Super Demo: Dataset-XML

Lex Jansen
Principal Software Developer @ SAS
CDISC XML Technologies Team
Dataset-XML

What is Dataset-XML?

- Formerly known as StudyDataSet-XML (SDS-XML)
- Alternative to SAS Version 5 Transport (XPT) format for datasets
- Based on CDISC ODM and Define-XML for representation of SDTM, SEND, ADaM or legacy datasets
- Capability to support CDISC data submissions to the FDA
- Based or aligned with Define-XML metadata
- Easy to transform to a dataset for analysis (SAS, R, ...)
Dataset-XML  SAS Version 5 Transport (XPT) limitations

- Dataset and Variable name length limitation (8)
- Dataset and Variable label length limitation (40)
- Variable data lengths limitation (200)
- Limited data types (Char, Num)
Dataset-XML  Dataset-XML - Status

- Final specification for version 1.0 has been released
- Includes sample Define-XML datasets with associated Define-XML file and XML schema
Dataset-XML
New Dataset-XML Standard v1.0 Now Available

The CDISC XML Technologies team is pleased to announce the release of the Dataset-XML v1.0 specification for production use. Dataset-XML, which was released for comment under the name “StudyDataSet-XML” but was renamed to avoid confusion with the CDISC SDS team, is a new standard used to exchange study datasets in an XML format. The purpose of Dataset-XML is to support the interchange of tabular data for clinical research applications using ODM-based XML technologies. The Dataset-XML model is based on the CDISC Operational Data Model (ODM) standard and should follow the metadata structure defined in the CDISC Define-XML standard.

Use this link to download the Dataset-XML 1.0 release package.
Transport Format for the Submission of Regulatory Study Data; Notice of Pilot Project

A Notice by the Food and Drug Administration on 11/27/2013

ACTION
Notice.

SUMMARY
The Center for Drug Evaluation and Research (CDER) and the Center for Biologies Evaluation and Research (CBER) in the Food and Drug Administration (FDA) are announcing a pilot project to evaluate the Clinical Data Interchange Standard Consortium (CDISC) Submission Data Standards (SDS) Extensible Markup Language (XML) transport format for the
Data and Metadata in Submissions Today

Data

SAS V5 XPT

Metadata

Define-XML
Dataset-XML as an Alternative to SAS XPT

- Data
  - Dataset-XML
  - ODM-based Standards

- Metadata
  - Define-XML
Relationship of Dataset-XML to other CDISC Standards

- Define-XML
  - Represents Dataset Metadata
  - Defined by SEND model SEND-IG

- ODM
  - Extended by Dataset-XML

- Dataset-XML
  - Represents Dataset Data
  - follows SDTM model SDTM-IG
  - follows ADaM model ADaM-IG
• Dataset-XML for Data Transport

- Convert SAS datasets to Dataset-XML
- Send Dataset-XML
- Receive Dataset-XML
- Convert to SAS datasets or load into a data warehouse

Data Transport
Benefits of a CDISC Dataset-XML Standard

• Open, non-proprietary standard without field width or dataset and variable naming restrictions of SAS V5 Transport files
• Supports representation of data relationships, metadata versions and audit trails
  • Note: not all of these will be available in the first release
• Harmonized with BRIDG, CDISC Controlled Terminology
• Data elements include references to metadata in Define-XML
• Straightforward implementation starting from SDTM data in SAS
• Supports FDA goal of encouraging open source reviewer tool development
• Facilitates Validation since both data and metadata share underlying technology
• Enables re-thinking some of the length restrictions in standards
Dataset-XML  Dataset-XML for Data Transport

<def:leaf ID="LF.AE" xlink:href="ae.xpt">
  <def:title>ae.xpt</def:title>
</def:leaf>

<def:leaf ID="LF.AE" xlink:href="ae.xml">
  <def:title>ae.xml</def:title>
</def:leaf>
Dataset-XML  Define-XML and Dataset-XML

```xml
<CDM ...>
  <Study OID="cdisc01">
    <GlobalVariables>
      <StudyName>CDISC01</StudyName>
      <StudyDescription>CDISC Test Study</StudyDescription>
      <ProtocolName>CDISC01</ProtocolName>
    </GlobalVariables>
    <MetaDataVersion OID="MDV.CDISC01.SDTMIG.3.1.2.SDTM.1.2" ... />
  </Study>
  <ItemGroupDef OID="IG.AE" 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="No"/>
    <ItemRef ItemOID="IT.AE.ARETRANS" 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>
</CDM>

<CDM ...>
  <ClinicalData StudyID="cdisc01" MetaDataVersionOID="MDV.CDISC01.SDTMIG.3.1.2.SDTM.1.2">
    <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="CDISC01.100000"/>
      ...
      <ItemData ItemOID="IT.AE.AETERM" Value="AGITATION"/>
      <ItemData ItemOID="IT.AE.AEDECOD" Value="Agitation"/>
      ...
    </ItemGroupData>
</CDM>
```

`define.xml`  
`ae.xml`
Dataset-XML Example - Subject Data

<?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="1.0.0"
    FileOID="www.cdisc.org.Studycdisc01-Define-XML_2.0.0(IG.AE)"
    PriorFileOID="www.cdisc.org.Studycdisc01-Define-XML_2.0.0"
    Originator="CDISC Dataset-XML Team" CreationDateTime="2014-04-01T09:31:03">
    <ClinicalData
        StudyID="cdisc01"
        MetaDataVersionOid="MDV.CDISC01.SDTMIG.3.1.2.SDTM.1.2">
        <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.AEENDT" Value="3"/>
            <ItemData ItemOid="IT.AE.AEENRF" Value="AFTER"/>
        </ItemGroupData>
    </ClinicalData>
</ODM>

<table>
<thead>
<tr>
<th>STUDYID</th>
<th>DOMAIN</th>
<th>USUBJID</th>
<th>AESEQ</th>
<th>AESPID</th>
<th>AETERM</th>
<th>AEMODIFY</th>
<th>AEDECOD</th>
<th>AEBODSYS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDISC01</td>
<td>AE</td>
<td>CDISC01.100008</td>
<td>1</td>
<td>1</td>
<td>AGITATED</td>
<td>AGITATION</td>
<td>Agitation</td>
<td>Psychiatric disorders</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AESEV</th>
<th>AESER</th>
<th>AEACN</th>
<th>AEREL</th>
<th>AESTDTC</th>
<th>AEENDTC</th>
<th>AESTDY</th>
<th>AEENDY</th>
<th>AEENRF</th>
</tr>
</thead>
<tbody>
<tr>
<td>MILD</td>
<td>N</td>
<td>DOSE NOT CHANGED</td>
<td>POSSIBLY RELATED</td>
<td>2003-05</td>
<td></td>
<td>3</td>
<td></td>
<td>AFTER</td>
</tr>
</tbody>
</table>
Fields that are not populated do not have any `<ItemData>` elements.

The following examples are **incorrect** in Dataset-XML:

```xml
<ItemData ItemID="IT.AE.AEENTC" Value=""/>
<ItemData ItemID="IT.AE.AEENTC"/>
<ItemData ItemID="IT.AE.AEENTC" IsNull="Yes"/>
```
Dataset-XML  Coming support in SAS

%datasetxml_write()

define.xml

%datasetxml_read()

SAS data sets

Dataset-XML

datasetxml.zip

2014-06-06
Dataset-XML  Available soon

support.sas.com/rnd/base/cdisc/cst/index.html

SAS Clinical Standards Toolkit

SAS® Clinical Standards Toolkit provides support of multiple CDISC standards, including SDTM (3.1.1, 3.1.2, 3.1.3, and 3.2), CRT-DDS (reading and creating define 1.0 xml files), Define-XML 2.0 (reading and creating define 2.0 xml files), ODM (reading and creating 1.3.0 and 1.3.1 xml files), AdA M 2.1, and SEND 3.0. This tool is the platform used by SAS® to support Health and Life Sciences industry data model standards.

The set of new functionality provided in the recent release of SAS Clinical Standards Toolkit 1.6 includes:

- Introduction of a set of Metadata Management tools to add, update, and delete Toolkit metadata
- Full support of CDISC-SDTM 3.2
- Introduction of the new CDISC Define-XML 2.0 standard:
  - A complete definition of the metadata model for CDISC Define-XML 2.0
  - Creation of a complete Define-XML 2.0 file based on study metadata, with study metadata examples from SDTM 3.1.2 and AdA M 2.1
  - Validation of a Define-XML 2.0 file against the XML schema definition, as published by CDISC
  - Import of a Define-XML 2.0 file into the SAS representation of the Define-XML 2.0 metadata model
• DEMO