
ORACLE®

CDISC Dataset-XML

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

Mike Collinson, Life Sciences Architect

Oracle Health Sciences

PhUSE SDE London
May 2015

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 Dataset-XML

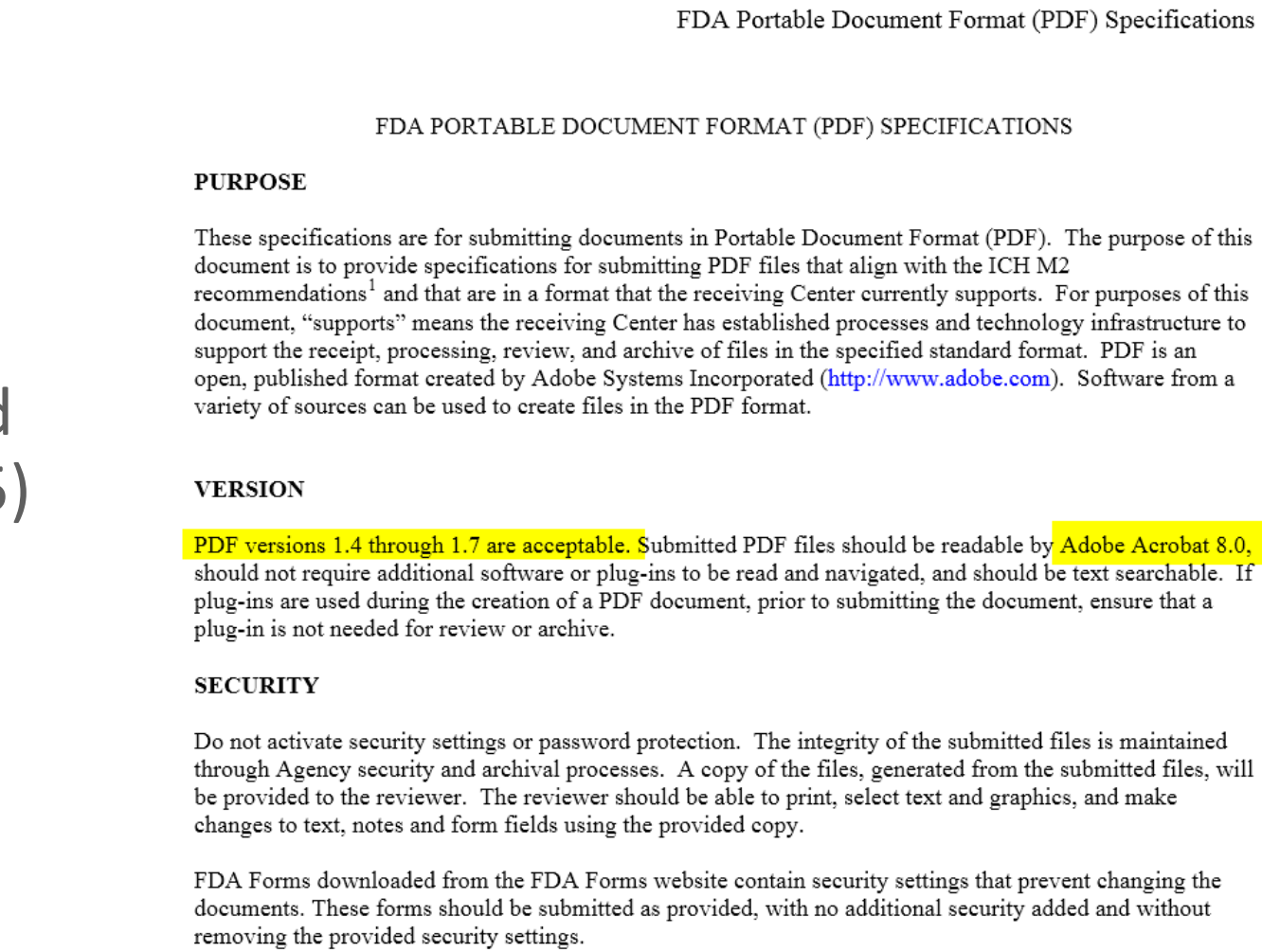
1. History of SAS®-XPT in Electronic Submisssions
2. CDISC Dataset-XML
3. Data Structure
4. Status and Future

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)



FDA Portable Document Format (PDF) Specifications

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 recommendations[1] 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

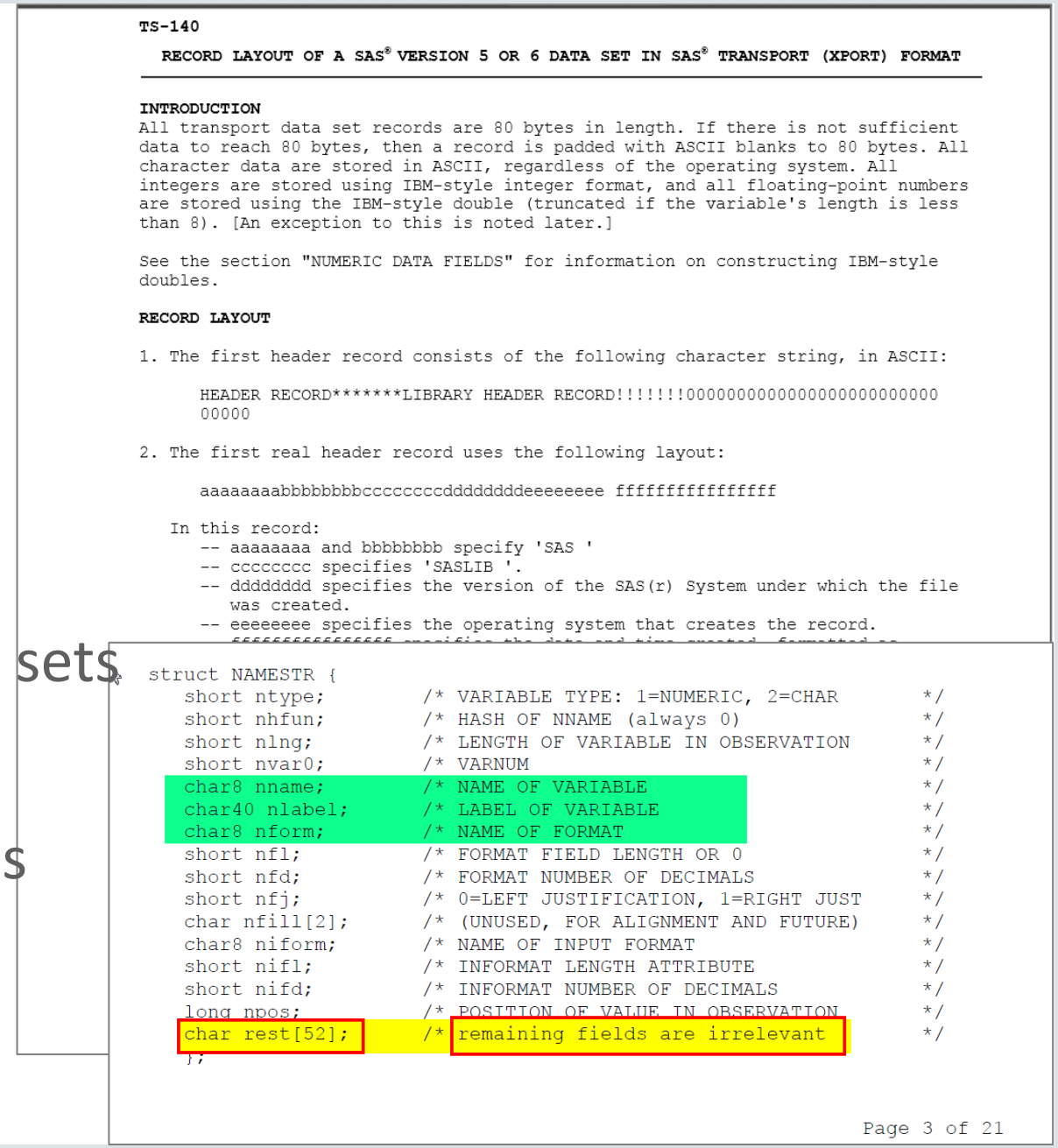
Do not activate security settings or password protection. The integrity of the submitted files is maintained through Agency security and archival processes. A copy of the files, generated from the submitted files, will be provided to the reviewer. The reviewer should be able to print, select text and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

What‘s wrong with XPT?

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

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



TS-140

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

INTRODUCTION

All transport data set records are 80 bytes in length. If there is not sufficient data to reach 80 bytes, then a record is padded with ASCII blanks to 80 bytes. All character data are stored in ASCII, regardless of the operating system. All integers are stored using IBM-style integer format, and all floating-point numbers are stored using the IBM-style double (truncated if the variable's length is less than 8). [An exception to this is noted later.]

See the section "NUMERIC DATA FIELDS" for information on constructing IBM-style doubles.

RECORD LAYOUT

1. The first header record consists of the following character string, in ASCII:

HEADER RECORD*******LIBRARY HEADER RECORD!!!!!!!0000000000000000000000000
00000

2. The first real header record uses the following layout:

aaaaaaaabbbbbbbbccccccccddddddddeeeeeeee ffffffffffffffff

In this record:
-- aaaaaaaa and bbbbbbbb specify 'SAS '
-- cccccccc specifies 'SASLIB '.
-- dddddddd specifies the version of the SAS(r) System under which the file was created.
-- eeeeeeee specifies the operating system that creates the record.

```
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 niform;       /* 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

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

Dataset-XML

- Originally called **Study Data Set XML (SDS-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.Studycdisc01-Define-XML_2.0.0(IG.CM)"
  PriorFileOID="www.cdisc.org.Studycdisc01-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>	- base ODM information as usual
< ReferenceData >	- study information, metadata version
<ItemGroupData>	- table/dataset level
<ItemData>	- 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.Studycdisc01-Define-XML_2.0.0(IG.TA)"
  PriorFileOID="www.cdisc.org.Studycdisc01-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		Dataset.XML
<Study OID = ...>	-->	<ClinicalData StudyOID = ...>
<MetaDataVersion OID = ...>	-->	<ClinicalData MetaDataVersionOID = ...>
<ItemGroupDef OID = ...>	-->	<ItemGroupData ItemGroupOID = ...>
<ItemRef ItemOID = ...>	-->	<ItemData ItemOID = ...>

Remark: For <ReferenceData> same as <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"/>
        … 4 →
        <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 ...
  <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"/>
      ... 4 →
      <ItemData ItemOID="IT.AE.AETERM" Value="AGITATED"/>
      <ItemData ItemOID="IT.AE.AEMODIFY" Value="AGITATION"/>
      <ItemData ItemOID="IT.AE.AEDECOD" Value="Agitation"/>
      ...
    </ItemGroupData>
```

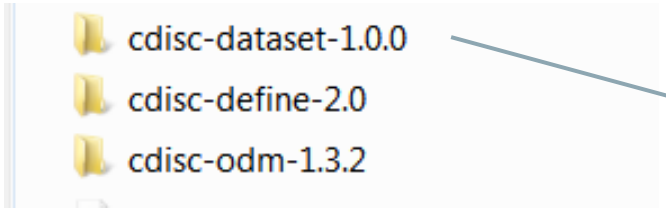
One XML file per (SDTM) domain

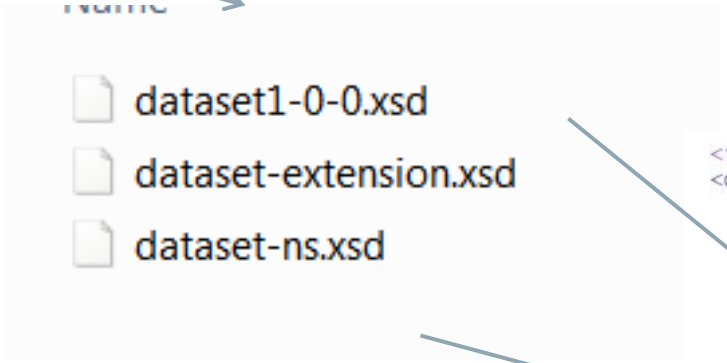
- One XML file per (SDTM) domain
- NULL and MISSING values not included as an ItemData element
- Same for <ReferenceData>

```
<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






```
<?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.Studycdisc01-Define-XML_2.0.0(IG.TA)"
  PriorFileOID="www.cdisc.org.Studycdisc01-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"/>
```

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

Tools

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

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

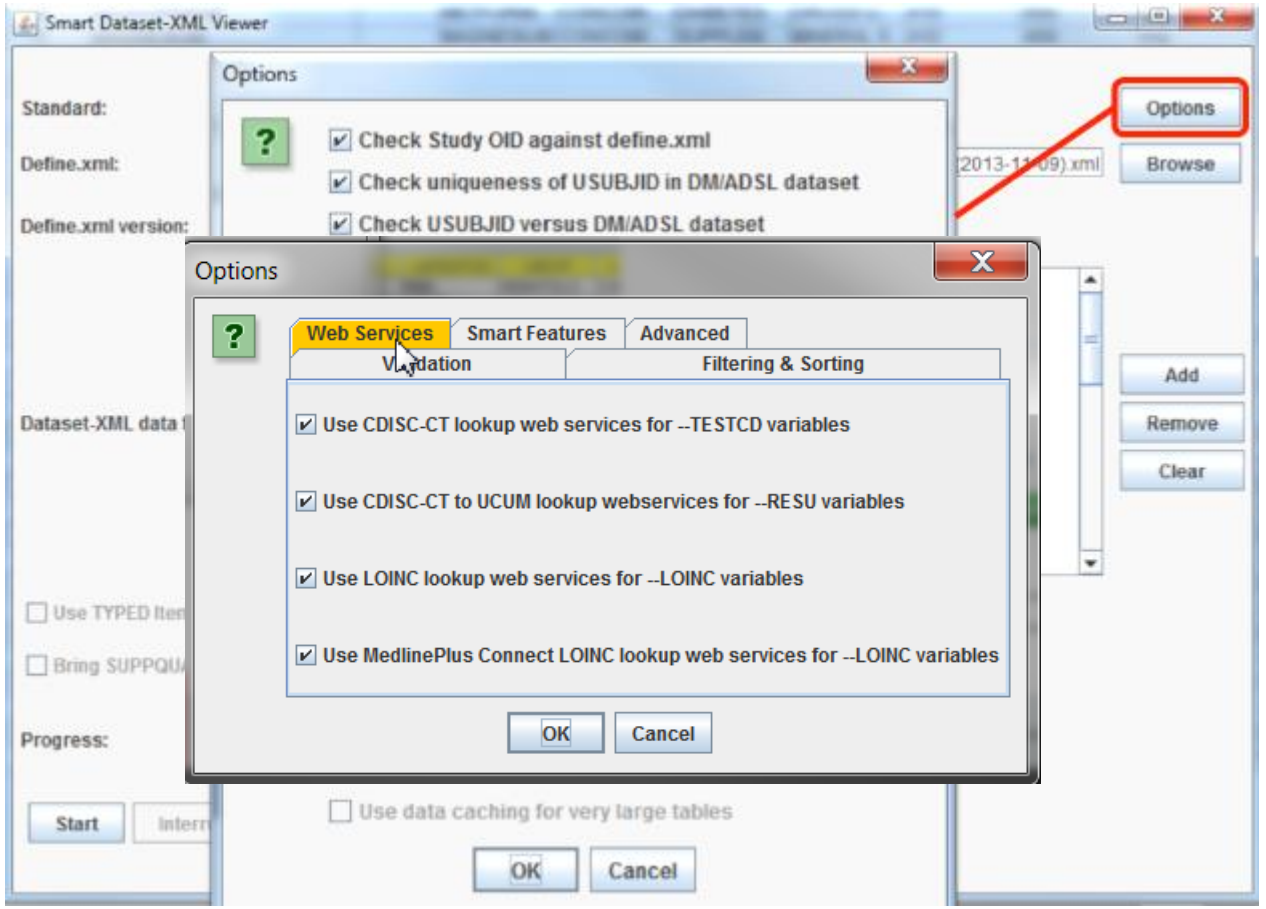
Dataset-XML Test Datasets

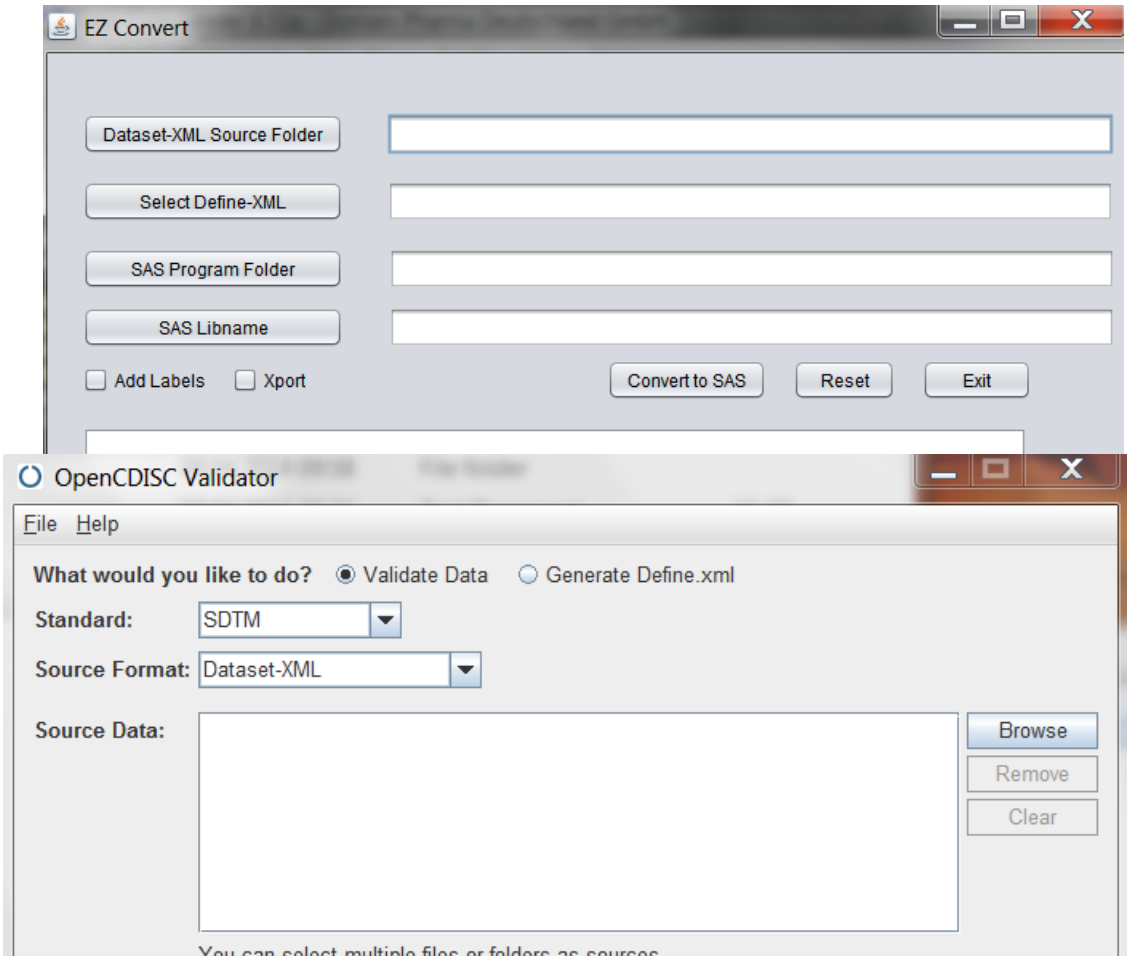
The following attached .zip files contain Dataset-XML datasets and a Define-XML file created by xml4pharma from the LZZT test datasets.

Files_from_LZZT_Pilot_2013_Dataset-XML_OK.zip

Tools & Ressourcess available

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





Status and Future

Where we are today?

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

Where we are today? (cont.)

- FDA have completed a pilot project

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

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.

Participation in this evaluation will be outside of the regulatory pathway and, as such, will not be used to make regulatory decisions. FDA expects that the pilot will assess the technical capability of SDS XML to exchange and archive regulatory study data in investigational new drug applications, new drug applications, and biologics licensing applications.

Question: XML, size and process in IT systems?

- Do not open an XML file in ‘notepad‘, instead use ‘vi(m)‘
- For transport you may compress the file (ASCII compress perfect)
- Serialize / and or process serial where possible
- In-Memory usage

- Will the xml files affect acceptance by the FDA Electronic Gateway?
- XML file size should not be a limitation for submission

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

Hardware and Software
Engineered to Work Together


ORACLE®