



# 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, ...)

STUDYID	DOMAIN	USUBJID	ASEQ	AESPID	AETERM	AEMODIFY	AEDECOD	AEBODSYS	AESEV	AESER	AEACN	AEREL	AESTDTC	AEENDTC	AESTDY	AEENDY	AEENR
CDISC01	AE	CDISC01.10...	1	1	AGITATED	AGITATION	Agitation	Psychiatric d...	MILD	N	DOSE NOT ...	POSSIBLY ...	2003-05		3		AFTER
CDISC01	AE	CDISC01.10...	2	2	ANXIETY		Anxiety	Psychiatric d...	MODERATE	N	DOSE NOT ...	POSSIBLY ...	2003-05-13		15		AFTER
CDISC01	AE	CDISC01.10...	3	3	DECREASE...		Decreased a...	Metabolism ...	MILD	N	DOSE NOT ...	POSSIBLY ...	2003-08-19	2003-09-15	113	140	
CDISC01	AE	CDISC01.10...	1	1	DIARRHEA		Diarrhoea	Gastrointesti...	MILD	N	DOSE NOT ...	NOT RELAT...	2004-01-06		84		AFTER
CDISC01	AE	CDISC01.10...	2	2	HEMORRHE...		Haemorrhoids	Gastrointesti...	MODERATE	N	DOSE NOT ...	NOT RELAT...	2004-01-06		84		AFTER
CDISC01	AE	CDISC01.10...	3	3	HEADACHE		Headache	Nervous syst...	MILD	N	DOSE NOT ...	NOT RELAT...	2004-01-27		105		AFTER
CDISC01	AE	CDISC01.10...	4	4	VOMIT	VOMITING	Vomiting	Gastrointesti...	MODERATE	N	DRUG INTE...	POSSIBLY ...	2004-02-03	2004-02-03	112	112	
CDISC01	AE	CDISC01.10...	5	5	VOMIT	VOMITING	Vomiting	Gastrointesti...	SEVERE	Y	DRUG INTE...	POSSIBLY ...	2004-02-04	2004-02-09	113	118	
CDISC01	AE	CDISC01.20...	1	1	ANXIETY		Anxiety	Psychiatric d...	SEVERE	N	DOSE NOT ...	POSSIBLY ...	2003-10-16	2003-10-20	17	21	
CDISC01	AE	CDISC01.20...	2	5	LEFT KNEE ...		Arthralgia	Musculoskel...	SEVERE	N	DRUG WITH...	NOT RELAT...	2004-02-02		126		AFTER
CDISC01	AE	CDISC01.20...	3	3	CONSTIPAT...		Constipation	Gastrointesti...	MODERATE	N	DOSE NOT ...	NOT RELAT...	2003-12-25		87		AFTER
CDISC01	AE	CDISC01.20...	4	4	TIREDFNESS		Fatigue	General diso...	SEVERE	N	DOSE NOT ...	POSSIBLY ...	2003-12-25		87		AFTER
CDISC01	AE	CDISC01.20...	5	2	NAUSEA IN...		Nausea	Gastrointesti...	SEVERE	N	DOSE NOT ...	POSSIBLY ...	2003-10-16	2003-10-20	17	21	
CDISC01	AE	CDISC01.20...	1	3	LIGHTHEAD		Dizziness	Nervous syst...	MILD	N	DOSE NOT ...	NOT RELAT...	2004-02-26	2004-02-26	140	140	
CDISC01	AE	CDISC01.20...	2	1	MUSCLE S...		Muscle spas...	Musculoskel...	MILD	N	DOSE NOT ...	NOT RELAT...	2004-01-05		88		AFTER
CDISC01	AE	CDISC01.20...	3	2	PALPITATIO...		Palpitations	Cardiac diso...	MILD	N	DOSE NOT ...	NOT RELAT...	2004-01-05		88		AFTER

## 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 CDISC Standards



[www.cdisc.org/dataset-xml](http://www.cdisc.org/dataset-xml)



ABOUT CDISC

STANDARDS & INNOVATIONS

RESOURCES

NEWS

EDUCATION & EVENTS

MEMBERSHIP

MEMBERS ONLY

## 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.

# Dataset-XML CDISC Standards



## Notice

### 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 Biologics 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

← Previous Document  
Next Document →

**LEGAL DISCLAIMER**

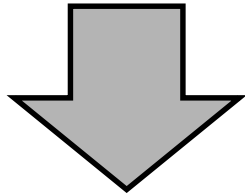
Font Controls + - A A

PDF DEV PRINT  
PUBLIC INSPECTION  
Publication Date:

Dataset-XML Dataset-XML

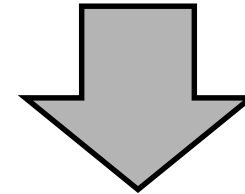
## Data and Metadata in Submissions Today

**Data**



**SAS V5 XPT**

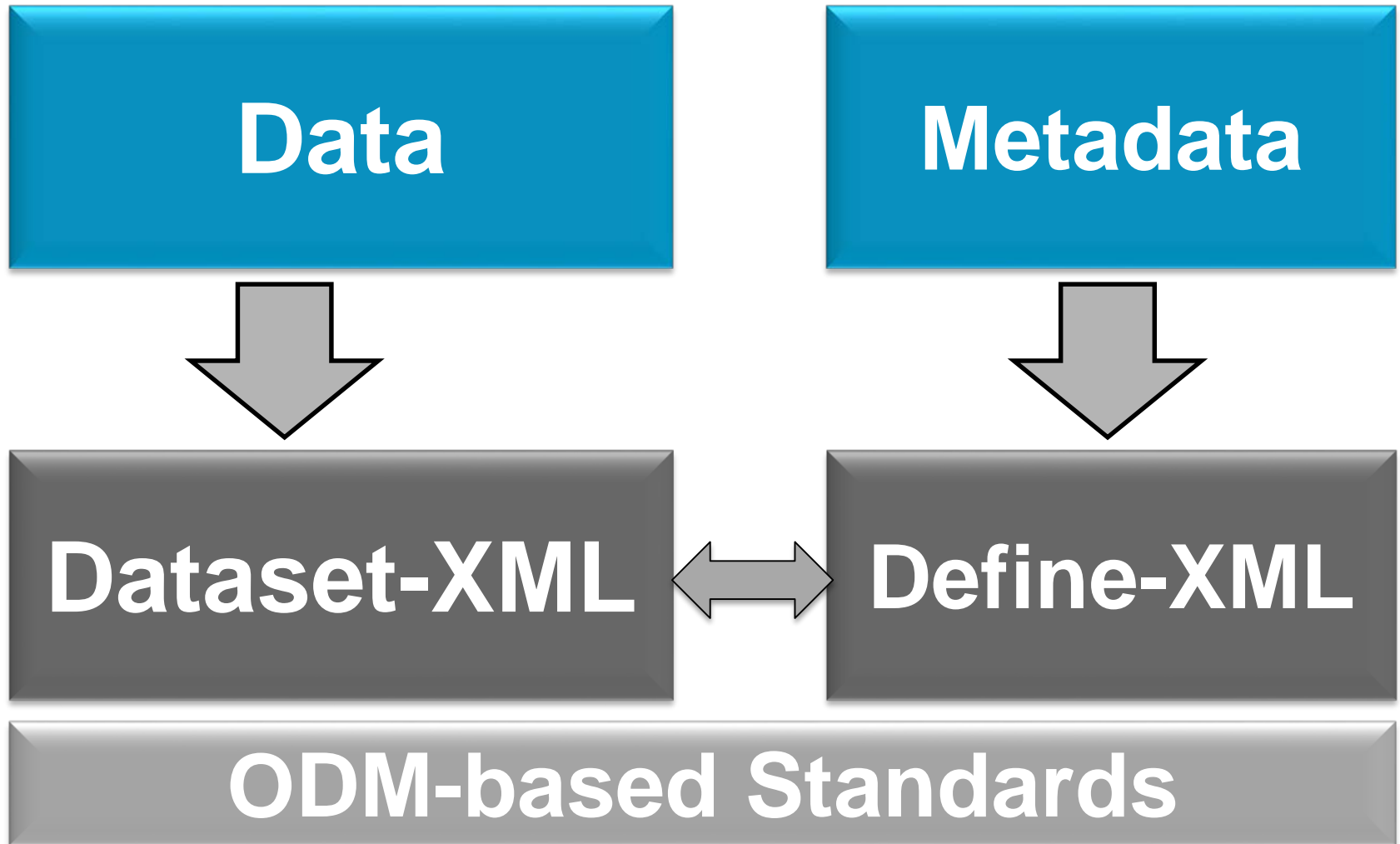
**Metadata**



**Define-XML**

Dataset-XML Dataset-XML

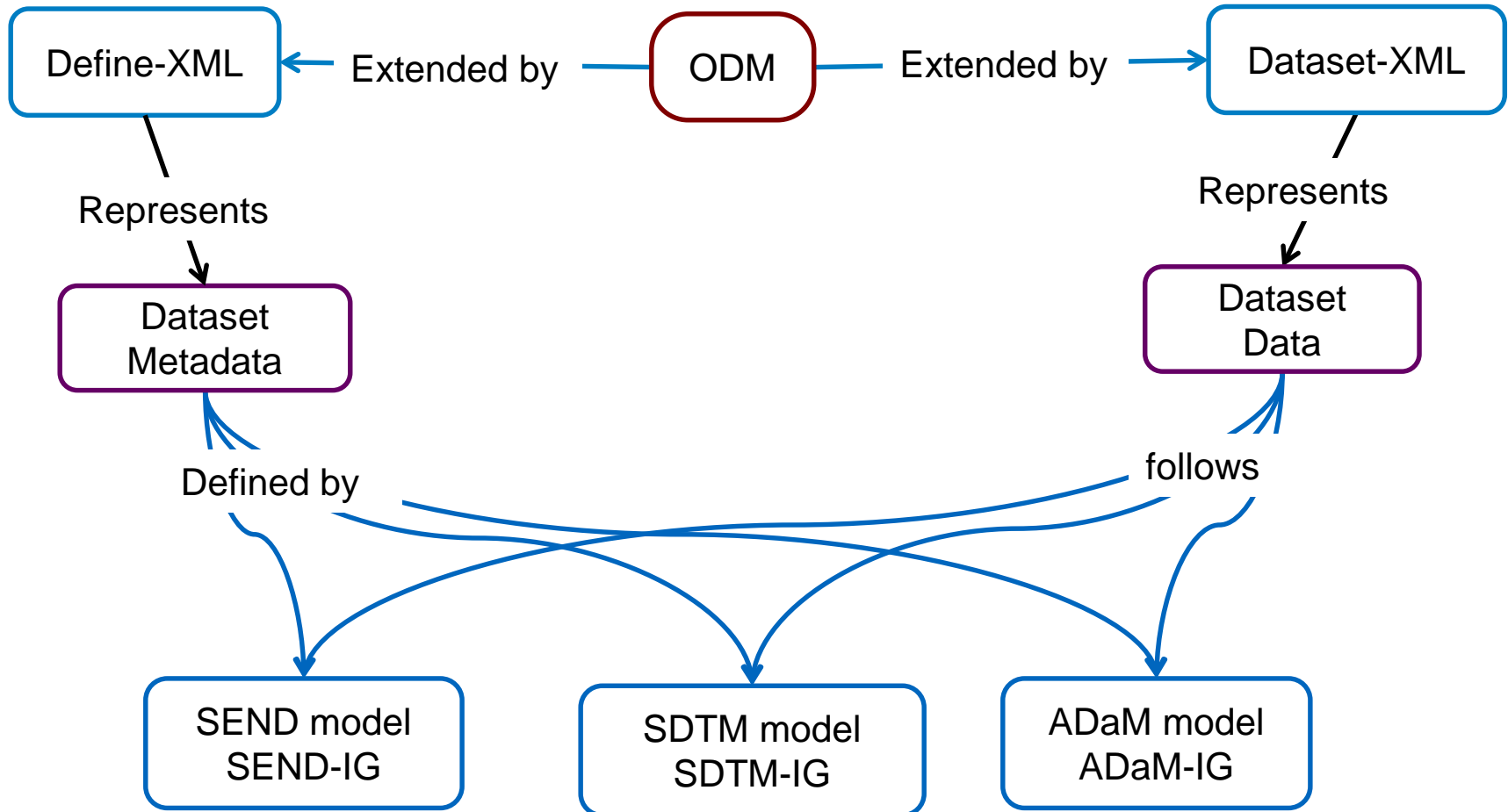
## Dataset-XML as an Alternative to SAS XPT





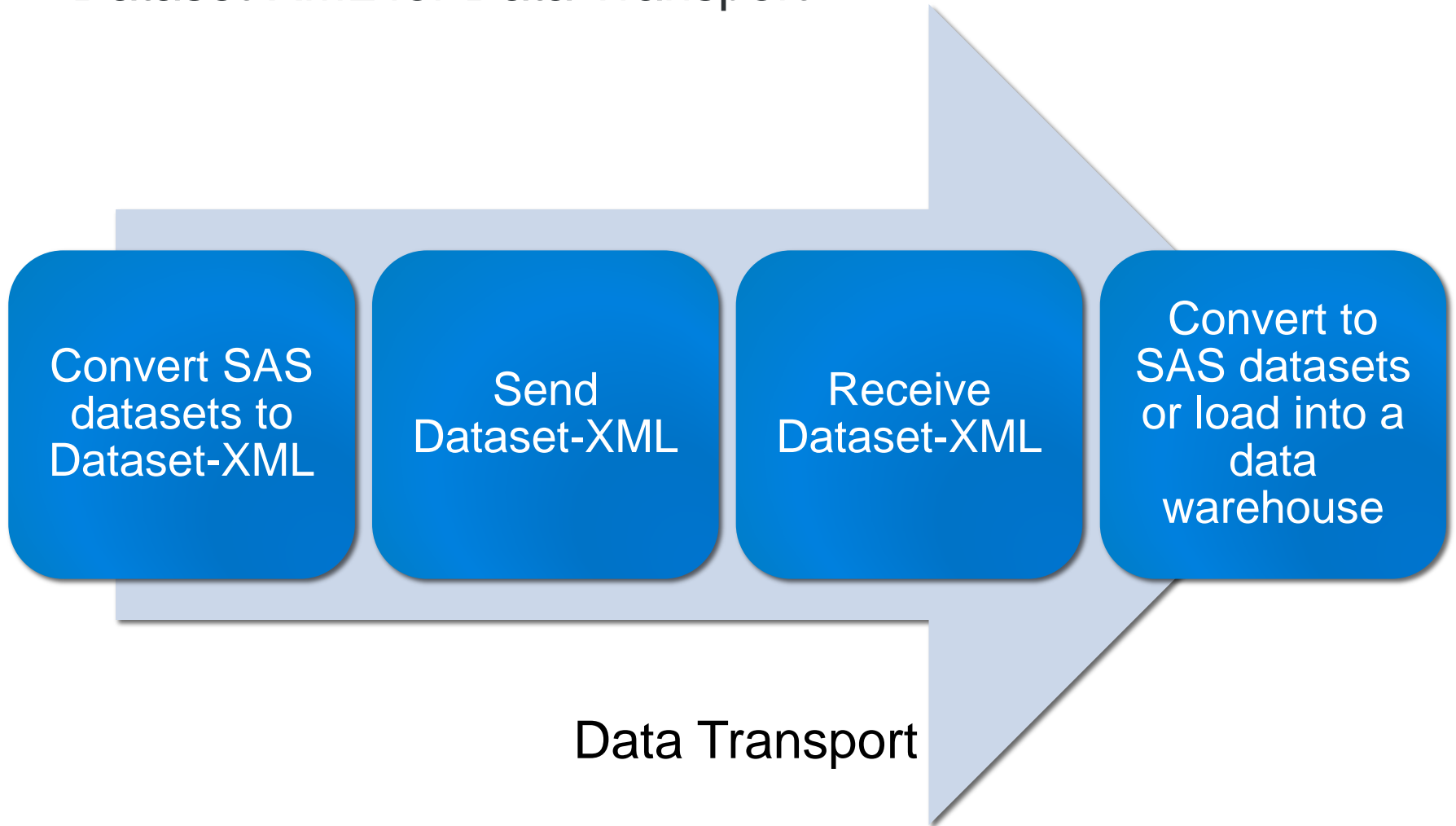
# Dataset-XML Dataset-XML

## Relationship of Dataset-XML to other CDISC Standards



## Dataset-XML Dataset-XML

- Dataset-XML for Data Transport

































## Dataset-XML Dataset-XML

### 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

 ae.xpt	SAS Xport Transport File
 blankcrf.pdf	Adobe Acrobat Document
 cm.xpt	SAS Xport Transport File
 complexalgorithms.pdf	Adobe Acrobat Document
 da.xpt	SAS Xport Transport File
 define.xml	XML Document
 define2-0-0.xsl	XSLT Stylesheet
 dm.xpt	SAS Xport Transport File
 ds.xpt	SAS Xport Transport File
 eg.xpt	SAS Xport Transport File
 ex.xpt	SAS Xport Transport File
 ie.xpt	SAS Xport Transport File
 lb.xpt	SAS Xport Transport File
 mh.xpt	SAS Xport Transport File
 pe.xpt	SAS Xport Transport File

 ae.xml	XML Document
 blankcrf.pdf	Adobe Acrobat Document
 cm.xml	XML Document
 complexalgorithms.pdf	Adobe Acrobat Document
 da.xml	XML Document
 define.xml	XML Document
 define2-0-0.xsl	XSLT Stylesheet
 dm.xml	XML Document
 ds.xml	XML Document
 eg.xml	XML Document
 ex.xml	XML Document
 ie.xml	XML Document
 lb.xml	XML Document
 mh.xml	XML Document
 pe.xml	XML Document

```
<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

```
<ODM ...
<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" ... >

  <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.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>
```

define.xml

```
<ODM ...
<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.AETERM" Value="AGITATION"/>
    <ItemData ItemOID="IT.AE.AEDECOD" Value="Agitation"/>
    ...
  </ItemGroupData>
```

ae.xml

# Dataset-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
    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.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>
  </ClinicalData>

```

	STUDYID	DOMAIN	USUBJID	AESEQ	AESPID	AETERM	AEMODIFY	AEDECOD	AEBODSYS
1	CDISC01	AE	CDISC01.100008	1	1	AGITATED	AGITATION	Agitation	Psychiatric disorders

	AESEV	AESER	AEACN	AEREL	AESTDTC	AEENDTC	AESTDY	AEENDY	AEENRF
1	MILD	N	DOSE NOT CHANGED	POSSIBLY RELATED	2003-05		3		AFTER

## Dataset-XML Fields not populated

	STUDYID	DOMAIN	USUBJID	AESEQ	AESPID	AETERM	AEMODIFY	AEDECOD	AEBODSYS
1	CDISC01	AE	CDISC01.100008	1	1	AGITATED	AGITATION	Agitation	Psychiatric disorders

	AESEV	AESER	AEACN	AEREL	AESTDTC	AEENDTC	AESTDY	AEENDY	AEENRF
1	MILD	N	DOSE NOT CHANGED	POSSIBLY RELATED	2003-05		3		AFTER

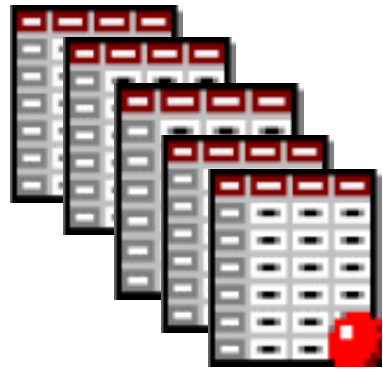
- Fields that are not populated do not have any <ItemData> elements
- The following examples are **incorrect** in Dataset-XML

```
<ItemData ItemOID="IT.AE.AEENTC" Value="" />
```

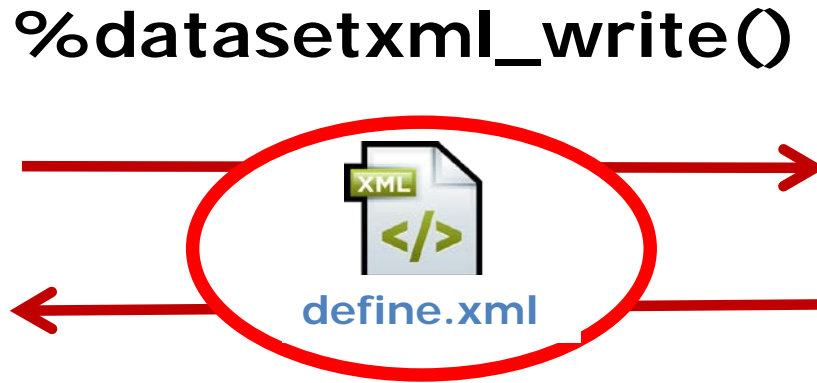
```
<ItemData ItemOID="IT.AE.AEENTC" />
```

```
<ItemData ItemOID="IT.AE.AEENTC" IsNull="Yes" />
```

# Dataset-XML Coming support in SAS



SAS data sets



Dataset-XML



datasetxml.zip



# Dataset-XML Available soon

[support.sas.com/rnd/base/cdisc/cst/index.html](http://support.sas.com/rnd/base/cdisc/cst/index.html)

SAS Clinical Standards Toolkit

support.sas.com Knowledge Base Support Training & Books Happenings Store Support Communities

## KNOWLEDGE BASE / FOCUS AREAS

### FOCUS AREAS

- Base SAS
  - CDISC
    - Clinical Standards Toolkit
    - PROC CDISC
  - DATA Step
  - ODS
  - SAS®
  - Universal Printing
  - XML Engine
  - Preproduction
- Graphics
- Enterprise Management Integration
- Migration
- Scalability & Performance
- Statistics & Operations Research
- SAS AppDev Studio  
2014-06-06

### Base SAS

CDISC Support

#### SAS® Clinical Standards Toolkit

The 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), ADaM 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 ADaM 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



datasetxml.zip

• **DEMO**