

Accessing the metadata from the define.xml using XSLT transformations

Lex Jansen

Octagon Research Solutions, Inc.

*Leading the Electronic Transformation
of Clinical R&D*



**PhUSE 2010
Berlin**

- Introduction
- define.xml: Regulatory landscape
- Data Definition Tables (pdf / xml)
- What is the define.xml
- Displaying the define.xml
- XSLT



Introduction



- This presentation is **NOT** about **CREATING** a define.xml file
- It is about how the information (metadata) in a define.xml file can be **USED**
- Before we can **USE** the metadata from the define.xml file, we need to be able to **ACCESS** that metadata
- This presentation has a focus on **XML** technologies (**XSLT**) to access that metadata



Define.xml: 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- Draft Guidance for Industry on Providing Regulatory Submissions in Electronic Format--Human Pharmaceutical Applications and Related Submissions.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 - [The Draft Guidance](#) 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
 - 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”*
- *... Include a reference to the style sheet as defined in the specification and place the corresponding style sheet in the same folder as the define.xml file ...*



November 2009 - Study Data Specifications v1.5:

- *"For datasets not prepared using the CDISC SDTM specifications, consult Appendix 2 for information concerning the preparation of a **define.pdf** data definition file."*
- Appendix 2 specifies a define.pdf specification similar to the 1999 guidance



Data Definition Tables in PDF



- Study Data Specifications:

"For datasets not prepared using the CDISC SDTM specifications, consult Appendix 2 for information concerning the preparation of a define.pdf data definition file"

Datasets for Study 1234		
Dataset	Description of Dataset	Location
AE	Adverse Events	m5/datasets/study1234/listings/ae.xpt
...	...	...



- "Sponsors should also provide a link to the appropriate annotated case report form file (blankcrf.pdf)"*

Study 1234 – Demographics Dataset Variables				
Variable	Label	Type	Codes	Comments ¹
USUBJID	Unique subject ID number	char		Demographics page 3
SEX	Sex of subject	char	f = female m = male	Demographics page 3
BDATE	Birth date	date		Demographics page 3
DUR	Duration of Treatment	num		Derived STOP DATE – START DATE
TRT	Assigned treatment group	num	0= placebo 5= 5mg/day	

¹Use footnotes for longer comments



Data Definition Tables in XML



- As of January 1, 2008: follow the eCTD guidance and document submitted data by including data definition tables (**define.xml**) and annotated case report forms (blankcrf.pdf)

```
<?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="Studydisc01"
  ODMVersion="1.2" FileType="Snapshot"
  CreationDateTime="2007-04-09T12:24:09">
  - <Study OID="cdisc01">
    - <GlobalVariables>
      <StudyName>CDISC01</StudyName>
      <StudyDescription>CDISC01 Test Study.</StudyDescription>
      <ProtocolName>CDISC01</ProtocolName>
    </GlobalVariables>
    <MetaDataVersion OID="CDISC.SDTM.3.1.1"
      Name="Study CDISC01, Data Definitions"
      Description="Study CDISC01, Data Definitions"
      def:DefineVersion="1.0.0"
      def:StandardName="CDISC SDTM"
      def:StandardVersion="3.1.1">
      ...
      <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.MH">
          <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>
```



```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<!-- ***** -->
<!-- File: define.xml -->
<!-- Date: 28-01-2005 -->
<!-- Author: Clinical Data Interchange Standards Consortium (CDISC) -->
<!-- Description: This is an example define.xml document which implements the Case -->
<!-- Report Tabulation Data Definition Specification Version 1.0.0. -->
<!-- ***** -->
<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:leaf ID="blankcrf" xlink:href="blankcrf.pdf">
```



... with a stylesheet

Study 1234, Data Definitions						
Annotated Case Report Form Supplemental Data Definitions Document Datasets Demographics (DM) Trial Elements (TE) Trial Arms (TA) Trial Visits (TV) Subject Elements (SE) Subject Visits (SV) Exposure (EX) Concomitant Medications (CM) Substance Use (SU) Adverse Events (AE) Disposition (DS) Medical History (MH) ECG Test Results (EG) Inclusion/Exclusion Exceptions (IE) Laboratory Tests (LB) Physical Examination (PE) Subject Characteristics (SC) Vital Signs (VS) Comments (CO) Value Level Metadata Computational Algorithms Controlled Terminology						
Datasets for Study 1234						
Dataset	Description	Class	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 violation per	Tabulation	USUBJID, DSTERM, DSSEQ	crt/datasets/1234/ds.xpt

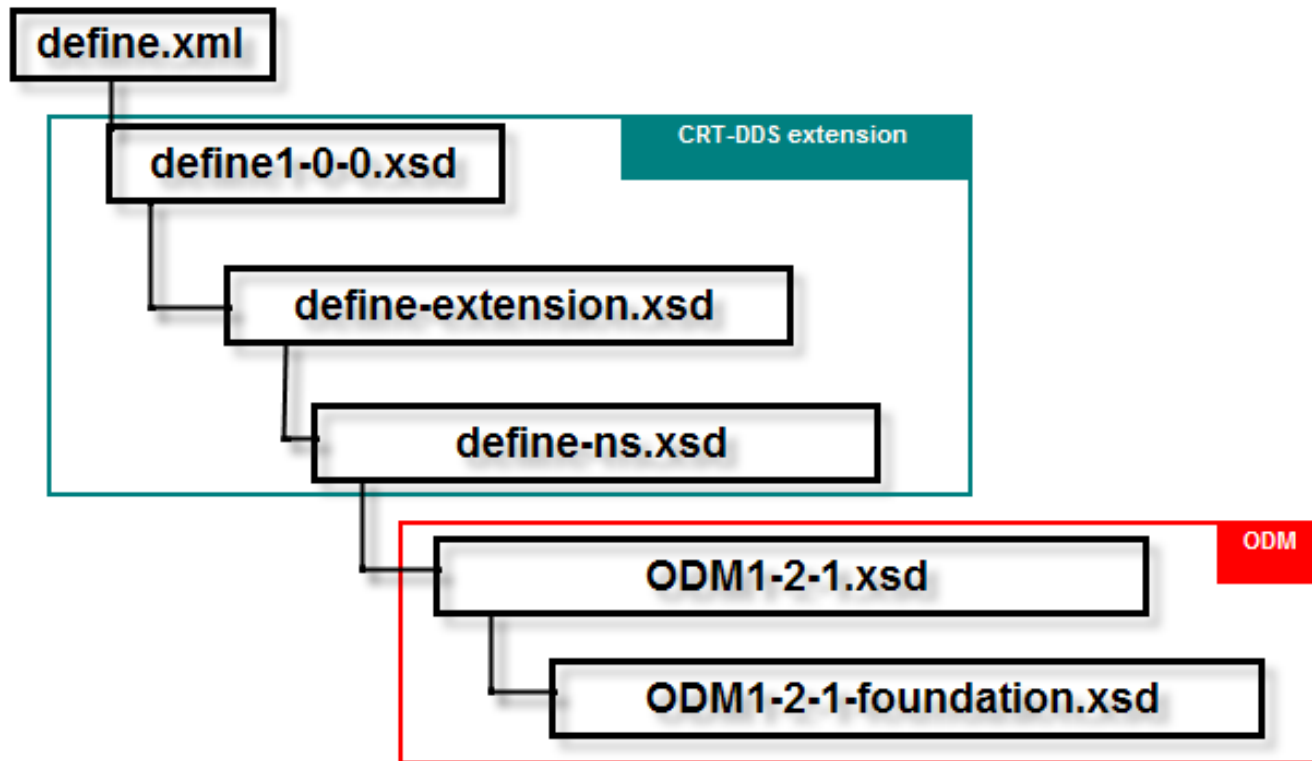
define.xml



- Case Report Tabulation Data Specification (CRT-DDS, or define.xml): Production version: **1.0.0**
- Extension of the CDISC Operational Data Model (**ODM**), an XML specification to facilitate the archival and interchange of the metadata and data for clinical research
- Maintained by CDISC's **XML Technologies Team** (formerly known as the ODM team)
- New define.xml version 2 in development with additional metadata support for SDTM and ADaM (results metadata)



XML schema definitions (XSD) describe the structure of the define.xml



 <http://www.cdisc.org/define-xml>



ABOUT CDISC

STANDARDS

RESOURCES

NEWS

EDUCATION & EVENTS

STANDARDS

SDTM

Operational Data
Model

Define.XML

Study/Trial Design
Model

LAB

ADaM

Protocol

Terminology

CDASH

SEND

CDISC SHARE

Therapeutic Area
Standards

Define.XML

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 here](#).

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

 <http://www.cdisc.org/define-xml>

Case Report Tabulation Data Definition Specification (define.xml)

Prepared by the
CDISC define.xml Team

Principal Editor: William Qubeck
Principal Contributors: Sally Cassells, Anthony Friebe, 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
2005-02-05	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 Friebe, William Qubeck, Sally Cassells

define.xml – Specifications

 <http://www.cdisc.org/odm>



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 GlobalVariables](#)
 - [3.4 StudyName](#)
 - [3.5 StudyDescription](#)
 - [3.6 ProtocolName](#)
 - [3.7 BasicDefinitions](#)
 - [3.8 MeasurementUnit](#)
 - [3.9 Symbol](#)
 - [3.10 TranslatedText](#)
- [4 Metadata Elements](#)



[ABOUT CDISC](#)

[STANDARDS](#)

[RESOURCES](#)

[NEWS](#)

[EDUCATION & EVENTS](#)

STANDARDS

[Study Data Tabulation Model](#)

[Operational Data Model](#)

[Define.XML](#)

[Study/Trial Design Model](#)

[LAB](#)

[ADaM](#)

[Protocol](#)

[Terminology](#)

[CDASH](#)

[SEND](#)

[CDISC SHARE](#)

[Therapeutic Area Standards](#)

Define.XML

XML Schema Validation for Define.xml White Paper

The XML Schema Validation for Define.xml white paper provides guidance on validating define.xml version 1.0 documents against the define.xml XML schemas. It proposes practices and tools to improve define.xml schema validation. The goal is to foster more consistent validation results in order to facilitate regulatory submissions and interchange of define.xml documents. The document does not include any define.xml specifications or recommendations for creating define.xml content.

[XML Schema Validation for Define.xml \(pdf\)](#)

This zip file packages 3 define.xml version 1.0 examples with copies of the schemas for testing define.xml validation.

[Define.xml Validation \(zip\)](#)



Displaying the define.xml



- **define.xml**
contains
metadata and is
machine
readable
- **define.xml**
becomes
human
readable with a
stylesheet

```
<?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 becomes human readable with an **XSL** stylesheet

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

... and looks even fancier with a different **stylesheet**

Study 1234, Data Definitions						
Datasets for Study 1234						
Dataset	Description	Class	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 violation per	Tabulation	USUBJID, DSTERM, DSSEQ	crt/datasets/1234/ds.xpt

iPad100%

Google

Study CDISC01, Data Definitions

PageTools

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, ETCDD	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

QWERTYUIOP

ASDFGHJKLGo

⬆ZXCVBNM, .⬆

.?123: / _ - .com .?123📱



```
<?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="Studycdisc01"
  ODMVersion="1.2"
  FileType="Snapshot"
  CreationDateTime="2007-04-09T12:24:09">
<Study OID="cdisc01">
  <GlobalVariables>
    <StudyName>CDISC01</StudyName>
    <StudyDescription>CDISC01 Test Study.</StudyDescription>
    <ProtocolName>CDISC01</ProtocolName>
  </GlobalVariables>
  <MetaDataVersion OID="CDISC.SDTM.3.1.1"
    Name="Study CDISC01, Data Definitions"
    Description="Study CDISC01, Data Definitions"
    def:DefineVersion="1.0.0"
    def:StandardName="CDISC SDTM"
    def:StandardVersion="3.1.1">
```

Displaying the define.xml



define.xml



define1-0-0.xsd



define-extension.xsd



define-ns.xsd



define1-0-0.xsl



define.css



icon1.gif



icon2.gif



icon3.gif



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

```
<table border="2" cellpadding="4">
  <tr>
    <th colspan="7" align="left" valign="top" height="20">Datasets for Study CDISC01</th>
  </tr><font face="Times New Roman" size="2"></font><tr align="center">
    <th align="center" valign="bottom">Dataset</th>
    <th align="center" valign="bottom">Description</th>
    <th align="center" valign="bottom">Class</th>
    <th align="center" valign="bottom">Structure</th>
    <th align="center" valign="bottom">Purpose</th>
    <th align="center" valign="bottom">Keys</th>
    <th align="center" valign="bottom">Location</th>
  </tr>
  <tr align="left" valign="top">
    <td>SV</td>
    <td><a href="#SV">Subject Visits</a></td>
    <td>TRIAL DESIGN</td>
    <td>One record per subject per actual visit</td>
    <td>Tabulation&nbsp;</td>
    <td>STUDYID, USUBJID, SVSTDTC, SVENDTC, VISITNUM, SVUPDES&nbsp;</td>
    <td><a href="sv.xpt">sv.xpt</a></td>
  </tr>
```

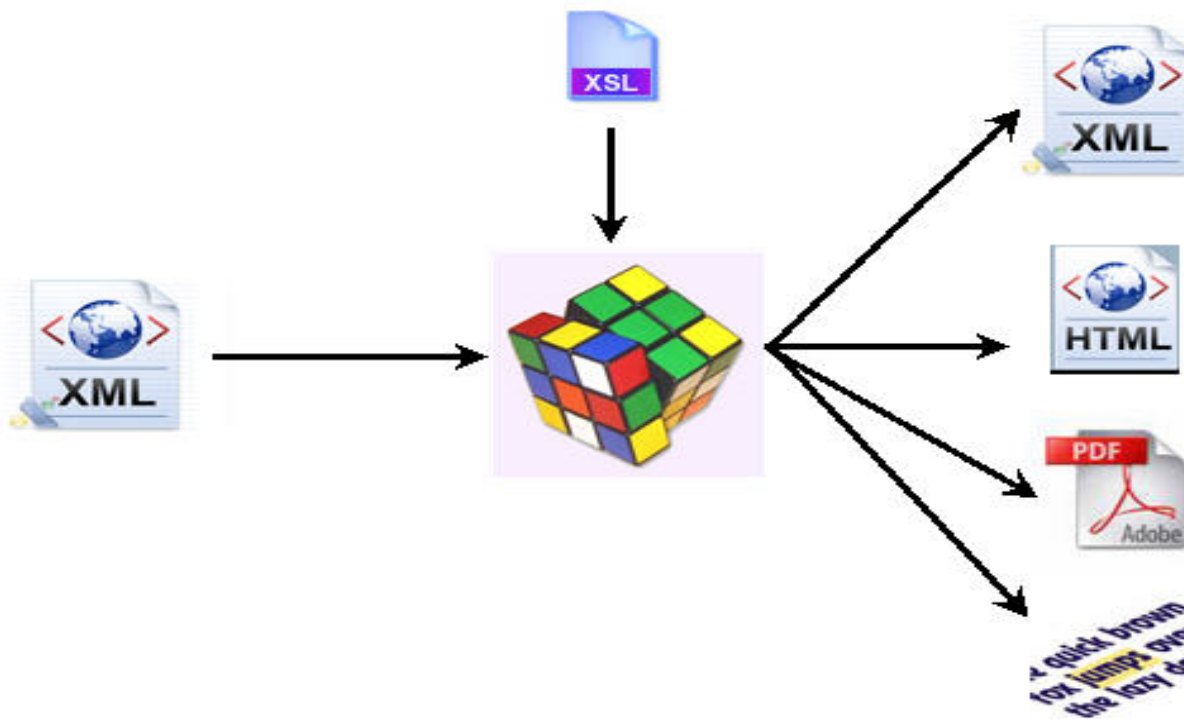
XSLT



- **eXtensible S**tylesheet **L**anguage **T**ransformations (XSLT) is a language that lets you convert XML documents into other XML documents, into HTML documents, or into any other text based document (like a SAS program), or even a PDF file
- XSLT is a language *"for transforming the structure and content of an XML document"*



- XSL transformations are like Rubik's cube!



•XML

•HTML

•PDF

•TEXT



- The mandatory "hello world" XML

Powerpoint would like to use
your current location

Don't Allow

OK

<?xml v

<doc>

<title>Hello World Test</title>

<greeting>Hello World</greeting>

</doc>

859-1"?>



- The mandatory "hello Berlin" XML

```
<?xml version="1.0" encoding="iso-8859-1"?>  
<doc>  
  <title>Hello Berlin Test</title>  
  <greeting>Hello Berlin</greeting>  
</doc>
```



- XSL stylesheet

```
<?xml version="1.0" encoding="iso-8859-1"?>
<xsl:stylesheet version="1.0"
  xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
  xmlns="http://www.w3.org/1999/xhtml">

  <xsl:output method="html" doctype-system="http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd"
    doctype-public="-//W3C//DTD XHTML 1.0 Transitional//EN" indent="yes"/>

  <xsl:template match="/">
    <html>
      <head>
        <title><xsl:value-of select="doc/title"/></title>
      </head>
      <body>
        <p><xsl:value-of select="doc/greeting"/></p>
      </body>
    </html>
  </xsl:template>
</xsl:stylesheet>
```



- Example: XML + XSL = HTML

```
<!DOCTYPE html
  PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
    "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <title>Hello Berlin Test</title>
  </head>
  <body>
    <p>Hello Berlin</p>
  </body>
</html>
```



Other examples of using the define.xml metadata



- Use dataset and variable information (type, length, label) to create zero-observation datasets that can serve as data conversion targets

```
<ItemGroupDef OID="DM"  
  Name="DM"  
  Repeating="No"  
  IsReferenceData="No"  
  SASDatasetName="DM"  
  Purpose="Tabulation"  
  def:Label="Demographics"
```

```
  <ItemRef ItemOID="DM.BRTHDTC"  
    OrderNumber="8"  
    Mandatory="No"  
    Role="RESULT QUALIFIER"  
    RoleCodeListOID="RoleCodeList"/>  
  <ItemRef ItemOID="DM.AGE"  
    OrderNumber="9"  
    Mandatory="No"  
    Role="RESULT QUALIFIER"  
    RoleCodeListOID="RoleCodeList"/>
```

```
<ItemDef OID="DM.BRTHDTC"  
  Name="BRTHDTC"  
  DataType="text"  
  Length="10"  
  Origin="CRF Page 3"  
  def:Label="Date/Time of Birth"  
>  
</ItemDef>  
<ItemDef OID="DM.AGE"  
  Name="AGE"  
  DataType="integer"  
  Length="2"  
  Origin="DERIVED"  
  Comment="Screening Date - Birth date"  
  def:Label="Age in AGEU at RFSTDTC"
```



DATASET TEMPLATES from the define.xml

```
CREATE TABLE DM(LABEL="Demographics") (  
  STUDYID CHAR(7) LABEL="Study Identifier",  
  DOMAIN CHAR(2) LABEL="Domain Abbreviation",  
  USUBJID CHAR(14) LABEL="Unique Subject Identifier",  
  SUBJID CHAR(6) LABEL="Subject Identifier for the Study",  
  RFSTDTC CHAR(10) LABEL="Subject Reference Start Date/Time",  
  RFENDTC CHAR(10) LABEL="Subject Reference End Date/Time",  
  SITEID CHAR(3) LABEL="Study Site Identifier",  
  BRTHDTC CHAR(10) LABEL="Date/Time of Birth",  
  AGE NUMERIC LABEL="Age in AGEU at RFSTDTC",  
  AGEU CHAR(5) LABEL="Age Units",  
  ...  
);
```



- Use codelist information (codes/decodes) to create a PROC FORMAT

```
<CodeList OID="FRAME" Name="FRAME" DataType="text" SASFormatName="$FRAME">  
  <CodeListItem CodedValue="S" def:Rank="1">  
    <Decode>  
      <TranslatedText xml:lang="en">Small</TranslatedText>  
    </Decode>  
  </CodeListItem>  
  <CodeListItem CodedValue="M" def:Rank="2">  
    <Decode>  
      <TranslatedText xml:lang="en">Medium</TranslatedText>  
    </Decode>  
  </CodeListItem>  
  <CodeListItem CodedValue="L" def:Rank="3">  
    <Decode>  
      <TranslatedText xml:lang="en">Large</TranslatedText>  
    </Decode>  
  </CodeListItem>  
  <CodeListItem CodedValue="XL" def:Rank="4">  
    <Decode>  
      <TranslatedText xml:lang="en">Extra large</TranslatedText>  
    </Decode>  
  </CodeListItem>  
</CodeList>
```



PROC FORMAT from the define.xml

```
PROC FORMAT FMTLIB;  
  VALUE $FRAME  
    "S" = "SMALL"  
    "M" = "MEDIUM"  
    "L" = "LARGE"  
    "XL" = "EXTRA LARGE"  
  ;  
  VALUE $SEX  
    "F" = "FEMALE"  
    "M" = "MALE"  
    "U" = "UNKNOWN"  
  ;  
  VALUE SMKCLAS  
    1 = "NEVER SMOKED"  
    2 = "SMOKER"  
    3 = "EX SMOKER"  
  ;  
RUN;
```



```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet version="1.0"
    xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
    xmlns:odm="http://www.cdisc.org/ns/odm/v1.2"
    xmlns:def="http://www.cdisc.org/ns/def/v1.0"
    xmlns:xlink="http://www.w3c.org/1999/xlink">

  <xsl:output method="text" indent="no" />

  <xsl:variable name="lf" select="'&#13;&#10;'" />

  <xsl:template match="/">

    <xsl:text>PROC FORMAT LIBRARY=WORK;</xsl:text>
    <xsl:value-of select="$lf" />

    <xsl:call-template name="CodeLists" />

    <xsl:text>RUN;</xsl:text>
  </xsl:template>
```



```
<CodeList OID="FRAME" Name="FRAME" DataType="text" SASFormatName="$FRAME">
  <CodeListItem CodedValue="S" def:Rank="1">
    <Decode>
      <TranslatedText xml:lang="en">Small</TranslatedText>
    </Decode>
  </CodeListItem>
```

```
<xsl:template name="CodeLists">
```

```
  <xsl:for-each select="/odm:ODM/odm:Study/odm:MetaDataVersion/odm:CodeList">
    <xsl:variable name="DataType" select="@DataType"/>
```

```
    <xsl:text> VALUE </xsl:text>
```

```
    <xsl:choose>
```

```
      <xsl:when test="@SASFormatName">
```

```
        <xsl:value-of select="@SASFormatName"/>
```

```
      </xsl:when>
```

```
      <xsl:otherwise>
```

```
        <xsl:if test="$DataType='text'">$</xsl:if>
```

```
        <xsl:value-of select="substring(@Name,1,7)"/>
```

```
      </xsl:otherwise>
```

```
    </xsl:choose>
```

```
  <xsl:value-of select="$If"/>
```



```
<CodeList OID="FRAME" Name="FRAME" DataType="text" SASFormatName="$FRAME">
  <CodeListItem CodedValue="S" def:Rank="1">
    <Decode>
      <TranslatedText xml:lang="en">Small</TranslatedText>
    </Decode>
  </CodeListItem>
```

```
<xsl:for-each select="./odm:CodeListItem">
  <xsl:sort select="concat(@def:Rank, ' ', @CodedValue)"/>
  <xsl:text> </xsl:text>
  <xsl:if test="$DataType='text'"></xsl:if>
  <xsl:value-of select="@CodedValue"/>
  <xsl:if test="$DataType='text'"></xsl:if>
  <xsl:text> = </xsl:text>
  <xsl:value-of select="odm:Decode/odm:TranslatedText/text()"/>
  <xsl:text> </xsl:text>
  <xsl:value-of select="$If"/>
</xsl:for-each>
```

```
<xsl:text> ;</xsl:text>
<xsl:value-of select="$If"/>
```

```
</xsl:for-each>
</xsl:template>
</xsl:stylesheet>
```



Running XSLT with SAS

Experimental

```
FILENAME XMLfile "define.xml";  
FILENAME XSLfile "ProcFormat.xml";  
FILENAME Outfile "ProcFormat.sas";
```

```
❏ PROC XSL IN=XMLfile XSL=XSLfile OUT=Outfile;  
RUN;
```

```
%INCLUDE "ProcFormat.sas";
```



```
1  FILENAME XMLfile "define.xml";
2  FILENAME XSLfile "ProcFormat.xsl";
3  FILENAME Outfile "ProcFormat.sas";
4
5  PROC XSL IN=XMLfile XSL=XSLfile OUT=Outfile;
6  RUN;
```

WARNING: This is an experimental release of the XSL procedure.

NOTE: PROCEDURE XSL used (Total process time):

real time	1.09 seconds
cpu time	0.07 seconds

```
7
8  %INCLUDE "ProcFormat.sas";
NOTE: Format $FRAME has been output.
NOTE: Format $SEX has been output.
NOTE: Format SMKCLAS has been output.
NOTE: Format $YESNO has been output.
NOTE: Format $YESNOUN has been output.
```

NOTE: PROCEDURE FORMAT used (Total process time):

real time	0.07 seconds
cpu time	0.00 seconds

<http://xml.apache.org/xalan-j/>

THE
APACHE XML
PROJECT

Xalan-Java Version 2.7.1

XML.APACHE.ORG

WWW.APACHE.ORG

WWW.W3.ORG w3c

- Home
- **Xalan-J 2.7.1**
- Charter
- What's New
- Release Notes
- Overview
- Download/Build
- Getting Started
- Using XSLTC
- FAQs
- Sample Apps
- Command Line
- Features
- Transform API
- XPath API
- Usage Patterns
- Xalan-J API

- [What is it?](#)
- [How about this release?](#)
- [For more information...](#)
- [Apache License Version 2.0](#)
- [Notice file](#)
- [Apache License Version 1.1](#)

What is it?

Xalan-Java is an XSLT processor for transforming XML documents into HTML, text, or other XML document types. It implements [XSL Transformations \(XSLT\) Version 1.0](#) and [XML Path Language \(XPath\) Version 1.0](#) and can be used from the command line, in an applet or a servlet, or as a module in other program.

```
x "java -jar C:\xalan-j_2_7_1\xalan.jar  
    -IN define.xml  
    -XSL ProcFormat.xsl  
    -OUT ProcFormat.sas  
    -PARAM library MyLib";
```



- In case you get serious about XSLT ...
- Get a good XML editor
 - Oxygen (<http://www.oxygenxml.com/>)
 - XMLSpy (<http://www.altova.com>) has some issues in validating define.xml
 - Check out the define.xml white paper on **<http://www.cdisc.org/define-xml>**





Lex Jansen
Senior Consultant,
Clinical Data Strategies

Octagon Research Solutions, Inc.
585 East Swedesford Road, Suite 200
Wayne, PA 19087

ljansen@octagonresearch.com

www.octagonresearch.com

**Find this paper and more than 11,000 other
SAS papers at <http://www.lexjansen.com>**

