

ORACLE®



CDISC Dataset-XML

A new Dataset Structure for Clinical Trial Data Transport
for Future Drug Submissions

Joerg Dillert

Consulting Solution Director
Oracle Health Sciences

PhUSE Annual Conference
October, 13th - 15th 2014

ORACLE

ORACLE
HEALTH SCIENCES

Safe Harbor Statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.



CDISC Europe Interchange 2015

4 – 8 May 2015

Location:

Basel, Switzerland.

CDISC Interchange, Europe, Call for ABSTRACTS

The Programme Committee is now accepting abstracts for presentations and posters at the CDISC EU Interchange in Basel, Switzerland on the 5th and 6th May 2015.

The hot topics we see for this year are:

Measureable Benefits of CDISC Implementations

Implementing CDISC standards (e.g. SDTM, terminology, CDASH) into your R&D process and measure and metrics of the benefits.

Practical Implementation of Therapeutic Area Standards

Therapeutic area specific topics and their impact on existing CDISC standards; practical implementation experiences

Protocol Metadata for the CDISC end-to-end implementation

Structured protocol metadata as a starting point to automate setup of clinical databases and downstream CDISC standards deliverables such as CRF's, datasets and reports. What is needed?

Main conference 5th/6th May 2015

Credits & Acknowledgement

CDISC Dataset-XML team

- Sam Hume, CDISC
- Sally Cassells, Next Step Clinical Systems LLC
- Kevin Burges, Formedix
- Jozef Aerts, XML4Pharma
- Lex Jansen, SAS Institute
- Marcelina Hungria, Dcore Group, LLC
- Mike Molter, D-Wise
- Peter Schaefer, Certara
- Paul Graham, Formedix
- Vanessa Nguyen, Baxter
- Veena Nataraj, Shire
- Vojtech Huser, NIH
- Priscilla Gathoni, Novartis
- Yoshiteru CHIBA, UMINCenter, Japan

CDISC Dataset-XML

- 1 ➤ Quick history of SAS[®]-XPT in Electronic Submissions
- 2 ➤ CDISC Dataset-XML
- 3 ➤ Data Structure
- 4 ➤ Demo
- 5 ➤ Status and Future

History of SAS[®]-XPT

... and the combined challenges

SAS[®]-XPT in Clinical Research and Electronic Submissions

- FDA, eCTD and Electronic Submissions
- Free available and documented formats: PDF and SAS[®]-XPT (V5)

Hmmmm?

- 10 years later ...
 - IT technology improvement
 - PDF version issues
 - FDA PDF v1.4-1.7
 - Security issues
 - SAS®-XPT for non SAS systems

FDA PORTABLE DOCUMENT FORMAT (PDF) SPECIFICATIONS

PURPOSE

These specifications are for submitting documents in Portable Document Format (PDF). The purpose of this document is to provide specifications for submitting PDF files that align with the ICH M2 recommendations¹ and that are in a format that the receiving Center currently supports. For purposes of this document, “supports” means the receiving Center has established processes and technology infrastructure to support the receipt, processing, review, and archive of files in the specified standard format. PDF is an open, published format created by Adobe Systems Incorporated (<http://www.adobe.com>). Software from a variety of sources can be used to create files in the PDF format.

VERSION

PDF versions 1.4 through 1.7 are acceptable. Submitted PDF files should be readable by Adobe Acrobat 8.0, should not require additional software or plug-ins to be read and navigated, and should be text searchable. If plug-ins are used during the creation of a PDF document, prior to submitting the document, ensure that a plug-in is not needed for review or archive.

SECURITY

Do not activate security settings or password protection. The integrity of the submitted files is maintained through Agency security and archival processes. A copy of the files, generated from the submitted files, will be provided to the reviewer. The reviewer should be able to print, select text and graphics, and make changes to text, notes and form fields using the provided copy.

FDA Forms downloaded from the FDA Forms website contain security settings that prevent changing the documents. These forms should be submitted as provided, with no additional security added and without removing the provided security settings.

The screenshot shows the SAS System Viewer interface. The main window displays the SAS log for a file named dm.xpt. The log contains several headers and data records. A Notepad window is open in the foreground, showing a large block of text that appears to be a corrupted or encoded version of the data, possibly containing Chinese characters and symbols.

- Sure, documented,
— but kind of binary

ORACLE®

RECORD LAYOUT OF A SAS® VERSION 5 OR 6 DATA SET IN SAS® TRANSPORT (XPORT) FORMAT

In this record, the string is the datetime modified. Most often, the datetime created and datetime modified will always be the same. Pad with ASCII blanks to

What's wrong with XPT?

- Variable name length limitation
 - Variable label length limitation
 - Variable data length limitation
 - No support for international characters sets
 - A kind of binary format (structs)
- > Hard to process on non-SAS® platforms

```
struct NAMESTR {  
    short ntype;      /* VARIABLE TYPE: 1=NUMERIC, 2=CHAR */  
    short nhfun;      /* HASH OF NNAME (always 0) */  
    short nlng;       /* LENGTH OF VARIABLE IN OBSERVATION */  
    short nvar0;      /* VARNUM */  
    char8 nname;      /* NAME OF VARIABLE */  
    char40 nlabel;    /* LABEL OF VARIABLE */  
    char8 nform;      /* NAME OF FORMAT */  
    short nfl;        /* FORMAT FIELD LENGTH OR 0 */  
    short nfd;        /* FORMAT NUMBER OF DECIMALS */  
    short nfj;        /* 0=LEFT JUSTIFICATION, 1=RIGHT JUST */  
    char nfill[2];    /* (UNUSED, FOR ALIGNMENT AND FUTURE) */  
    char8 niorm;       /* NAME OF INPUT FORMAT */  
    short nifl;       /* INFORMAT LENGTH ATTRIBUTE */  
    short nifd;       /* INFORMAT NUMBER OF DECIMALS */  
    long npos;        /* POSITION OF VALUE IN OBSERVATION */  
    char rest[52];    /* remaining fields are irrelevant */  
};
```

Page 3 of 21

<https://support.sas.com/techsup/technote/ts140.pdf>

Requirements for a change

- A new, modern dataset transport format of regulated data
- Easy to process (with free tools) and faster software development
- Support tabular datasets including SDTM, ADaM, SEND
- Support of international character sets

CDISC Dataset-XML

... as the answer

Where does the name come from?

- Original name was

Study Data Set XML (SDS-XML)

- To avoid confusion changed to

Dataset-XML

Dataset-XML

- Based on CDISC ODM and Define-XML
 - Define.XML V 2.0 or later is recommended for Dataset-XML
- Related to SDTM, ADaM, SEND and any kind of tabular data in the future

What will change for submission / eCTD?

For the end user?

- Old

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: 2013-03-03T17:04:44

Stylesheet version: 2013-04-24

Tabulation Datasets for Study CDISC01 (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, ARMCD, TAETORD	ta.xpt	
TE	Trial Elements	TRIAL DESIGN	One record per planned Element	Tabulation	STUDYID, ETCD	te.xpt	
TI	Trial Inclusion/Exclusion Criteria	TRIAL DESIGN	One record per I/E criterion	Tabulation	STUDYID, IETESTCD	ti.xpt	
TS	Trial Summary	TRIAL DESIGN	One record per trial summary parameter value	Tabulation	STUDYID, TSPARMCD, TSSEQ	ts.xpt	
TV	Trial Visits	TRIAL DESIGN	One record per planned Visit per Arm	Tabulation	STUDYID, VISITNUM, ARMCD	tv.xpt	
DM	Demographics	SPECIAL PURPOSE	One record per subject	Tabulation	STUDYID, USUBJID	dm.xpt	See Reviewer's Guide, Section 2.1 Demographics Reviewers Guide

What will change for submission / eCTD?

Nothing special

- New

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: 2013-11-09T22:11:15

Stylesheet version: 2013-03-04

Tabulation Datasets for Study CDISC01 (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, ARMCD, TAETORD	ta.xml	
TE	Trial Elements	TRIAL DESIGN	One record per planned Element	Tabulation	STUDYID, ETCD	te.xml	
TI	Trial Inclusion/Exclusion Criteria	TRIAL DESIGN	One record per I/E criterion	Tabulation	STUDYID, IETESTCD	ti.xml	
TS	Trial Summary	TRIAL DESIGN	One record per trial summary parameter value	Tabulation	STUDYID, TSPARMCD, TSSEQ	ts.xml	
TV	Trial Visits	TRIAL DESIGN	One record per planned Visit per Arm	Tabulation	STUDYID, VISITNUM, ARMCD	tv.xml	
DM	Demographics	SPECIAL PURPOSE	One record per subject	Tabulation	STUDYID, USUBJID	dm.xml	See Reviewer's Guide, Section 2.1 Demographics Reviewers Guide
SE	Subject Elements	SPECIAL	One record per actual	Tabulation	STUDYID, USUBJID,	se.xml	

Data Structure

Bring some light on it – no magic

ODM is my base for the structure (1) - Clinical Data

<ODM>

- base ODM information as usual

<ClinicalData>

- study information, metadata version

<ItemGroupData>

- table/dataset level

<ItemData>

- data points, items

For all clinical domains, like AE, DM, CM etc.

First Clinical Data Example

```
<?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:data="http://www.cdisc.org/ns/Dataset-XML/v1.0"
  FileType="Snapshot"
  ODMVersion="1.3.2"
  data:DatasetXMLVersion="1.0"
  FileOID="www.cdisc.org.Studydisc01-Define-XML_2.0.0 (IG.CM) "
  PriorFileOID="www.cdisc.org.Studydisc01-Define-XML_2.0.0"
  Originator="CDISC Dataset-XML Team"
  CreationDateTime="2014-03-20T21:45:33">
  <ClinicalData
    StudyOID="cdisc01"
    MetadataVersionOID="MDV.CDISC01.SDTMIG.3.1.2.SDTM.1.2">
    <!-- Dataset (CM) -->
    <ItemGroupData ItemGroupOID="IG.CM" data:ItemGroupDataSeq="1">
      <ItemData ItemOID="IT.STUDYID" Value="CDISC01"/>
      <ItemData ItemOID="IT.CM.DOMAIN" Value="CM"/>
      <ItemData ItemOID="IT.USUBJID" Value="CDISC01.100008"/>
      <ItemData ItemOID="IT.CM.CMSEQ" Value="1"/>
      <ItemData ItemOID="IT.CM.CMTRT" Value="PROCARDIA XL"/>
      ...
    </ItemGroupData>
    ...
  </ClinicalData>
</ODM>
```

ODM is my base for the structure (2) – Reference Data

<ODM>

< **ReferenceData** >

<ItemGroupData>

<ItemData>

- base ODM information as usual
- study information, metadata version
 - table/dataset level
- data points, items

For all reference / trial metadata domains, like TA, TV etc

Reference Data example

```
<?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:data="http://www.cdisc.org/ns/Dataset-XML/v1.0"
  FileType="Snapshot"
  ODMVersion="1.3.2"
  data:DatasetXMLVersion="1.0"
  FileOID="www.cdisc.org.Studydisc01-Define-XML_2.0.0(IG.TA)"
  PriorFileOID="www.cdisc.org.Studydisc01-Define-XML_2.0.0"
  Originator="CDISC Dataset-XML Team"
  CreationDateTime="2014-03-20T21:45:33" >
  <ReferenceData
    StudyOID="cdisc01"
    MetadataVersionOID="MDV.CDISC01.SDTMIG.3.1.2.SDTM.1.2">
    <!-- Dataset (TA) -->
    <ItemGroupData ItemGroupOID="IG.TA" data:ItemGroupDataSeq="1">
      <ItemData ItemOID="IT.STUDYID" Value="CDISC01"/>
      <ItemData ItemOID="IT.TA.DOMAIN" Value="TA"/>
      <ItemData ItemOID="IT.TA.ARMCD" Value="PLACEBO"/>
      <ItemData ItemOID="IT.TA.ARM" Value="Placebo"/>
      <ItemData ItemOID="IT.TA.TAETORD" Value="1"/>
      <ItemData ItemOID="IT.TA.ETCD" Value="SCREEN"/>
      <ItemData ItemOID="IT.TA.ELEMENT" Value="Screening"/>
      <ItemData ItemOID="IT.TA.EPOCH" Value="SCREEN"/>
    </ItemGroupData>
    ""
  </ReferenceData>
</ODM>
```

Define.XML are my metadata description

Relation of OIDS

Define.XML

<Study **OID** = ...>

<MetaDataVersion **OID** = ...>

<ItemGroupDef **OID** = ...>

<ItemRef **ItemOID** = ...>

-->

-->

-->

-->

Dataset.XML

<ClinicalData **StudyOID** = ...>

<ClinicalData **MetaDataVersionOID** = ...>

<ItemGroupData **ItemGroupOID** = ...>

<ItemData **ItemOID** = ...>

Remark: For <ReferenceData> same like <ClinicalData>

Define.XML and Dataset.XML OID mapping

```
<ODM ...
  <Study OID="cdisc01"> ← 1
    <GlobalVariables>
      <StudyName>CDISC01</StudyName>
      <StudyDescription>CDISC Test Study</StudyDescription>
      <ProtocolName>CDISC01</ProtocolName>
    </GlobalVariables>
    <MetadataVersion OID="MDV.CDISC01.SDTMIG.3.1.2.SDTM.1.2" ...> ← 2
    ...
    <ItemGroupDef OID="IG.AE" ← 3
      Domain="AE" Name="AE" Repeating="Yes" IsReferenceData="No"
      SASDatasetName="AE" Purpose="Tabulation"
      def:Structure="One record per adverse event per subject" def:Class="EVENTS"
      def:ArchiveLocationID="LF.AE">
      <Description>
        <TranslatedText xml:lang="en">Adverse Events</TranslatedText>
      </Description>
      <ItemRef ItemOID="IT.STUDYID" OrderNumber="1" Mandatory="Yes" KeySequence="1"/>
      <ItemRef ItemOID="IT.AE.DOMAIN" OrderNumber="2" Mandatory="Yes"/>
      <ItemRef ItemOID="IT.USUBJID" OrderNumber="3" Mandatory="Yes" KeySequence="2"
        MethodOID="MT.USUBJID"/>
      ...
      <ItemRef ItemOID="IT.AE.AETERM" OrderNumber="6" Mandatory="Yes"/>
      <ItemRef ItemOID="IT.AE.AEMODIFY" OrderNumber="7" Mandatory="No"/>
      <ItemRef ItemOID="IT.AE.AEDECOD" OrderNumber="8" Mandatory="Yes" KeySequence="3"/>
      ...
      <def:leaf ID="LF.AE" xlink:href="ae.xml">
        <def:title>ae.xml</def:title>
      </def:leaf>
    </ItemGroupDef>
  ...
</ODM>
```

```
<ODM ...
  <ClinicalData
    StudyOID="cdisc01" ← 1
    MetadataVersionOID="MDV.CDISC01.SDTMIG.3.1.2.SDTM.1.2" ← 2
    <ItemGroupData ItemGroupOID="IG.AE" ← 3 data:ItemGroupDataSeq="1">
      <ItemData ItemOID="IT.STUDYID" Value="CDISC01"/>
      <ItemData ItemOID="IT.AE.DOMAIN" Value="AE"/>
      <ItemData ItemOID="IT.USUBJID" Value="CDISC01.100008"/>
      ...
      <ItemData ItemOID="IT.AE.AETERM" Value="AGITATED"/>
      <ItemData ItemOID="IT.AE.AEMODIFY" Value="AGITATION"/>
      <ItemData ItemOID="IT.AE.AEDECOD" Value="Agitation"/>
      ...
    </ItemGroupData>
  </ClinicalData>
</ODM>
```

One XML file per (SDTM) domain

- One XML file per (SDTM) domain
- NULL and MISSING values not included as an ItemData element
- Same for <ReferenceData>

```
<ClinicalData
  StudyOID="cdisc01"
  MetadataVersionOID="MDV.CDISC01.SDTMIG.3.1.2.SDTM.1.2">
  <!-- Dataset (AE) -->
  <ItemGroupData ItemGroupOID="IG.AE" data:ItemGroupDataSeq="1">
    <ItemData ItemOID="IT.STUDYID" Value="CDISC01"/>
    <ItemData ItemOID="IT.AE.DOMAIN" Value="AE"/>
    <ItemData ItemOID="IT.USUBJID" Value="CDISC01.100008"/>
    <ItemData ItemOID="IT.AE.AESEQ" Value="1"/>
    <ItemData ItemOID="IT.AE.AESPID" Value="1"/>
    <ItemData ItemOID="IT.AE.AETERM" Value="AGITATED"/>
    <ItemData ItemOID="IT.AE.AEMODIFY" Value="AGITATION"/>
    <ItemData ItemOID="IT.AE.AEDECOD" Value="Agitation"/>
    <ItemData ItemOID="IT.AE.AEBODSYS" Value="Psychiatric disorders"/>
    <ItemData ItemOID="IT.AE.AESEV" Value="MILD"/>
    <ItemData ItemOID="IT.AE.AESER" Value="N"/>
    <ItemData ItemOID="IT.AE.AEACN" Value="DOSE NOT CHANGED"/>
    <ItemData ItemOID="IT.AE.AEREL" Value="POSSIBLY RELATED"/>
    <ItemData ItemOID="IT.AE.AESTDTC" Value="2003-05"/>
    <ItemData ItemOID="IT.AE.AESTDY" Value="3"/>
    <ItemData ItemOID="IT.AE.AEENRF" Value="AFTER"/>
  </ItemGroupData>
  ...
  <ItemGroupData ItemGroupOID="IG.AE" data:ItemGroupDataSeq="16">
    <ItemData ItemOID="IT.STUDYID" Value="CDISC01"/>
    <ItemData ItemOID="IT.AE.DOMAIN" Value="AE"/>
    <ItemData ItemOID="IT.USUBJID" Value="CDISC01.200002"/>
    <ItemData ItemOID="IT.AE.AESEQ" Value="3"/>
    <ItemData ItemOID="IT.AE.AESPID" Value="2"/>
    <ItemData ItemOID="IT.AE.AETERM" Value="PALPITATIONS INTERMITTENT"/>
    <ItemData ItemOID="IT.AE.AEDECOD" Value="Palpitations"/>
    <ItemData ItemOID="IT.AE.AEBODSYS" Value="Cardiac disorders"/>
    <ItemData ItemOID="IT.AE.AESEV" Value="MILD"/>
    <ItemData ItemOID="IT.AE.AESER" Value="N"/>
    <ItemData ItemOID="IT.AE.AEACN" Value="DOSE NOT CHANGED"/>
    <ItemData ItemOID="IT.AE.AEREL" Value="NOT RELATED"/>
    <ItemData ItemOID="IT.AE.AESTDTC" Value="2004-01-05"/>
    <ItemData ItemOID="IT.AE.AESTDY" Value="88"/>
    <ItemData ItemOID="IT.AE.AEENRF" Value="AFTER"/>
  </ItemGroupData>
</ClinicalData>
```

XML schemas and extensions

For Technical Reference

cdisc-dataset-1.0.0
cdisc-define-2.0
cdisc-odm-1.3.2

dataset1-0-0.xsd
dataset-extension.xsd
dataset-ns.xsd

```
<?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:data="http://www.cdisc.org/ns/Dataset-XML/v1.0"
  FileType="Snapshot"
  ODMVersion="1.3.2"
  data:DatasetXMLVersion="1.0"
  FileOID="www.cdisc.org.Studydisc01-Define-XML_2.0.0(IG.TA)"
  PriorFileOID="www.cdisc.org.Studydisc01-Define-XML_2.0.0"
  Originator="CDISC Dataset-XML Team"
  CreationDateTime="2014-03-20T21:45:33" >
  <ReferenceData
    StudyOID="cdisc01"
    MetadataVersionOID="MDV.CDISC01.SDTMIG.3.1.2.SDTM.1.2">
    <!-- Dataset (TA) -->
    <ItemGroupData ItemGroupOID="IG.TA" data:ItemGroupDataSeq="1">
      <ItemData ItemOID="IT.STUDYID" Value="CDISC01"/>
      <ItemData ItemOID="IT.TA.DOMAIN" Value="TA"/>
      <ItemData ItemOID="IT.TA.ARMCD" Value="PLACEBO"/>
      <ItemData ItemOID="IT.TA.ARM" Value="Placebo"/>
      <ItemData ItemOID="IT.TA.TAETORD" Value="1"/>
    </ItemGroupData>
  </ReferenceData>
</ODM>
```

Data Types

- All data types follow the definition in the define.xml specification.
- For converting floats, the define.xml should contain entries in the 'SignificantDigits' and 'Length' attributes for the item definition in <ItemDef> element.

Tools & Ressourcess available

http://wiki.cdisc.org/display/PUB/CDISC+Dataset-XML+Resources

PUBLIC

CDISC Dataset-XML Resources

3

Added by Sam Hume, last edited by Sam Hume on Jul 17, 2014 (view change) show comment

Tools

Introduction

This page contains information on many of the tools currently available or under development for use with Dataset-XML files. Dataset-XML is a new CDISC XML Technologies standard developed as a data transport format for representing SDTM, SEND, ADaM, and legacy datasets. Dataset-XML functions as an alternative to SAS Version 5 Transport (XPT) for the transmission of datasets. More information can be found on this standard at the [CDISC web site](#).

Dataset-XML Tool Summary

Name	Description	Provided By	Links
XPT2DatasetXML	- Transforms XPT datasets into Dataset-XML datasets - Freely available	XML4Pharma	Available under the Smart SDS-XML View project on source forge
Smart Dataset-XML Viewer	- Similar to the SAS Viewer, but with additional functionality - Supports working with Define-XML + Dataset-XML files - Supports SDTM, SEND, and ADaM data - Basic validation - Open source	Univ. Appl. Sciences FH Joanneum Graz - eHealth	- The application and tutorial is available under the Smart SDS-XML View project on source forge Youtube video on the Smart Dataset-XML Viewer
EZ Convert	- Converts Dataset-XML files into SAS datasets - Supports Define-XML Version 1 or Version 2 - Open Source	@ Sally Cassells	EZConvert Demonstration video Beta version of EZConvert
SAS Clinical Standards Toolkit	- Dataset-XML support (writing/reading/validation) will be part of the next release of SAS® Clinical Standards Toolkit. Updated information will be published at the SAS web site. - Support for Dataset-XML is available as a pre-production package that contains SAS macros, XML schema files, sample data, and sample programs to support the following functionality: - Creating Dataset-XML files from SAS data sets - Creating SAS data sets from Dataset-XML files - Validating Dataset-XML files against an XML schema - Comparing original SAS data sets with SAS data sets created from Dataset-XML files - These macros are standalone and do not require SAS® Clinical Standards Toolkit.	SAS Institute Inc.	SAS Clinical Standards Toolkit SAS Macros to support Dataset-XML v1.0.0
OpenCDISC v1.5	OpenCDISC v1.5 works with Dataset-XML files and Define-XML v2.0	OpenCDISC	OpenCDISC.org
R4CDISC	R4CDISC package includes functions for reading Dataset-XML and Define-XML files.	Ippei Akiya	CRAN project page with downloads Reference manual

Dataset-XML Test Datasets

The following attached .zip files contain Dataset-XML datasets and a Define-XML file created by xml4pharma from the LZTZ test datasets.

[Files_from_LZZT_Pilot_2013_Dataset-XML_OK.zip](#)

Status and Future

Question: XML, size and process in IT systems?

- Answer: You should not open an XML file in 'notepad', instead use 'vi(m)'
 - Answer: For transport you may compress the file (ASCII compress perfect)
 - Answer: Serialize / and or process serial
 - Answer: In-Memory usage
-
- XML file size should not be a limit

Where we are (all) today?

- Scope: Functionality required to replace XPT files for FDA submission
- XPT limitations removed for a modern, vendor neutral format
- One file per dataset
- Supports SDTM concepts incl. SuppQuals, RELREC and tabular structures in general

Where we are (all) today? (cont.)

- Support international character sets

- FDA have announced a pilot project

<http://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/ucm380756.htm>

- Several tools developed

2. CDISC SDS-XML Study Data Exchange Pilot evaluation.

Following a Federal Register Notice, FDA held a public meeting on November 5, 2012, entitled "Regulatory New Drug Review: Solutions for Study Data Exchange Standards." The purpose of the public meeting was to solicit input from industry, technology vendors, and other members of the public regarding the advantages and disadvantages of current and emerging open, consensus-based standards for the exchange of regulated study data. FDA indicated, in the Notice and at the meeting, based on feedback received at the public meeting and other information sources, it would undertake further requirements analysis in support of expected evaluation projects.

FDA envisions several pilot projects conducted to evaluate new transport formats. The purpose of this pilot project is to obtain additional experience with CDISC SDS XML format. A successful pilot may allow CDER and CBER to routinely receive study data that employ CDISC SDS XML format as the transport format once an alternatives analysis is completed. As part of this pilot, FDA would like to have sponsors participate in the preparation and submission of previously submitted study datasets using the SDS XML transport format.

Participate
regulatory
archive
applications.

Page Last Updated: 02/05/2014

Future

- Incremental transfers
- Audit records
- Different ways of grouping data
 - All in one file, one file per subject, one file per domain per subject
- Typed data
- Inline Supplemental Qualifiers (once in SDTM)
- Enables SDTM to remove 200 char limit on fields

Demo

Study Dataset XML-Viewer

01-701-1145	1145			
01-701-1015	1146	2013-05-...	2013-06-...	2013-05-...
01-701-1148	1148	2013-08-...	2014-02-...	2013-08-...
01-701-1153	1153	2013-09-...	2014-04-...	2013-09-...

ERROR: Duplicate USUBJID in DM dataset (USUBJID)

2013-03-32	2013-02-...	2013-03-...
2014-07-09	2014-07-...	2014-07-...

WARNING: Non-existent date (RFENDTC)

USUBJID	SUBJID	RFSTDTC	RFXSTDTC	RFXENDTC
01-701-1015	1015	2014-01-02	2014-01-02	2014-07-02
01-701-1023	1023	2012-08-05	2012-08-02	2012-09-01
01-701-1028	01-701-1015 (USUBJID)			1-14
01-701-1033	First date of study treatment exposure = Thu Jan 02 2014			3-31
01-701-1034	Last date of study treatment exposure = Wed Jul 02 2014			2-30

LBDTC	...	LBTMSHI	ENDPOINT
2014-09-25T15:40	...	0.3	
2014-10-21T09:52	...	0.4	
2014-11-18T12:23	...	0.4	
2014-12-...	2014-10-21T09:52 (LBDTC)		Y
2014-12-...	LBDY = 113		

COMPLT24	COMPLT8	EFFICACY	ITT	SAFE
Y	Label: Completers of Week 24 Population Flag			Y
	Mandatory = No			Y
Y	Datatype: text			Y
	Length: 1			Y
Y	CodeList: YN			Y

DM	RELREC	AE	DS	LB	SUPPAE
STUDYID		RDOMAIN		USUBJID	
CDISC01		AE		CDISC01.200001	
CDISC01		DS		CDISC01.200001	

Message



There are 2 related records with RELID '1':
Domain = AE - USUBJID = CDISC01.200001 - AESPID = 5
Domain = DS - USUBJID = CDISC01.200001 - DSSEQ = 3

DM	RELREC	AE	DS	LB	SUPPAE
STUDYID	DOMAIN	USUBJID	AESEQ	AESPID	AETERM
CDISC01	AE	CDISC01.100008	1	1	AGITATED
CDISC01	AE	CDISC01.100008	2	2	ANXIETY
CDISC01	AE	CDISC01.100008	3	3	DECREAS...
CDISC01	AE	CDISC01.100014	1	1	DIARRHEA
CDISC01	AE	CDISC01.100014	2	2	HEMORRH...
CDISC01	AE	CDISC01.100014	3	3	HEADACHE
CDISC01	AE	CDISC01.100014	4	4	VOMIT
CDISC01	AE	CDISC01.100014	5	5	VOMIT
CDISC01	AE	CDISC01.200001	1	1	ANXIETY
CDISC01	AE	CDISC01.200001	2	5	LEFT KNE...

LBSTRESN	LBSTRESU	LBSTNRLO	LBSTNRHI	LBNRIND
38	g/L	33	49	NORMAL
34	U/L	35	115	LOW
50	U/L	35	115	NORMAL
41	U/L	35	115	NORMAL

ATTENTION: value is out of normal range (LBSTRESN)

Study Dataset XML-Viewer WebService version

- Calls external WebServices for more information
- Idea is to call CDISC SHARE in the future

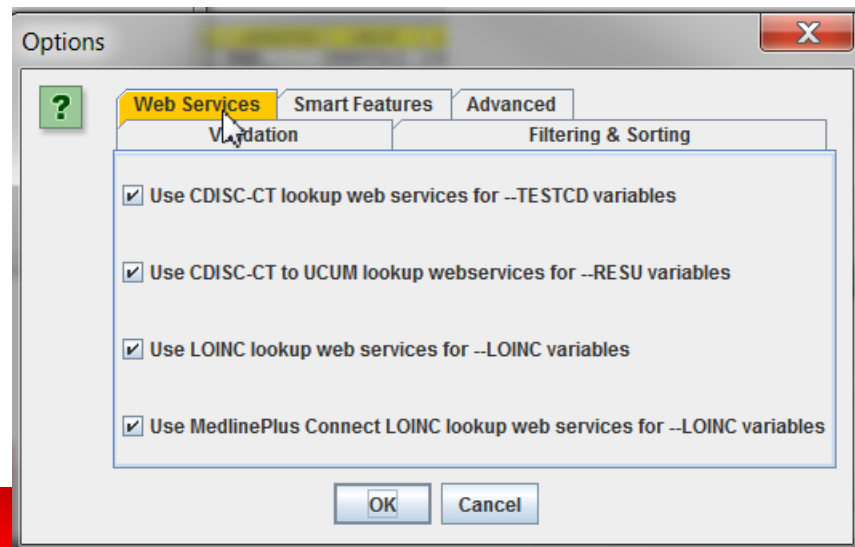
Q	LBTESTCD	LBCAT	L
	RBC	HEMATOLO...	3.9
	RBC	HEMATOLO...	4.0
	RBC (LBTESTCD)	LO...	3.9
	RBC Erythrocytes	LO...	4.8
	RBC NCI: C51946	LO...	4.9
	RBC	HEMATOLO...	4.7
	RBC	HEMATOLO...	4.4

LBLOINC	LBBFL	VISITNUM	VISIT	VISITDY	LBOTC
26453-1		7	WEEK 6	42	2014-06-04...
26453-1		8	WEEK 8	56	2014-06-18...
26453-1		10	WEEK 16	112	2014-08-06...
26453-1	26453-1 (LBLOINC)				2013-11-24...
26453-1	LOINC Name: Erythrocytes.NCnc:PtBld.Qn				2013-12-23...
26453-1	LOINC Common Name: Erythrocytes [#volume] in Blood				2014-01-04...
26453-1	Example UCUM Units: 10*6/uL				2014-01-20...
26453-1		8	WEEK 8	56	2014-02-03...
26453-1		9	WEEK 12	84	2014-02-16...
26453-1	Y	1	SCREENIN...	-7	2013-09-26...

DIABP

CDISC Definition:

The blood pressure after the contraction of the heart while the chambers of the heart refill with blood. (NCI)



MedlinePlus
Trusted Health Information for You

A service of the U.S. National Library of Medicine
NIH National Institutes of Health

Home About MedlinePlus Site Map FAQs Contact Us

Search MedlinePlus GO

Health Topics Drugs & Supplements Videos & Cool Tools ESPAÑOL

RBC count

An RBC count is a blood test that measures how many red blood cells (RBCs) you have.

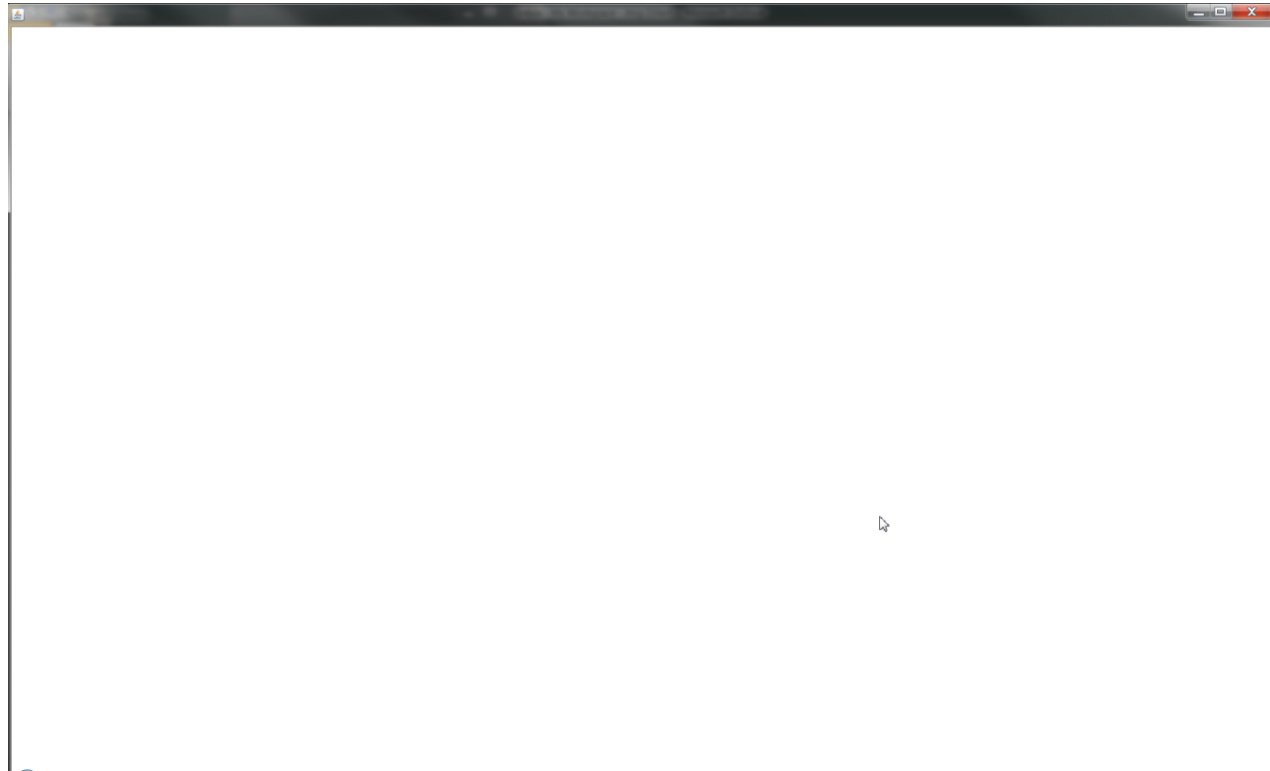
RBCs contain [hemoglobin](#), which carries oxygen. How much oxygen your body tissues get depends on how many RBCs you have and how well they work.

MedlinePlus Topics

- [Blood](#)
- [Blood Disorders](#)

Study Dataset XML-Viewer Webservice version

- Problem to see anything?
- VPN? PROXY?



Study Dataset XML-Viewer – 64bit JAVA recommended

```
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation. All rights reserved.

d:\PhUSE demo\Smart Dataset-XML Webservices>Smart_Dataset-XML_Viewer.bat

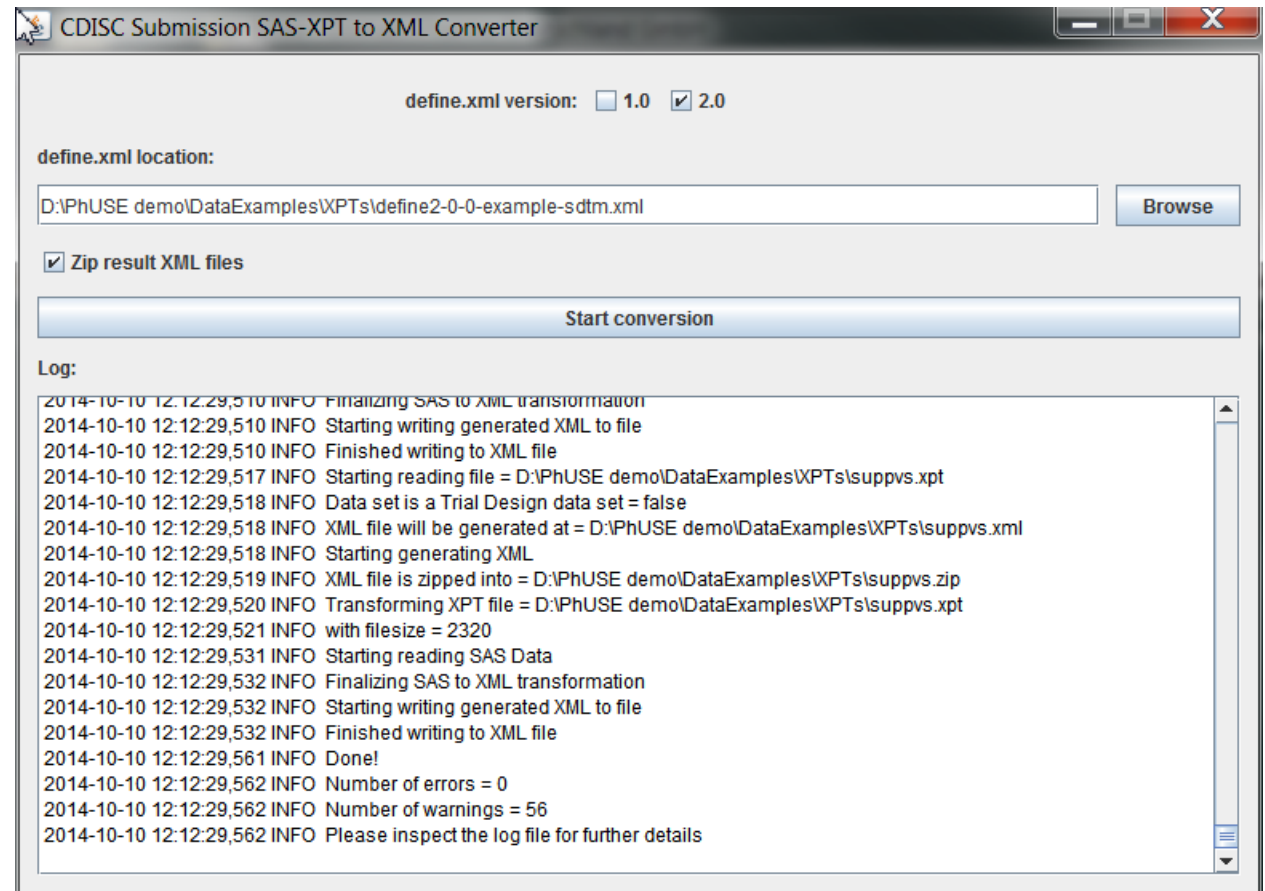
d:\PhUSE demo\Smart Dataset-XML Webservices>#set JAVA_HOME=c:\jdk1.6.0_12
'#set' is not recognized as an internal or external command,
operable program or batch file.

d:\PhUSE demo\Smart Dataset-XML Webservices>#set path=.;\bin;.
'#set' is not recognized as an internal or external command,
operable program or batch file.

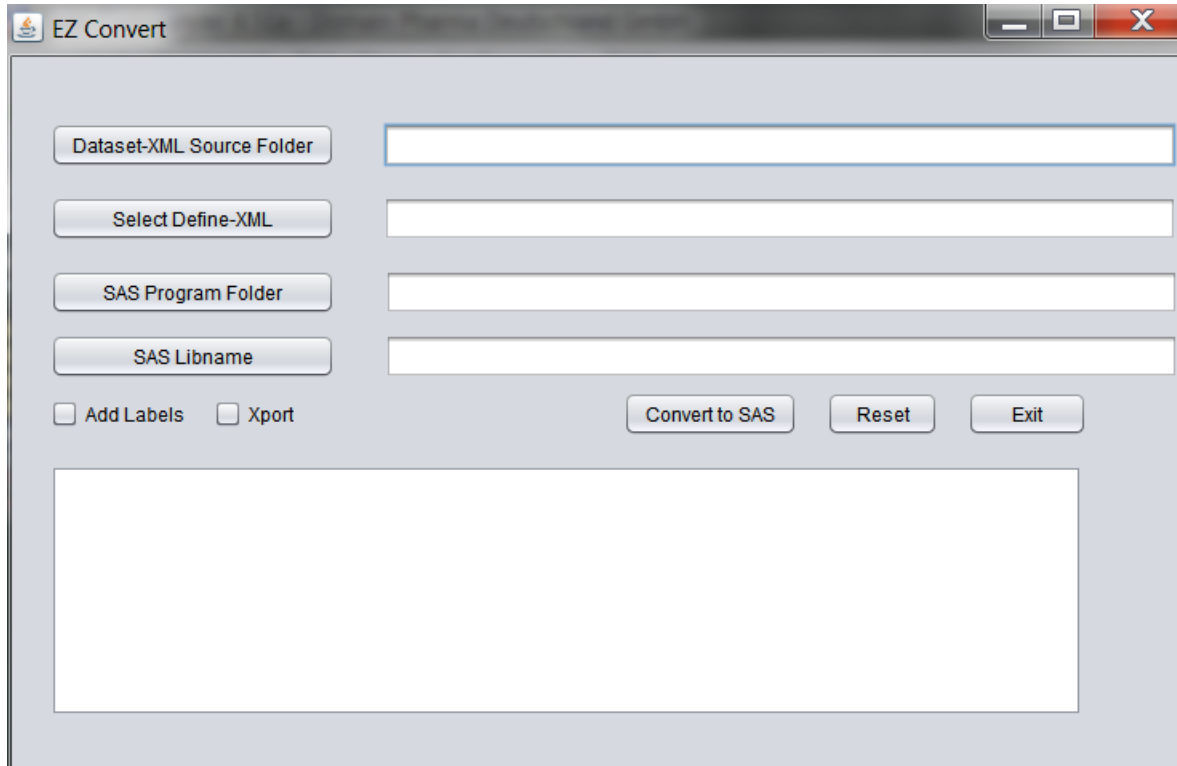
d:\PhUSE demo\Smart Dataset-XML Webservices>set CLASSPATH=.\Smart_Dataset-XML_Viewer.jar;.\vtd-xml.jar;.\log4j-1.2.13.jar;.\zip4j_1.3.2.jar;.\jersey-client-1.18.jar;.\jersey-core-1.18.jar;.\jsr311-api-1.1.1.jar;.\asm-3.1.jar

d:\PhUSE demo\Smart Dataset-XML Webservices>java -Xms256M -Xmx2400M -cp .\Smart_Dataset-XML_Viewer.jar;.\vtd-xml.jar;.\log4j-1.2.13.jar;.\zip4j_1.3.2.jar;.\jersey-client-1.18.jar;.\jersey-core-1.18.jar;.\jsr311-api-1.1.1.jar;.\asm-3.1.jar edu.fhjoanneum.ehealth.smartdatasetxmlviewer.gui.GUI
Error occurred during initialization of VM
Could not reserve enough space for object heap
Error: Could not create the Java Virtual Machine.
Error: A fatal exception has occurred. Program will exit.
d:\PhUSE demo\Smart Dataset-XML Webservices>
```

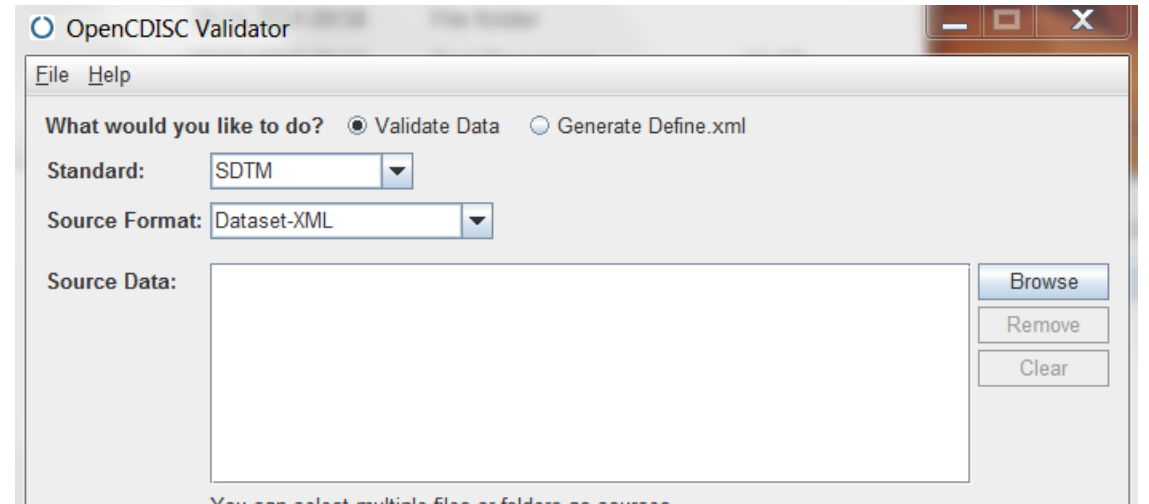
XPT2DataSetXML Converter



EZ Convert – OpenCDISC Validator



The EZ Convert application window features a title bar with the text "EZ Convert" and standard window controls. The main interface includes four input fields on the right, each preceded by a button: "Dataset-XML Source Folder", "Select Define-XML", "SAS Program Folder", and "SAS Libname". Below these fields are two checkboxes, "Add Labels" and "Xport", and three buttons: "Convert to SAS", "Reset", and "Exit". A large empty rectangular box is positioned at the bottom of the window.



The OpenCDISC Validator application window has a title bar with the text "OpenCDISC Validator" and standard window controls. It includes a menu bar with "File" and "Help". The main area contains a section titled "What would you like to do?" with two radio buttons: "Validate Data" (selected) and "Generate Define.xml". Below this are two dropdown menus: "Standard:" with "SDTM" selected, and "Source Format:" with "Dataset-XML" selected. A "Source Data:" label is followed by a large empty text area. To the right of this area are three buttons: "Browse", "Remove", and "Clear". At the bottom of the window, a line of text reads: "You can select multiple files or folders as sources."

No Magic!

Hardware and Software

Engineered to Work Together

ORACLE®