SAS Super Demo

Supporting Dataset-XML with the SAS® Clinical Standards Toolkit

Lex Jansen
Principal Software Developer @ SAS
CDISC XML Technologies Team
**Dataset-XML**  What is Dataset-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, ...)

---

<table>
<thead>
<tr>
<th>STUDYD</th>
<th>TRT</th>
<th>ROUTE</th>
<th>AESEQ</th>
<th>AESSPID</th>
<th>AETERM</th>
<th>AEMODEFO</th>
<th>AECODED</th>
<th>AEOBSSYS</th>
<th>AEOCT</th>
<th>AEM</th>
<th>AECON</th>
<th>AEREL</th>
<th>AESDTC</th>
<th>AEDNTG</th>
<th>AESDY</th>
<th>AESND</th>
<th>AESEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSOC01</td>
<td>AE</td>
<td>C0031</td>
<td>1</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>AE1</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
</tr>
<tr>
<td>DSOC01</td>
<td>AE</td>
<td>C0031</td>
<td>2</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>AE1</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
</tr>
<tr>
<td>DSOC01</td>
<td>AE</td>
<td>C0031</td>
<td>3</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>AE1</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
</tr>
<tr>
<td>DSOC01</td>
<td>AE</td>
<td>C0031</td>
<td>4</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>AE1</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
</tr>
<tr>
<td>DSOC01</td>
<td>AE</td>
<td>C0031</td>
<td>5</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>AE1</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
<td>ADJ13</td>
</tr>
</tbody>
</table>
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)
- Very limited international character support (only ASCII)
- Documented, but not easy to read/write
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
- Straightforward implementation starting from data in SAS
- Supports FDA goal of encouraging open source reviewer tool development
- Facilitates Validation since both data and metadata share underlying XML technology
- Enables re-thinking some of the length restrictions in standards
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
- FDA conducted successful pilot last year

- Presentation
  SAS Tools for Working with Dataset-XML files
  Tuesday 13:15
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
Dataset-XML  Dataset-XML

Data and Metadata in Submissions Today

Data

Data

Metadata

SAS V5 XPT

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

Data

Dataset-XML

Metadata

Define-XML

ODM-based Standards
Relationship of Dataset-XML to other CDISC Standards

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

- ODM
  - Extended by Dataset-XML
  - Represents Dataset Data

- Dataset-XML
  - Follows SDTM model SDTM-IG
  - Follows ADaM model ADaM-IG

- SEND model SEND-IG
- SDTM model SDTM-IG
- 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
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>
<Study OID="cdisc01">
  <GlobalVariables>
    <StudyName>CDISC01</StudyName>
    <StudyDescription>CDISC Test Study</StudyDescription>
    <ProtocolName>CDISC01</ProtocolName>
  </GlobalVariables>
  <MetaDataVersion OID="MDV.CDISCO1.SDTMIG.3.1.2.SDTM.1.2" ... />

  <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="5" Mandatory="No"/>
    <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>
</Study>

<ClinicalData StudyOID="cdisc01" MetaDataVersionOID="MDV.CDISCO1.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="CDISC01.100009"/>
    ...
    <ItemData ItemOID="IT.AE.AETERM" Value="AGITATION"/>
    <ItemData ItemOID="IT.AE.AEDECOD" Value="Agitation"/>
    ...
  </ItemGroupData>
</ClinicalData>
Dataset-XML Example - Subject Data

```xml
<?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
    StudyOID="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.AEDEV" 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.AEENDTC" Value="2003-05"/>
      <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>AEDEV</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></td>
<td>3</td>
<td>AFTER</td>
</tr>
</tbody>
</table>
Dataset-XML Support in SAS

SAS data

%datasetxml_write()

%xml_validate()

%datasetxml_read()

%cstutilcomparedatasets()

define.xml

Dataset-XML
Expected differences

• Date- and time-related columns may get a different length, since they do not have a length defined in the Define-XML metadata.

• Small differences in precision can be expected around the machine precision for numeric variables that represent real numbers.

• Character data that contains leading spaces or trailing spaces may lose the leading and trailing spaces.
**Dataset-XML**  Available Now

[Image: support.sas.com/rnd/base/cdisc/cst/index.html]

**Base SAS**

**SAS® Clinical Standards Toolkit**

- Introduction
- Ordering the Toolkit
- Toolkit Documentation
  - Papers
    - Version 1.7
    - Version 1.6
    - Version 1.5
    - Version 1.4 and Earlier
- Discussion Forum

**Introduction**

The SAS® Clinical Standards Toolkit provides support of multiple CDISC standards, including SDTM (3.1.2, 3.1.3, and 3.2), CRT-DDS (reading and creating define creating define 2.0 XML files), Dataset-XML (creating Dataset-XML files from SAS data sets and creating SAS data sets from Dataset-XML files), ODM (reading ar 2.1, CDASH 1.1, SEND 3.0, and validating XML files against an XML schema file. This tool is the platform used by SAS® to support Health and Life Sciences industry.

The set of new functionality provided in the recent release of SAS Clinical Standards Toolkit 1.7 includes:

- Introduction of a set of migration tools, previously offered as pre-production software, to help migrate from one version of the SAS Clinical Standards Tool to CDISC CDASH 1.1.
- Enhanced support for creating an initial version of the six SAS source metadata data sets (source_study, source_tables, source_columns, source_code, source_documents) that serve as input for creating a Define-XML 2.0 file.
- Implementation of the CDISC Dataset-XML 1.0 data standard that can be used to transport CDISC SDTM, SEND, and ADaM data sets for submission to regulatory bodies.
- Macros to support Dataset-XML v1.0.0 on support.sas.com for information about a standalone version of the macros that support the CDISC-Dataset XML.
- Reduced and consolidated validation_data sets for SDTM 3.1.2, 3.1.3, 3.2, and ADaM 2.1.
- Updated support for CDISC NCI controlled terminology.
- New framework macros have been added, including cstutilmanagemetadatasize, cstutilcomparemetadatasasdefine, cstutilsqlcolumndefinition, cstutilsqlg_cstutilregistercstsubtypes, and cstutilxmvalidate.
Sample 53447: SAS® Macros to support Dataset-XML v1.0.0

This package contains 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

Documentation is available in this file that is part of the ZIP file: SAS-Dataset-XML-v1.0.0-support.pdf

Users are encouraged to use this new functionality. To help guide future development, post feedback on the SAS in Health Care Related Fields and Clinical Trials community:

http://communities.sas.com/community/support-communities/sas_in_health_care_related_fields

Note: These macros are standalone and do not require SAS® Clinical Standards Toolkit.
DEMO