDISC SDS Metadata Team (2007):
define.xml + XSL style sheet = html

Study CDISC01, Data Definitions

- Annotated Case Report Form - Datasets - Value Level Metadata - Computational Algorithms - Controlled Terminology - Code Lists - AEACTF - AERELF - AESEVRF - CFREQF - CGIIMP - CMCAT_F - CROUTEF - CSIND_F - CUNITF - DSCAT_F - DSDECOD_F - EDUCATNF - ETHNIC2F - FAMILYF - GENSYMP - HLN - IECAT_F - LBCAT_F - MURDRCYC_F	<table> <tr> <th colspan="2">CMCAT_F, Reference Name (CMCAT_F)</th></tr> <tr> <th>Code Value</th><th>Code Text</th></tr> <tr> <td>Concomitant Medications</td><td>Concomitant Medications</td></tr> <tr> <td>Psychotropic Drug Treatment History</td><td>Psychotropic Drug Treatment History</td></tr> <tr> <th colspan="2">CROUTEF, Reference Name (CROUTEF)</th></tr> <tr> <th>Code Value</th><th>Code Text</th></tr> <tr> <td>PO</td><td>PO</td></tr> <tr> <td>SC</td><td>SC</td></tr> <tr> <td>IM</td><td>IM</td></tr> <tr> <td>IV</td><td>IV</td></tr> <tr> <td>Rectal</td><td>Rectal</td></tr> <tr> <td>Topical</td><td>Topical</td></tr> <tr> <td>Nasal</td><td>Nasal</td></tr> <tr> <td>Inhaled</td><td>Inhaled</td></tr> <tr> <td>SL</td><td>SL</td></tr> <tr> <td>Other</td><td>Other</td></tr> </table>	CMCAT_F, Reference Name (CMCAT_F)		Code Value	Code Text	Concomitant Medications	Concomitant Medications	Psychotropic Drug Treatment History	Psychotropic Drug Treatment History	CROUTEF, Reference Name (CROUTEF)		Code Value	Code Text	PO	PO	SC	SC	IM	IM	IV	IV	Rectal	Rectal	Topical	Topical	Nasal	Nasal	Inhaled	Inhaled	SL	SL	Other	Other
CMCAT_F, Reference Name (CMCAT_F)																																	
Code Value	Code Text																																
Concomitant Medications	Concomitant Medications																																
Psychotropic Drug Treatment History	Psychotropic Drug Treatment History																																
CROUTEF, Reference Name (CROUTEF)																																	
Code Value	Code Text																																
PO	PO																																
SC	SC																																
IM	IM																																
IV	IV																																
Rectal	Rectal																																
Topical	Topical																																
Nasal	Nasal																																
Inhaled	Inhaled																																
SL	SL																																
Other	Other																																

Define.xml in the SDTM/ADaM CDISC pilot I



- SDTM / ADaM Pilot (published 2008): adds **Analysis** metadata
[define.xml](#) + XSL style sheet = html

	Links
	Reviewer's Guide
	Annotated Case Report Form
	Analysis Results Metadata
	Analysis Datasets
	SDTM Datasets
	Computational Algorithms
	Code Lists
	Discrete Value Listings

Links for Study CDISC_Pilot

[Reviewer's Guide](#)

[Analysis Results Metadata](#)

[Analysis Datasets](#)

[SDTM Datasets](#)

Analysis Results Metadata (Summary) for Study CDISC_Pilot

[Table 14-1.01 - Summary of Populations](#)

[Table 14-1.02 - Summary of End of Study Data](#)

[Table 14-1.03 - Summary of Number of Subjects by Site](#)

[Table 14-2.01 - Summary of Demographic and Baseline Characteristics](#)

[Table 14-3.01 - Primary Endpoint Analysis: ADAS Cog \(11\) - Change from Baseline to Week 24 -- LOCF](#)

[Table 14-3.02 - Primary Endpoint Analysis: CIBIC+ - Summary at Week 24 -- LOCF](#)

[Table 14-3.03 - ADAS Cog \(11\) - Change from Baseline to Week 8 -- LOCF](#)

[Table 14-3.04 - CIBIC+ - Summary at Week 8 -- LOCF](#)

- SDTM / ADaM Pilot (published 2008): adds **Analysis** metadata
[define.xml](#) + XSL style sheet = html

Analysis Results Metadata

- Table 14-1.01
- Table 14-1.02
- Table 14-1.03
- Table 14-2.01
- Table 14-3.01
- Table 14-3.02
- Table 14-3.03
- Table 14-3.04
- Table 14-3.05
- Table 14-3.06
- Table 14-3.07
- Table 14-3.08
- Table 14-3.09
- Table 14-3.10
- Table 14-3.11
- Table 14-3.12
- Table 14-3.13
- Table 14-4.01
- Table 14-5.01
- Table 14-5.02
- Table 14-6.01
- Table 14-6.02
- Table 14-6.03
- Table 14-6.04
- Table 14-6.05

Analysis	Table 14-6.01 - Summary Statistics for Continuous Laboratory Values
Description	Summary statistics for each laboratory parameter (numeric results), provided for each scheduled visit, Safety population
Reason	pre-specified in SAP
Data References	Hematology Lab Analysis (ADLBH) [where SAFETY='Y'], Chemistry Lab Analysis (ADLBC) [where SAFETY='Y']
Documentation	SAP Section 11.5 , SAP Template 15 , Variables summarized are LBSTRESN and CHSTRESN

Go to the [Analysis Results Metadata Summary](#)

Analysis	Table 14-6.02 - Frequency of Normal and Abnormal (Beyond Normal Range) Laboratory Values During Treatment
Description	For each lab parameter, counts subjects based on most extreme value during treatment - categorized by whether that value is below, above, or within normal range, Safety population
Reason	pre-specified in SAP
Data References	Hematology Lab Analysis (ADLBH) [where SAFETY='Y'], Chemistry Lab Analysis (ADLBC) [where SAFETY='Y']
Documentation	SAP Section 11.5 , SAP Template 16 , To select a single record for summary use LBTRMXFL. This flag is programmed so that the record with the flag set to 'Y' is the most extreme value for the subject. Therefore if that record is within normal range then no other visits for the subject are outside normal range. The normal range flag variable is LBNRIND.

- SDTM / ADaM Pilot (published 2008): adds **Analysis** metadata
[define.xml](#) + XSL style sheet = html

Hematology Lab Analysis Dataset (ADLBH)							adlbh.xpt
Variable	Label	Type	Controlled Terms or Format	Computational Algorithm or Method	Origin	Role	Comment
USUBJID	Unique Subject Identifier	text			DM	Identifier	Data from ADSL
VISITNUM	Visit Number	float			LB	Timing	Data from LB
LBTESTCD	LAB Test or Examination Short Name	text	LBTESTCD		LB	Identifier	Data from LB
STUDYID	Study Identifier	text			DM	Identifier	Data from ADSL
DOMAIN	Domain Abbreviation	text		COMP ADLBH DOMAIN	created here	Identifier	Data from ADLBH
SUBJID	Subject Identifier for the Study	text			DM	Identifier	Data from ADSL
VISIT	Visit Name	text			LB	Timing	Data from LB
ANLDY	Analysis Day	float		COMP ADLBH ANLDY	created here	Timing	Data from ADLBH , ADSL
IRDT	Date of	date		COMP ADLBH IRDT	created	Timing	Data from LB

- CDISC SDTM / ADaM Pilot project report:

*“A major **issue** identified by the regulatory review team was the **difficulty in printing the Define file**. The style sheet used in the pilot submission package was developed with the **primary target of web browser rendering**, which is not readily suited to printing. Reviewers who attempted to print the Define file found that the file did not fit on portrait pages, that page breaks were not clean, and that printing only a portion of the file was difficult. Opening the document in another application (e.g., Microsoft Word) provided a work-around, but was not an option that was user friendly or efficient.”*



- CDISC SDTM / ADaM Pilot project report:

*“This problem could be viewed as an **implementation issue** that sponsors will need to handle, after discussing the issue with their FDA reviewers. For example, a sponsor might choose to provide two versions of the style sheet – **XML for viewing** and **PDF for printing**. Ideally, a reminder of the issue would be included somewhere in the CRT-DDS guidance (e.g., a note that consideration be given to how the sponsor will respond to a request from reviewers for a print-friendly version of the style sheet). It should be noted that the **regulatory review team for the pilot project emphasized that the ability to print the document would be essential for the future use of XML files.**”*



Printing the define.xml



- The **PDF** format is the *de facto standard* for **printable** documents on the web
- PDF is platform independent (no browser issues ...)
- How can we create a PDF file from the define.xml ???



How to create the PDF rendition of define.xml ?

- Original metadata in SAS → SAS ODS PDF → define.pdf
- *But what if we only have the SAS .XPT files and the define.xml ?*
- Use XML based tools to convert XML to PDF (FOP, XSL Formatting Objects Processor)
- *Possible, but very complicated to develop in-house when not familiar with XML technology.*



SAS Solution:

- Convert the XML hierarchy to a relational data model in the form of (2-dimensional) SAS data sets
- Once we have the define.xml content in SAS datasets, we can use SAS to create a PDF rendition (with ODS PDF)



- **How to Convert** the XML hierarchy to a relational data model in the form of (2-dimensional) SAS data sets
- **Solution: SAS XML Mapper**
- **SAS XML Mapper:**
free stand-alone Java client application available on the SAS product distribution disks
- Uses XPATH to create a MAP file that maps hierarchical XML to rows and columns in SAS



- relations in the define.xml

```
<ItemGroupDef OID="MH"
  Name="MH" Repeating="Yes" IsReferenceData="No"
  Purpose="Tabulation" def:Label="Medical History"
  def:Structure="One record per medical history event per subject"
  def:DomainKeys="STUDYID, USUBJID, MHCAT, MHTERM, MHSTDTC"
  def:Class="EVENTS" def:ArchiveLocationID="Location.AE">
  <ItemRef ItemOID="STUDYID" OrderNumber="1" Mandatory="Yes"
    Role="IDENTIFIER" RoleCodeListOID="RoleCodeList" />
```

...

```
<def:leaf ID="Location.MH" xlink:href="mh.xpt">
  <def:title>mh.xpt</def:title>
</def:leaf>
</ItemGroupDef>
```

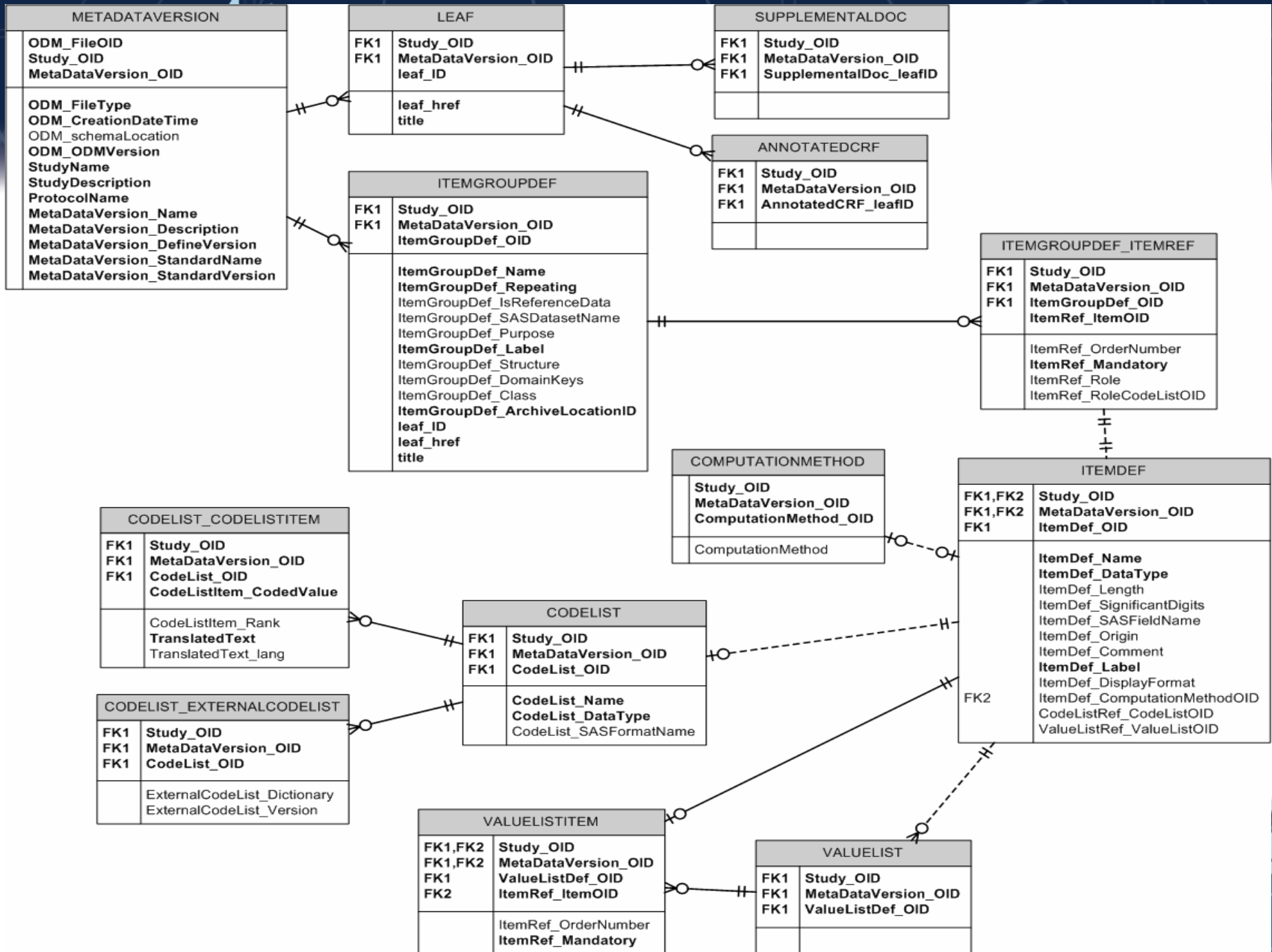
...

```
<ItemDef OID="STUDYID" Name="STUDYID" DataType="text" Length="7"
  Origin="CRF Page 3" def:Label="Study Identifier" />
```



define.xml as a relational model





SAS XML Mapper



XML Mapper

File Tools Help

Condensed Full Schema

define.xml structure

ItemDef [389] {389}

Attributes [1] {389}

OID [1] {389}

Name [1] {389}

DataType [1] {389}

Length [1] {389}

Origin [1] {389}

Comment [1] {107}

def.Label [1] {389}

SignificantDigits [1] {14}

def.ComputationMethodOID [1] {2}

CodeListRef [1] {62}

Attributes [1] {62}

CodeListOID [1] {62}

def.ValueListRef [1] {19}

Attributes [1] {19}

ValueListOID [1] {19}

CodeList [29] {29}

Properties Format Condition Enumeration Class XMLMap Settings Output

Name ItemDef_OID

Description OID

Path /ODM/Study/MetaDataVersion/ItemDef/@OID

End Path

Begin/End

Retain Replace

ItemGroupDef

ItemGroupDef_ItemRef

ItemDef

Study_OID

MetaDataVersion_OID

ItemDef_OID

ItemDef_Name

ItemDef_DataType

ItemDef_Length

ItemDef_SignificantDigits

ItemDef_SASFieldName

ItemDef_Origin

ItemDef_Comment

ItemDef_Label

ItemDef_DisplayFormat

ItemDef_ComputationMethodOID

CodeListRef_CodeListOID

ValueListRef_ValueListOID

SAS data set Metadata

XML source XML Schema source XMLMap SAS Code Example Table view Contents Validate Log

Table: ItemDef Row: 24 / 389 Columns: 6 / 15

SAS formats and informats are not applied to this view.

	ItemDef_OID	ItemDef_Name	ItemDef_D...	ItemDef_Le...	ItemDef_S...	ItemDef...	ItemDef_Origin	ItemDef_Comment
16	SUPPVS.QNAM.VSCSI...	VSCSIND	text	2			DERIVED	Only created if value qualifies as pote
17	AE.AESEQ	AESEQ	integer	1			DERIVED	Sequential number uniquely identifiyr
18	AE.AESPID	AESPID	text	2			SPONSOR DEFINED	ID of original SAS dataset
19	AE.AETERM	AETERM	text	25			CRF Page 34	
20	AE.AEMODIFY	AEMODIFY	text	9			SPONSOR DEFINED	
21	AE.AEDECOD	AEDECOD	text	18			DERIVED	MedDRA version 8.0
22	AE.AEBODSYS	AEBODSYS	text	52			DERIVED	MedDRA version 8.0
23	AE.AESEV	AESEV	text	8			CRF Page 34	

XMLMap loaded: DefineXML_v01.map

0 37 0 2 13 96

Define.xml -> SAS datasets

itemgroupdef.sas7bdat

	Study_OID	MetaDataVersion_OID	ItemGroupDef_OID	ItemGroupDef_Name	ItemGroupDef_Repeated	ItemGroupDef_Updated	ItemGroupDef_Purpose	ItemGroupDef_Label	ItemGroupDef_Structure	ItemGroupDef_DomainKeys	ItemGroupDef_Class	ItemGroupDef_Ar
11	cdisc01	CDISC.SDTM.3.1.1	AE	AE	Yes	No	Tabulation	Adverse Events	One record per	STUDYID, USUBJID	EVENTS	Locati
12	cdisc01	CDISC.SDTM.3.1.1	DS	DS	Yes	No	Tabulation	Disposition	One record per	STUDYID, USUBJID	EVENTS	Locati
13	cdisc01	CDISC.SDTM.3.1.1	MH	MH	Yes	No	Tabulation	Medical History	One record per	STUDYID, USUBJID	EVENTS	Locati
14	cdisc01	CDISC.SDTM.3.1.1	DA	DA	Yes	No	Tabulation	Drug Accountability	One record per	STUDYID, USUBJID	FINDINGS	Locati
15	cdisc01	CDISC.SDTM.3.1.1	EG	EG	Yes	No	Tabulation	ECG	One record per	STUDYID, USUBJID	FINDINGS	Locati

itemgroupdef_itemref.sas7bdat

	Study_OID	MetaDataVersion_OID	ItemGroupDef_OID	ItemRef_ItemOID	ItemRef_OrderNumber	ItemRef_Mandatory	ItemRef_Role	ItemRef_RoleCodeListOID
107	cdisc01	CDISC.SDTM.3.1.1	AE	DOMAIN	2	Yes	IDENTIFIER	RoleCodeList
108	cdisc01	CDISC.SDTM.3.1.1	AE	USUBJID	3	Yes	IDENTIFIER	RoleCodeList
109	cdisc01	CDISC.SDTM.3.1.1	AE	AE.AESEQ	4	Yes	IDENTIFIER	RoleCodeList
110	cdisc01	CDISC.SDTM.3.1.1	AE	AE.AESPID	5	No	IDENTIFIER	RoleCodeList
111	cdisc01	CDISC.SDTM.3.1.1	AE	AE.AETERM	6	Yes	TOPIC	RoleCodeList
112	cdisc01	CDISC.SDTM.3.1.1	AE	AE.AEMODIFY	7	No	SYNONYM QUALIFI	RoleCodeList
113	cdisc01	CDISC.SDTM.3.1.1	AE	AE.AEDECOD	8	Yes	SYNONYM QUALIFI	RoleCodeList

itemdef.sas7bdat

	Study_OID	MetaDataVersion_OID	ItemDef_OID	ItemDef_Name	ItemDef_DataType	ItemDef_Length	ItemDef_Origin	ItemDef_Comment	ItemDef_Label	ItemDef_DisplayFormat	ItemDef_ComputationMethodOID	CodeListRef_CodeListOID
18	cdisc01	CDISC.SDTM.3.1.1	AE.AESPID	AESPID	text	2	SPONSOR DEFINED	ID of orig	Sponsor-Defined Ident			
19	cdisc01	CDISC.SDTM.3.1.1	AE.AETERM	AETERM	text	25	CRF Page 34		Reported Term for the			
20	cdisc01	CDISC.SDTM.3.1.1	AE.AEMODIFY	AEMODIFY	text	9	SPONSOR DEFINED		Modified Reported Ter			
21	cdisc01	CDISC.SDTM.3.1.1	AE.AEDECOD	AEDECOD	text	18	DERIVED	MedDRA ver	Dictionary-Derived Te			AEDICT_F
22	cdisc01	CDISC.SDTM.3.1.1	AE.AEBODSYS	AEBODSYS	text	52	DERIVED	MedDRA ver	Body System or Organ			
23	cdisc01	CDISC.SDTM.3.1.1	AE.AESEV	AESEV	text	8	CRF Page 34		Severity/Intensity			AESEVRF
24	cdisc01	CDISC.SDTM.3.1.1	AE.AESER	AESER	text	1	CRF Page 34		Serious Event			NY
25	cdisc01	CDISC.SDTM.3.1.1	AE.AEACN	AEACN	text	30	CRF Page 34		Action Taken with Stu			AEACTF

Validating the define.xml



Some process used Metadata
(data sets, variables, codelists)
to create:

- define.xml
- SAS transport files



define.xml



Validating the define.xml:

1. well-formedness
2. Against Schema
3. Against CRT-DDS Specification
4. Against SAS transport files
5. Against SDTM spec (“mandatory”)



Validating the define.xml:

1. well-formedness
2. Against Schema
3. Against CRT-DDS Specification
4. Against SAS transport files
5. Against SDTM spec ("mandatory")

Many XML based tools can do this

XML schema (1.0) can not do this.
Schema 1.1 ?? Schematron ??

XML based tools ????



Define.xml -> SAS datasets: Validation

Some process used Metadata
(data sets, variables, codelists)
to create:

- define.xml
- SAS transport files



define.xml



Use SAS XML Mapper to convert
define.xml to SAS data sets



define.xml as
SAS datasets

VALIDATION



use SAS ODS
to create define.pdf



Example define.pdf in FDA / CDISC Pilot II

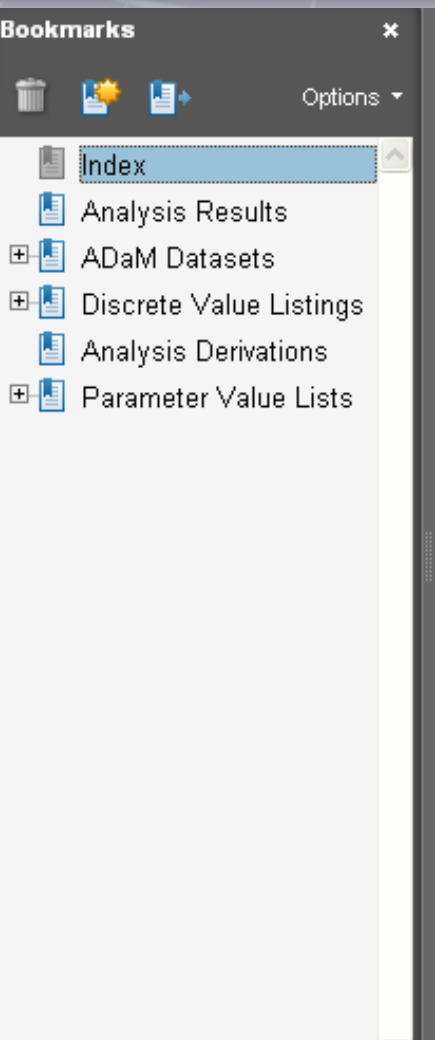


- PDF rendition is being used in the 2nd CDISC FDA Integrated Safety Data Pilot



Define.xml Tasks

- Worked with define.xml team to provide support for row level metadata
 - Link ValueLists to MethodDefs
- Adding extension for Analysis Results Metadata based on Pilot 1 findings
- New style sheets and a PDF rendering of the define.xml for human readable format



ADaM metadata for study 1

ODM Filetype	Snapshot
ODM File OID	CDISC Pilot 2 study 1
ODM Schema Location	http://www.cdisc.org/ns/odm/v1.3 ../util/cp02.xsd
ODM Version	1.3
ODM CreationDateTime	2009-09-28T13:28:34
ODM AsOfDateTime	2009-09-08T13:25:54
Study Name	CDISC_Pilot2_1
Study Description	CDISC Pilot 2 Study 1
ProtocolName	CDISCPILLOT02study1
MetaData Version OID	CDISC.SDTM.3.1.0
MetaData Version Name	CDISC_Pilot2_1, Data Definitions
MetaData Version Description	Study CDISC_Pilot2_1, Data Definitions
MetaData Version Define Version	1.0.1
MetaData Version Standard Name	CDISC SDTM
MetaData Version Standard Version	3.1.1



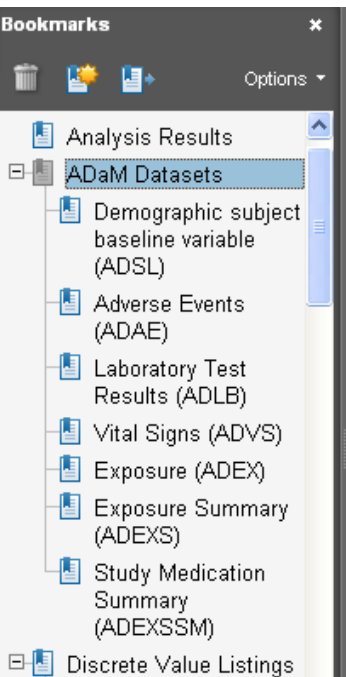
Bookmarks x

Options ▾

- Index
- Analysis Results**
- ADaM Datasets
- Discrete Value Listings
- Analysis Derivations
- Parameter Value Lists

Display Identifier	Display Name	Result Identifier	Analysis Variable	Reason	Dataset	Selection Criteria
Table 14-3.3	Exposure by Gender (Safety Population)	DOSCU	DOSCU	Pre-specified in SAP	ADSL	SAFFL eq 'Y'; class SEXN;
Table 14-3.3	Exposure by Gender (Safety Population)	DOSMAX	DOSMAX	Pre-specified in SAP	ADSL	SAFFL eq 'Y'; class SEXN;
Table 14-4.1.1	Total Treatment Emergent Events Based on Patients Experiencing the Event (Safety Population)	DECODECUR	DECODECUR	Pre-specified in SAP	ADAE	"trtemrfi='Y' and saffi='Y'"
Table 14-4.1.1	Total Treatment Emergent Events Based on Patients Experiencing the Event (Safety Population)	PARAM	PARAM	Pre-specified in SAP	ADLB	anlfi='Y' and avisitn greater than 0 and anrind in ('L','H') and saffi='Y'
Table 14-4.1.1	Total Treatment Emergent Events Based on Patients Experiencing the Event (Safety Population)	PARAM	PARAM	Pre-specified in SAP	ADVS	anlfi='Y' and avisitn greater than 0 and anrind in ('L','H') and saffi='Y'
Table 14-4.1.2	Total Treatment Emergent Events by Age Category Based on Number of Patients Experiencing the Event (Safety Population)	DECODECUR	DECODECUR	Pre-specified in SAP	ADAE	"trtemrfi='Y' and saffi='Y'; byvar=agegrp"
Table 14-4.1.2	Total Treatment Emergent Events by Age Category Based on Number of Patients Experiencing the Event (Safety Population)	PARAM	PARAM	Pre-specified in SAP	ADLB	anlfi='Y' and avisitn greater than 0 and anrind in ('L','H') and saffi='Y'; byvar=agegrp"





ADaM Datasets for Study CDISC_Pilot2_1

Dataset	Description	Structure	Keys	Location
ADSL	Demographic subject baseline variable	One record per subject	USUBJID	ADSL.xpt
ADAE	Adverse Events	One row per subject per event	STUDYID, USUBJID, AESEQ, AETERM, AEDECOD, AESTDTC	ADAE.xpt
ADLB	Laboratory Test Results	One record per subject per timepoint per labtest	STUDYID, USUBJID, LBSEQ	ADLB.xpt
ADVS	Vital Signs	One record per subject per timepoint per vital sign per position	STUDYID, USUBJID, VSSEQ	ADVS.xpt
ADEX	Exposure	One record per constant dosing interval per subject	STUDYID, USUBJID, EPOCHN, EPOCH, ASTDT, AENDT, EXSTDTC, EXENDTC	ADEX.xpt
ADEXS	Exposure Summary	One record per EPOCH per subject	STUDYID, USUBJID, EPOCHN, EPOCH, ASTDT, AENDT, EXSTDTC, EXENDTC	ADEXS.xpt
ADEXSSM	Study Medication Summary	One record per subject	STUDYID, USUBJID, ASTDT, AENDT, EXSTDTC, EXENDTC	ADEXSSM.xpt



Bookmarks ✕

Options ▾

- Index
- Analysis Results
- ADaM Datasets
 - Demographic subject baseline variable (ADSL)
 - Adverse Events (ADAE)
 - Laboratory Test Results (ADLB)
 - Vital Signs (ADVS)**
 - Exposure (ADEX)
 - Exposure Summary (ADEXS)
 - Study Medication Summary (ADEXSSM)
- Discrete Value Listings
- Analysis Derivations
- Parameter Value Lists

Vital Signs Dataset (ADVS, [ADVS.xpt](#))

Variable	Label	Type	Length	Format	Code List / Controlled Terms	Derivation	Comments
STUDYID	Study Identifier	text	6			ADVS_STUDYID	
USUBJID	Unique Subject Identifier	text	22			ADVS_USUBJID	
VSSEQ	Sequence Number	float	8			ADVS_VSSEQ	
EPOCH	Epoch	text	20			ADVS_EPOCH	
VISITNUM	Visit Number	float	8			ADVS_VSTPTNUM	
VISIT	Visit Name	text	50			ADVS_VSTPT	
VSDTC	Date/Time of Measurements	text	19			ADVS_VSDTC	
ADT	Analysis Date	float	8	IS8601DA.		ADVS_ADT	
ADTM	Analysis Date/Time	float	8	IS8601DT.		ADVS_ADTM	
ARFSTDT	Analysis Reference Start Date	float	8	IS8601DA.		ADVS_ARFSTDT	
ADY	Analysis Study Day	float	8			ADVS_ADY	
AVISITN	Analysis Visit Number	float	8			ADVS_AVISITN	
AVISIT	Analysis Visit	text	50			ADVS_AVISIT	
ANLFL	Analyzed Record Flag	text	1		YNULL	ADVS_ANLFL	
ABLFL	Baseline Record Flag	text	1		YNULL	ADVS_ABLFL	
PARAM	Parameter Description	text	200		VSPARAM	ADVS_PARAM	
PARAMCD	Parameter Code	text	8		VSPARACD	ADVS_PARAMCD	



Bookmarks ✕

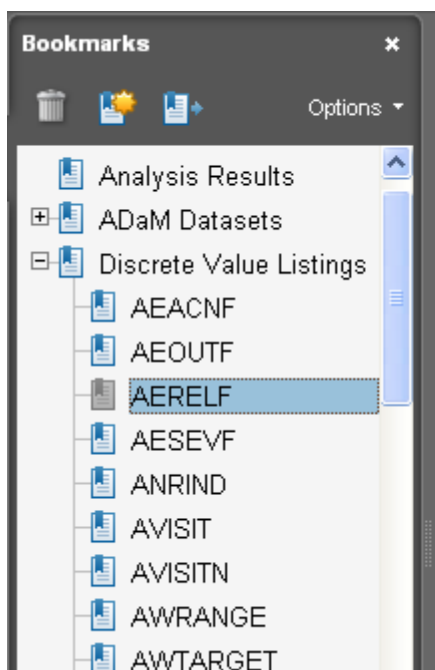
Options ▾

- Index
- Analysis Results
- ADaM Datasets
 - Demographic subject baseline variable (ADSL)
 - Adverse Events (ADAE)
 - Laboratory Test Results (ADLB)
 - Vital Signs (ADVS)
 - Exposure (ADEX)
 - Exposure Summary (ADEXS)
 - Study Medication Summary (ADESSM)
- Discrete Value Listings
- Analysis Derivations
- Parameter Value Lists
 - ADLB_PARAMCD
 - ADVS_PARAMCD

Parameter List ADVS_PARAMCD

Parameter Value	Parameter Related Variable	Type	Length	Format	Code List / Controlled Terms	Derivation	Comments
Context: In ADVS where PARAMCD is SYSBPST	PARAM	text	40		VSSTSYS		
Context: In ADVS where PARAMCD is DIABPST	PARAM	text	40		VSSTDIA		
Context: In ADVS where PARAMCD is HRST	PARAM	text	40		VSSTHR		
Context: In ADVS where PARAMCD is SYSBPSI	PARAM	text	40		VSSISYS		
Context: In ADVS where PARAMCD is DIABPSI	PARAM	text	40		VSSIDIA		
Context: In ADVS where PARAMCD is HRSI	PARAM	text	40		VSSIHR		
Context: In ADVS where PARAMCD is HEIGHT	PARAM	text	40		VSHEIGHT		
Context: In ADVS where PARAMCD is WEIGHT	PARAM	text	40		VSWEIGHT		

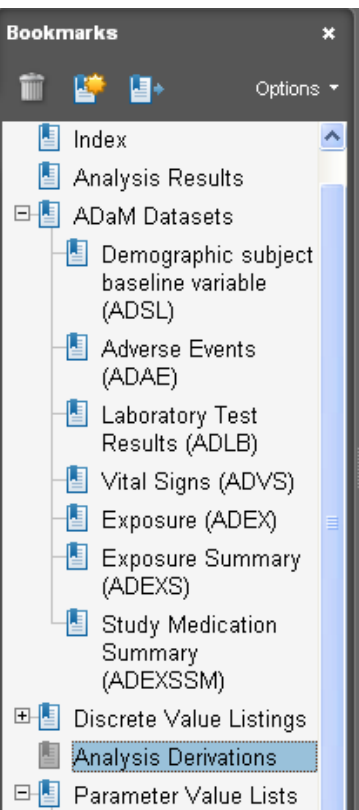




Enumerated Code List AERELF

AERELF, Reference Name (CODELISTC3)	
Valid Values	
DEFINITELY NOT	
PROBABLY NOT	
PROBABLY	
POSSIBLY	





Computational Algorithms (Methods) for Study CDISC_Pilot2_1

Method	Type	Description
ADAE_AEACN	Computation	Copied from AE.AEACN
ADAE_AEBODSYS	Computation	Derived from QNAME in SUPQUAL
ADAE_AEDECOD	Computation	Copied from AE.AEDECOD
ADAE_AEDOSE	Computation	Derived from EX.EXDOSE just before AESTDT
ADAE_AEDOSEU	Computation	Copied from EX.EXDOSE
ADAE_AEDUR	Computation	Convert AEDUR into unit of days, if Duration not recoded, calculate as AE end date-AE start date+1; if AE end date missing calculate as reference end date-AE start date+2
ADAE_AEDURN	Computation	Convert AEDUR into unit of days
ADAE_AEDURU	Computation	Hardcoded to 'Days'
ADAE_AEENDTC	Computation	Copied from AE.AEENDTC
ADAE_AEENDY	Computation	Copied from AE.AEENDY
ADAE_AEOUT	Computation	Copied from AE.AEOUT
ADAE_AEPOCH	Computation	compare the AE start date vs the treatment start date in each period to determine the EPOCH
ADAE_AEREL	Computation	Copied from AE.AEREL
ADAE_AERELN	Computation	Code AE.AEREL to numeric
ADAE_AESEQ	Computation	Copied from AE.AESEQ
ADAE_AESER	Computation	Copied from AE.AESER





Lex Jansen
Senior Consultant,
Clinical Data Strategies

Octagon Research Solutions, Inc.
585 East Swedesford Road, Suite 200
Wayne, PA 19087
Phone: (610) 535-6500 Ext. 417
Fax: (610) 535-6515

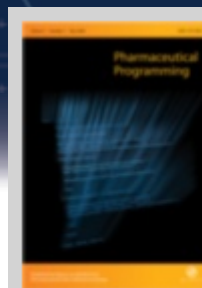
ljansen@octagonresearch.com

www.octagonresearch.com



**Find this paper and more than 10,000 other
SAS papers at <http://www.lexjansen.com>**

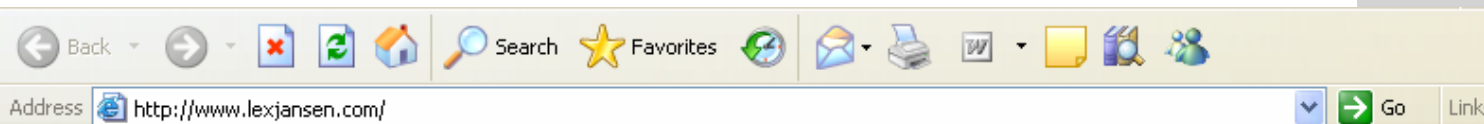




Pharmaceutical
Programming:
Journal

[Call for Papers](#)

[More
information](#)



Google™ Custom Search

Search

This searches **10019** SAS papers from SAS Global Forum, SUGI, PharmaSUG, NESUG, SESUG, PhUSE, WUSS, MWSUG, PNWSUG and SCSUG. Alternatively, search for a word based on [title, author or keywords](#).

4195 SAS papers presented at [SGF/SUGI 1996-2009](#).

1191 SAS papers presented at [PharmaSUG 2000-2009](#).

1810 SAS papers presented at [NESUG 1997-2009](#).

1087 SAS papers presented at [SESUG 1999-2008](#).

357 SAS papers presented at [Phuse 2005-2008](#).
-- PhUSE 2009 preview.

787 SAS papers presented at [WUSS 2003-2009](#).

207 SAS papers presented at [MWSUG 2001, 2005-2008](#).

188 SAS papers presented at [PNWSUG 2004-2009](#).

197 SAS papers presented at [SCSUG 2003-2007](#).

163 SAS papers related to [CDISC](#).
Easy access to the [CDISC Forum](#).

