

Using the SAS® Clinical Standards Toolkit 1.5 to import CDISC ODM files

Lex Jansen, SAS Institute Inc., Cary, North Carolina, USA

ABSTRACT

The CDISC Operational Data Model (ODM) is a vendor neutral, platform independent XML format for interchange and archival of clinical study data. The model represents study metadata, administrative metadata, reference data and subject data associated with a clinical trial. The ODM format is defined by an XML schema and a specification.

The SAS® Clinical Standards Toolkit (CST) is a framework of SAS macros, metadata and configuration files to support clinical research activities. The CDISC ODM standard is one of the standards that are supported by CST. To enable ODM support, SAS has defined a relational data model that represents the ODM model as SAS data sets. This paper provides an introduction to the structure and content of ODM files and shows how the SAS Clinical Standards Toolkit supports:

- Reading ODM files.
- Schema level validation of an ODM file
- Validating structure and content of the SAS representation of ODM
- Extraction of clinical data and reference data into SDTM SAS data sets
- Importing CDISC NCI Controlled Terminology ODM files

The paper also looks at the inner workings and underlying technologies of the SAS Clinical Standards Toolkit with respect to XML based standards.

INTRODUCTION

This paper discusses release 1.5 of the SAS Clinical Standards Toolkit [1]. The SAS Clinical Standards Toolkit 1.5 is supported with SAS 9.3M2. This release of the Toolkit supports reading and writing of files that conform to the Operational Data Model standard version 1.3.0 and 1.3.1.

SAS® CLINICAL STANDARDS TOOLKIT

The SAS Clinical Standards Toolkit (CST) focuses primarily on supporting clinical research activities. These activities involve the discovery and development of new pharmaceutical and biotechnology products and medical devices. These activities occur from project initiation through product submission and throughout the full product lifecycle.

The SAS Clinical Standards Toolkit initially focuses on standards defined by the Clinical Data Interchange Standards Consortium (CDISC). CDISC is a global, open, multidisciplinary, nonprofit organization that has established standards to support the acquisition, exchange, submission, and archival of clinical research data and metadata. The CDISC mission is to develop and support global, platform-independent data standards that enable information-system interoperability, which, in turn, improves medical research and related areas of health care. The SAS Clinical Standards Toolkit is not limited to supporting CDISC standards. In time, the SAS Clinical Standards Toolkit will support other evolving industry-standard data models. The SAS Clinical Standards Toolkit framework is designed to support the specification and use of any user-defined standard.

The term "toolkit" connotes a collection of tools, products, and solutions. The SAS Clinical Standards Toolkit provides a set of standards and functionality that will evolve and grow with future product updates and releases. Customer requirements and expectations of the SAS Clinical Standards Toolkit will play a key role in the deciding what functionality to provide in future releases.

The SAS Clinical Standards Toolkit 1.5 includes support for the following CDISC standards:

- SDTM 1.3.1, 3.1.2 and 3.1.3
- An initial implementation of the CDISC SEND 3.0 standard, including definition of all domains and columns.
- ADaM 2.1 (ADSL, Basic Data Structure, ADAE) and Analysis Results Metadata templates, as well as new validation checks in support of ADAE and ADTTE.
- CRT-DDS 1.0 (define.xml), including define.pdf and Value Level metadata support.

- ODM 1.3.0 and ODM 1.3.1, including support for the extraction of ODM Clinical data and ODM Reference data into SAS data sets.
- The implementation of CT 1.0.0, a tool to support the import of NCI CDISC Controlled Terminology in the ODM XML format into SAS data sets and SAS format catalogs.
- CDISC Controlled Terminology package that includes the cumulative set of terminology as posted to the [NCI FTP](#) site as of January 2013.
- Furthermore, a set of tools was introduced to validate the SAS Clinical Standards Toolkit metadata itself ("Internal validation").

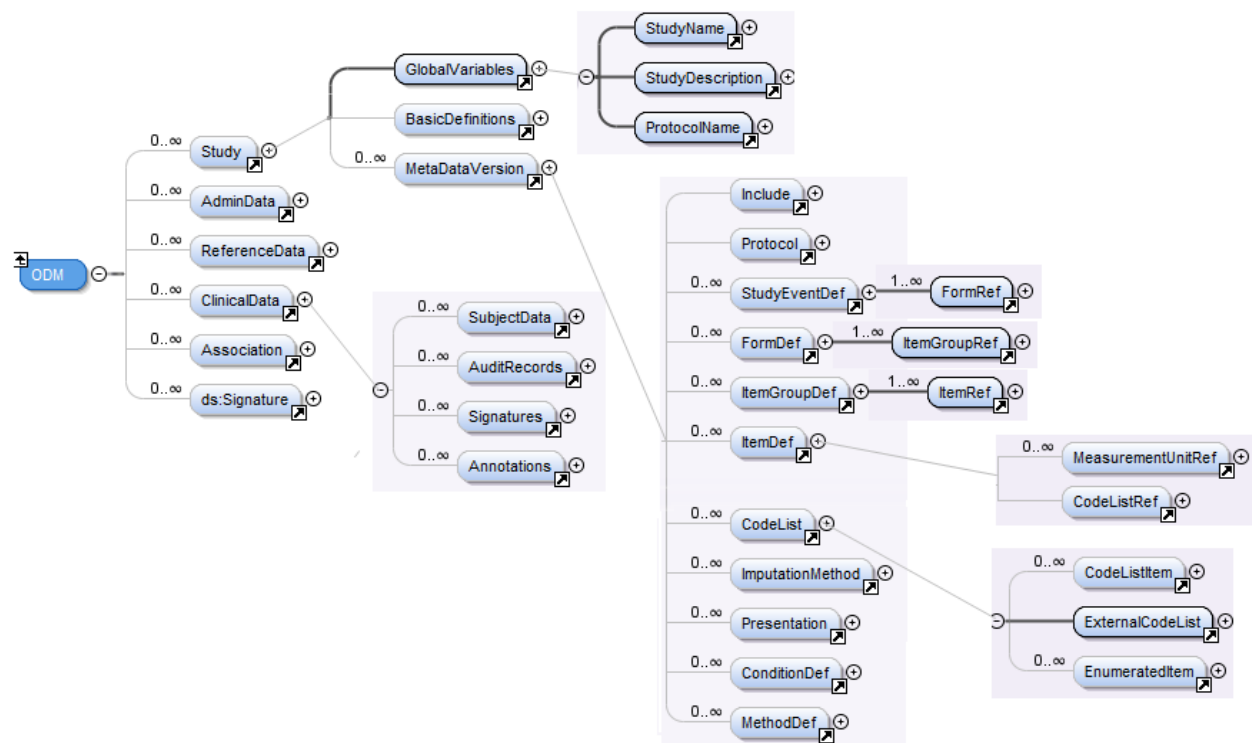
Each SAS Clinical Standards Toolkit standard provides a SAS representation of the published source guidelines or source specification. The SAS representation is designed to serve as a model or template of the source specification. Two key design requirements shaped the implementation of the SAS Clinical Standards Toolkit standards.

- 1) Each supported standard is represented in one or more SAS files. This facilitates these points:
 - It provides SAS users with an implementation of data models and standards that are based on SAS.
 - It enables you to use SAS routines to assess how well any user-defined set of data and metadata conforms to the standard.
 - It enables you to use SAS code to read and derive files in other formats (for example, XML).
 Each SAS Clinical Standards Toolkit standard is an optimized reference standard from a SAS perspective.
- 2) You are able to define your own customized standards, or you are able to modify existing SAS standards.

OPERATIONAL DATA MODEL (ODM)

The CDISC Operational Data Model (ODM) is a vendor neutral, platform independent XML based format for the interchange and archival of clinical study data. The model represents study metadata, administrative metadata, reference data and subject data associated with a clinical trial. The ODM format is defined by an XML schema and a specification. Display 1 partially shows the ODM 1.3.1 element structure. For a full understanding of the ODM structure one has to study the ODM specification and the XML schema that defines ODM 1.3.1 [2].

Display 1 ODM: partial structure



SAS® CLINICAL STANDARDS TOOLKIT AND ODM

SAS® CLINICAL STANDARDS TOOLKIT AND ODM

The SAS Clinical Standards Toolkit 1.5 provides the following support of the CDISC ODM standard (1.3.0 and 1.3.1):

- reading and representing in SAS a complete ODM XML file (specific limitations are noted below)
- building an odm.xml file from a SAS representation of the ODM standard
- schema-level validation of an odm.xml file
- validating the structure and content of the SAS representation of an odm.xml file
- Extraction of ODM Clinical data and ODM Reference data into SAS data sets
- A sample tool to identify unsupported ODM elements and attributes.

The SAS Clinical Standards Toolkit 1.5 will NOT support the following CDISC ODM 1.3.0 and ODM 1.3.1 functionality:

- Reading or writing the DigitalSignatures section of the ODM
- Vendor or customer extensions of the ODM
- Processing is limited to a single ODM file (i.e., use of PriorFileOID to reference another file is ignored).
- Full file metadata is expected in each file.
- Effective support only for ODM FileType="Snapshot". The SAS Clinical Standards Toolkit 1.5 makes no attempt to process multiple transactions per data point; multiple transactions are saved in the SAS ODM representation for subsequent processing

The SAS PROC CDISC procedure supports bi-directional conversion of data content contained in a CDISC ODM XML document to and from data sources that are accessible through SAS. However, PROC CDISC supports CDISC ODM version 1.2 only. PROC CDISC is maintenance only – no new features will be added. The SAS Clinical Standards Toolkit is in active development.

With the production releases of the SAS Clinical Standards Toolkit 1.5, the CDISC ODM reference standard will support reading and representing in SAS a complete odm.xml file, building an odm.xml file, extracting data from an ODM file, and validating the structure and content of the SAS representation of an odm.xml file. In addition, it will validate the structural integrity of the ODM XML file. To support this functionality, supplemental files include the following global standards library files:

- A SAS format catalog (odmfmtcat*.sas7bcat) in the formats folder provides valid values for selected columns in the tables of the SAS representation of ODM.
- The Messages data set in the messages folder provides error messaging for all Validation Master checks.
- The Validation Master data set in the validation/control folder contains the super-set of checks validating the structure and content of the tables that form the SAS representation of the ODM model.
- SAS code in the macros folder provides CDISC ODM-specific code that augments code that is provided in the primary SAS Clinical Standards Toolkit autocall library (!sasroot/cstframework/sasmacro).

It is this set of files, in whole or in part, that defines the CDISC ODM reference standard in SAS.

SAS® CLINICAL STANDARDS TOOLKIT AND ODM – THE SAS DATA MODEL FOR ODM

The highly-structured nature of ODM data requires that any mapping to a relational format includes a large number of data sets, with enforcement of foreign-key relationships to help preserve the intended non-relational object structure. In the SAS Clinical Standards Toolkit, key relationships will be enforced when validating the ODM data sets.

Field lengths in the CDISC ODM data sets are consistent by core data type. CDISC has not specified any limit to the length of most character fields. Arbitrary lengths have been chosen by data type. These lengths are listed in the following table. In the table, standard data types are distilled into core data types. To be safe, larger lengths have been chosen to ensure that no data loss occurs in the SAS Clinical Standards Toolkit pre-installed data sets. Production tables might be compressed using SAS mechanisms to preserve disk space.

Table 1 CDISC ODM Default Lengths by Data Type

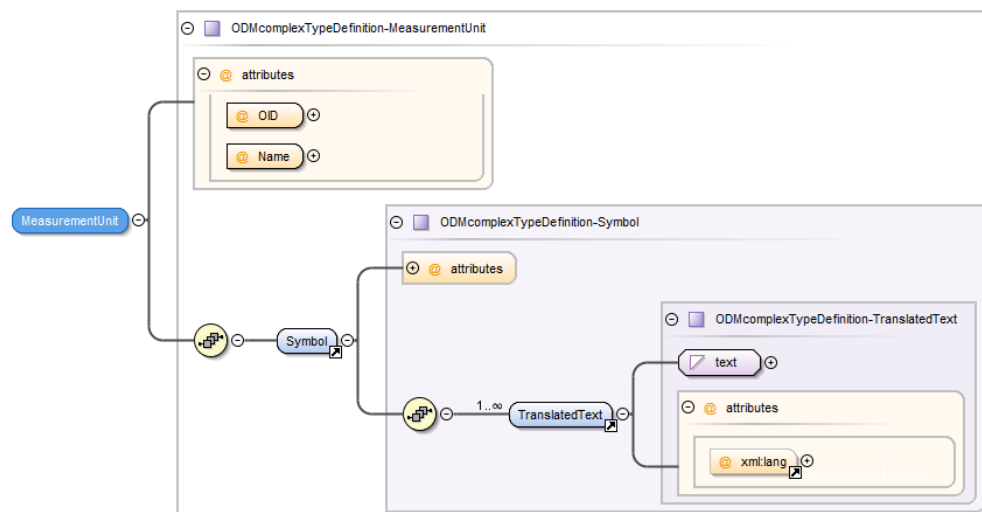
Type Name	Length	Description
oid	64	A unique object identifier or a reference
text	2000	A character field that can accommodate a large number of characters
name	128	A descriptive identifier
value	512	An item of collected or reference data
path	512	An absolute or relative file system path or URL

When mapping the ODM XML schema to SAS data sets, in general, elements become SAS data sets and attributes become columns in the data sets. When there is a 1-n relation between an XML element and its sub-element, the element and its sub-element become different SAS data sets.

The example in Display 2 illustrates this concept. The physical unit of measure for a data item or value has a unique identifier (OID) and a name (Name) and a symbol providing a human-readable name for the measurement unit that is appropriate for a particular language. TranslatedText elements typically occur in a series, presenting a set of alternative textual renditions for different languages. To avoid ambiguity, a particular language tag must not occur more than once in a series of TranslatedText elements. Also, it is not permitted to have more than one TranslatedText element without an xml:lang attribute within the same parent.

The MeasurementUnit element and its attributes and TranslatedText sub-element are mapped to 2 SAS data sets: MeasurementUnits and MuTranslatedText.

Display 2 ODM to SAS data set mapping



```

<BasicDefinitions>
  <MeasurementUnit OID="MeasurementUnits.OID.BEAT_MIN" Name="BEAT_MIN">
    <Symbol>
      <TranslatedText xml:lang="en">beats/minute</TranslatedText>
      <TranslatedText xml:lang="fr-CA">bat/minute</TranslatedText>
    </Symbol>
  </MeasurementUnit>
  <MeasurementUnit OID="MeasurementUnits.OID.CM" Name="CM">
    <Symbol>
      <TranslatedText xml:lang="en">CM</TranslatedText>
      <TranslatedText xml:lang="fr-CA">CM</TranslatedText>
    </Symbol>
  </MeasurementUnit>
  <MeasurementUnit OID="MeasurementUnits.OID.MMHG" Name="MMHG">
    <Symbol>
      <TranslatedText xml:lang="en">mmHG</TranslatedText>
      <TranslatedText xml:lang="fr-CA">mmHG</TranslatedText>
    </Symbol>
  </MeasurementUnit>
  <MeasurementUnit OID="MeasurementUnits.OID.YRS" Name="YEARS">
    <Symbol>
      <TranslatedText xml:lang="de">Jahren</TranslatedText>
      <TranslatedText xml:lang="en">Years of age</TranslatedText>
      <TranslatedText xml:lang="fr-CA">Ans</TranslatedText>
    </Symbol>
  </MeasurementUnit>
</BasicDefinitions>

```

VIEWTABLE: Srcdata.Measurementunits			
	OID	Name	FK_Study
1	MeasurementUnits.OID.BEAT_MIN	BEAT_MIN	Study.OID
2	MeasurementUnits.OID.CM	CM	Study.OID
3	MeasurementUnits.OID.MMHG	MMHG	Study.OID
4	MeasurementUnits.OID.YRS	YEARS	Study.OID

VIEWTABLE: Srcdata.Mutranslatedtext			
	TranslatedText	lang	FK_MeasurementUnits
1	beats/minute	en	MeasurementUnits.OID.BEAT_MIN
2	bat/minute	fr-CA	MeasurementUnits.OID.BEAT_MIN
3	CM	en	MeasurementUnits.OID.CM
4	CM	fr-CA	MeasurementUnits.OID.CM
5	mmHG	en	MeasurementUnits.OID.MMHG
6	mmHG	fr-CA	MeasurementUnits.OID.MMHG
7	Jahren	de	MeasurementUnits.OID.YRS
8	Years of age	en	MeasurementUnits.OID.YRS
9	Ans	fr-CA	MeasurementUnits.OID.YRS

The complete listing of the 76 SAS data sets that comprise the ODM 1.3.1 standard can be found in Appendix 1.

SAS® CLINICAL STANDARDS TOOLKIT AND ODM – PROCESSING XML

The SAS Clinical Standards Toolkit uses an intermediate XML format when transforming an ODM XML file into SAS data sets or transforming SAS data sets into an ODM XML file. This intermediate format is a very 'flat' representation of the table data in XML and will consist of extremely simple markup describing the tables and column data.

This very simple XML 'cube' format captures all tabular data and is sufficient as input to the XML transformation step, which will consume the 'cube' XML file and produce a standards-compliant ODM XML file. Display 3 shows the overall structure of the XML 'cube' document.

Display 3 XML 'cube'

```
<LIBRARY>
  <TableName0>
    <Column0>some data</Column0>
    <Column1>some data</Column1>
    <Column2>some data</Column2>
  </TableName0>
  <TableName1>
    <AColumn0>some data</AColumn0>
    <AColumn1>some data</AColumn1>
  </TableName1>
  <TableName1>
    <AColumn0>some data</AColumn0>
    <AColumn1/>
  </TableName1>
</LIBRARY>
```

We refer to this sort of XML representation of relational table data as a 'cube' XML format, so named because the XML is three levels (or 'dimensions') deep:

- All XML documents must have exactly one root element. In our format, the root element will always be named 'LIBRARY', and will have no additional attributes.
- Each child element of the root will have as its name the name of a data set. Each occurrence of such an element represents a single row/observation in that data set. Therefore, if a data set contains more than one row of data, multiple elements having the same name will exist at this level. Although the order in which distinct data set names appear in this file is unimportant, all elements representing rows in the same data set MUST appear consecutively. Data set name case must be preserved (i.e. data set names cannot be produced in all caps).
- The third, and deepest, level of elements in the XML document represents the rows within the data set represented by the parent of each. The name of each element at this level is the same as the name (case-sensitive) of the SAS data set column. For each observation, exactly one element must be produced for each row in the data set, regardless of whether there is data in that column for the row currently being represented. A cell that contains no data will be represented as an empty XML element, as shown above with the <AColumn1/> element. Otherwise, the element will contain the character data from the corresponding data set cell, exactly as it had been represented in the staging data sets.

Display 4 shows an example of this XML 'cube' representing the MeasurementUnits example as we already saw in Display 2.

Display 4 XML 'cube' example

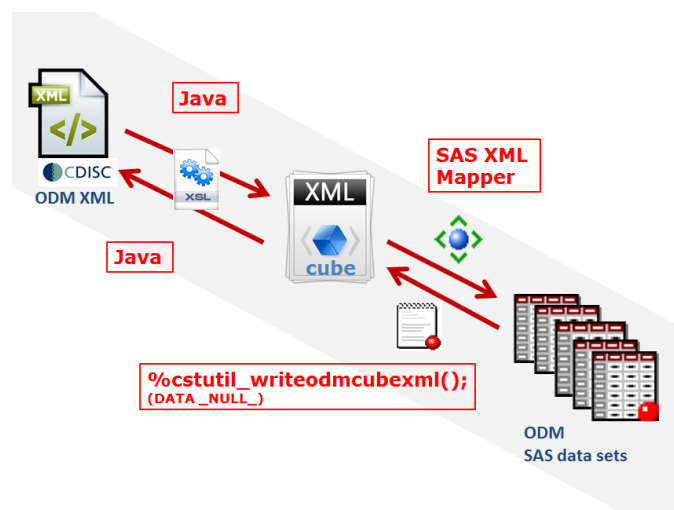
```
<MeasurementUnits>
  <OID>MeasurementUnits.OID.BEAT_MIN</OID>
  <Name>BEAT_MIN</Name>
  <FK_Study>Study.OID</FK_Study>
</MeasurementUnits>
<MeasurementUnits>
  <OID>MeasurementUnits.OID.CM</OID>
  <Name>CM</Name>
  <FK_Study>Study.OID</FK_Study>
</MeasurementUnits>
<MeasurementUnits>
  <OID>MeasurementUnits.OID.MMHG</OID>
  <Name>MMHG</Name>
  <FK_Study>Study.OID</FK_Study>
</MeasurementUnits>
<MeasurementUnits>
  <OID>MeasurementUnits.OID.YRS</OID>
  <Name>YEARS</Name>
  <FK_Study>Study.OID</FK_Study>
</MeasurementUnits>
<MUTranslatedText>
  <TranslatedText>beats/minute</TranslatedText>
  <lang>en</lang>
  <FK_MeasurementUnits>MeasurementUnits.OID.BEAT_MIN</FK_MeasurementUnits>
</MUTranslatedText>
<MUTranslatedText>
  <TranslatedText>bat/minute</TranslatedText>
  <lang>fr-CA</lang>
  <FK_MeasurementUnits>MeasurementUnits.OID.BEAT_MIN</FK_MeasurementUnits>
</MUTranslatedText>
```

A Java process is invoked (through the SAS data step JavaObj component) that leverages XSL technology to transform between the 'cube' XML and the ODM XML file. The SAS Clinical Standards Toolkit uses the Xerces Java parser to validate the ODM XML file and to obtain informational messages about the validation result.

The SAS macro `cstutil_writeodmcubexml` transforms the SAS data sets that represent an ODM XML file to the 'cube' XML file by using a simple `DATA _NULL_` step. By using a `DATA _NULL_` step instead of the SAS XML engine we do not have to make any data type inferences. This `DATA _NULL_` step is driven by the tables and columns metadata that define the ODM standard in the SAS Clinical Standards Toolkit.

The SAS data sets that represent the ODM XML file can be created from the 'cube' XML file by utilizing SAS XML Mapper technology. The XML Map used by the SAS XML Mapper is completely determined by the metadata that describes the SAS representation of the ODM XML standard, i.e. the *reference_tables* and *reference_columns* SAS data sets. Display 5 gives an overview of the various transformations and the technologies used.

Display 5 Transformations between ODM XML and SAS



SAS® CLINICAL STANDARDS TOOLKIT – THE GENERIC PROCESS

The SAS Clinical Standards Toolkit Global Library contains the SAS metadata definition of all supported standards. This metadata definition, defined primarily as the *reference_tables* and *reference_columns* data sets, can be used as defined or may be modified or used as a template to build your own SAS representation of the SAS-supplied CDISC standards or your own customized standard. Sample driver modules are provided with each standard that is supported by the SAS Clinical Standards Toolkit. These driver modules can be copied and modified to reflect your own data and metadata sources and the target location for any SAS process output. These drivers all follow the same general process workflow:

- Set any process global macro variable values.
Global macro variables are set by utilizing a so-called properties file that has name-value pairs.
- Define a root path for input and output files
- Create or reference a *SASReferences* data set that defines all input/output files
- Call the *cstutil_processsetup* macro that confirms a valid *SASReferences* structure, allocates any SAS librefs and filerefs, set the process-specific macro autocall and format search paths
- Call the primary macro that performs the task of interest (such as validation or reading/writing an ODM XML file)
- Optionally perform any session cleanup

Perhaps the single most important file used by the SAS Clinical Standards Toolkit is the *SASReferences* data set. The *SASReferences* data set captures all the input and output file library and file references associated with any given process.

Displays 6a, 6b and 6c give examples of *SASReference* tables for various ODM 1.3.1 related processes.

Display 6a Sample *SASReferences* SAS data set (ODM 1.3.1 Read ODM XML Process)

VIEWTABLE: Work.SasReferences										
standard	type	subtype	SASref	reftype	iotype	filetype	allowoverwrite	path	order	memname
CST-FRAMEWORK	messages		messages	libref	input	dataset	N		1	
CDISC-ODM	messages		odmmsg	libref	input	dataset	N		2	
CDISC-ODM	autocall		odmcode	fileref	input	folder	N		1	
CDISC-ODM	sourcedata		srcdata	libref	output	folder	Y	c:/cstSampleLibrary/cdisc-odm-1.3.1-1.5/derived/data		
CDISC-ODM	sourcecmetadata	table	srcmeta	libref	output	dataset	Y	c:/cstSampleLibrary/cdisc-odm-1.3.1-1.5/derived/metadata		source_tables.sas7bdat
CDISC-ODM	sourcecmetadata	column	srcmeta	libref	output	dataset	Y	c:/cstSampleLibrary/cdisc-odm-1.3.1-1.5/derived/metadata		source_columns.sas7bdat
CDISC-ODM	referencecmetadata	table	refmeta	libref	input	dataset	N			
CDISC-ODM	referencecmetadata	column	refmeta	libref	input	dataset	N			
CDISC-ODM	results	results	results	libref	output	dataset	Y	c:/cstSampleLibrary/cdisc-odm-1.3.1-1.5/results		read_results.sas7bdat
CDISC-ODM	externabxml	xml	odmxml	fileref	input	file	N	c:/cstSampleLibrary/cdisc-odm-1.3.1-1.5/sourcexml	1	odm_sample.xml
CDISC-ODM	referencexml	map	odmmap	fileref	input	file	N	c:/cstSampleLibrary/cdisc-odm-1.3.1-1.5/referencexml	1	odm_map
CDISC-ODM	properties	initialize	inprop	fileref	input	file	N		1	

Display 6b Sample *SASReferences* SAS data set (ODM 1.3.1 Validation Process)

VIEWTABLE: Work.SasReferences										
standard	type	subtype	SASref	reftype	iotype	filetype	allowoverwrite	path	order	memname
CST-FRAMEWORK	messages		messages	libref	input	dataset	N		1	
CST-FRAMEWORK	template		csttmplt	libref	input	folder	N		2	
CDISC-ODM	messages		odmmsg	libref	input	dataset	N		2	
CDISC-ODM	lookup		lookup	libref	input	dataset	N			
CDISC-ODM	template		odmtmpl	libref	input	folder	N		1	
CDISC-ODM	autocall		auto1	fileref	input	folder	N		1	
CDISC-ODM	control	reference	cntrl_s	libref	both	dataset	Y	C:\Users\VRJANS~1\CAR\AppData\Local\Temp\SAS Temporary Files\TD7468_L72371_		sasreferences
CDISC-ODM	control	validation	cntrl_v	libref	input	dataset	N	c:/cstSampleLibrary/cdisc-odm-1.3.1-1.5/control		validation_control
CDISC-ODM	fmtsearch		odmfmt	libref	input	catalog	N		1	
CDISC-ODM	referencecontrol	validation	refcntrl	libref	input	dataset	N			
CDISC-ODM	referencecmetadata	table		libref	input	dataset	N			
CDISC-ODM	referencecmetadata	column		libref	input	dataset	N			
CDISC-ODM	sourcecmetadata	table	srcmeta	libref	input	dataset	N	c:/cstSampleLibrary/cdisc-odm-1.3.1-1.5/metadata		source_tables
CDISC-ODM	sourcecmetadata	column	srcmeta	libref	input	dataset	N	c:/cstSampleLibrary/cdisc-odm-1.3.1-1.5/metadata		source_columns
CDISC-ODM	sourcedata		srcdata	libref	input	folder	N	c:/cstSampleLibrary/cdisc-odm-1.3.1-1.5/data		
CDISC-ODM	results	validationresults	results	libref	output	dataset	Y	c:/cstSampleLibrary/cdisc-odm-1.3.1-1.5/results		validation_results
CDISC-ODM	results	validationmetrics	results	libref	output	dataset	Y	c:/cstSampleLibrary/cdisc-odm-1.3.1-1.5/results		validation_metrics
CDISC-ODM	properties	initialize	inprop	fileref	input	file	N		1	
CDISC-ODM	properties	validation	valprop	fileref	input	file	N		2	

Display 6c Sample SASReferences SAS data set (ODM 1.3.1 Extract domain data from ODM Process)

VIEWTABLE: Work.Sasreferences											
standard	type	subtype	SASref	reftype	iotype	filetype	allowoverwrite	path	order	memname	
CST-FRAMEWORK	messages		messages	libref	input	dataset	N		1		
CDISC-ODM	messages		odmmsg	libref	input	dataset	N		2		
CDISC-ODM	autocall		auto1	fileref	input	folder	N		1		
CDISC-ODM	control	reference	cntl_s	libref	both	dataset	Y	C:\Users\FRJANS~1\CAR\AppData\Local\Temp\SAS Temporary Files_TD7468_L72371_		sasreferences	
CDISC-ODM	fmtsearch		odmfmt	libref	input	catalog	N	c:/cstSampleLibrary/cdisc-odm-1.3.1-1.5/derived/formats	1	odmfmtcat_en	
CDISC-ODM	sourcedata		srcdata	libref	input	folder	N	c:/cstSampleLibrary/cdisc-odm-1.3.1-1.5/derived/data			
CDISC-SDTM	targetdata		trgdata	libref	output	folder	Y	c:/cstSampleLibrary/cdisc-odm-1.3.1-1.5/derived/domains		AE	
CDISC-ODM	results	results	results	libref	output	dataset	Y	c:/cstSampleLibrary/cdisc-odm-1.3.1-1.5/results		extract_results.sas7bdat	
CDISC-ODM	properties	initialize	inprop	fileref	input	file	N		1		

Some general observations about the sample SASReferences data set shown above:

- The type and subtype values are used by the SAS Clinical Standards Toolkit code to find the fileref or libref associated with a particular input or output file type.
- The SASref may be any user-defined value.
- An empty path tells the SAS Clinical Standards Toolkit to look for the default value from the standard-specific metadata in the Global Library (usually a Global Library location).

SAS® CLINICAL STANDARDS TOOLKIT AND ODM – READING ODM

The SAS code for reading an ODM XML file is given below in a shortened form:

```
%let _cstStandard=CDISC-ODM;
%let _cstStandardVersion=1.3.1;
%let _cstVersion=;

%cst_setStandardProperties(_cstStandard=CST-FRAMEWORK,_cstSubType=initialize);

%cst_createdsfromtemplate(_cstStandard=CST-FRAMEWORK,_cstType=control,
                          _cstSubType=reference,_cstOutputDS=work.sasreferences);

proc sql;
  insert into work.sasreferences
  values ("CST-FRAMEWORK" "1.2" "messages" "" "messages"
         "libref" "input" "dataset" "N" "" "" "1" "" "")
  values ("& _cstStandard" "& _cstStandardVersion" "messages" "" "odmmsg"
         "libref" "input" "dataset" "N" "" "" "2" "" "")
  values ("& _cstStandard" "& _cstStandardVersion" "autocall" "" "odmcode"
         "fileref" "input" "folder" "N" "" "" "1" "" "")
  values ("& _cstStandard" "& _cstStandardVersion" "sourcedata" "" "srcdata"
         "libref" "output" "folder" "Y" "" "&studyOutputPath/derived/data"
         . "" "")
  values ("& _cstStandard" "& _cstStandardVersion" "sourcemetadata" "table" "srcmeta"
         "libref" "output" "dataset" "Y" "" "&studyOutputPath/derived/metadata"
         "source_tables.sas7bdat" "")
  values ("& _cstStandard" "& _cstStandardVersion" "sourcemetadata" "column" "srcmeta"
         "libref" "output" "dataset" "Y" "" "&studyOutputPath/derived/metadata"
         "source_columns.sas7bdat" "")
  values ("& _cstStandard" "& _cstStandardVersion" "referencemetadata" "table" "refmeta"
         "libref" "input" "dataset" "N" "" "" . "" "")
  values ("& _cstStandard" "& _cstStandardVersion" "referencemetadata" "column" "refmeta"
         "libref" "input" "dataset" "N" "" "" . "" "")
  values ("& _cstStandard" "& _cstStandardVersion" "results" "results" "results"
         "libref" "output" "dataset" "Y" "" "&studyOutputPath/results"
         "read results.sas7bdat" "")
  values ("& _cstStandard" "& _cstStandardVersion" "externalxml" "xml" "odmxml"
         "fileref" "input" "file" "N" "" "&studyRootPath/sourcexml"
         "odmfile" "")
  values ("& _cstStandard" "& _cstStandardVersion" "referencexml" "map" "odmmap"
         "fileref" "input" "file" "N" "" "&studyRootPath/referencexml"
         "odm.map" "")
  values ("& _cstStandard" "& _cstStandardVersion" "properties" "initialize" "inprop"
         "fileref" "input" "file" "N" "" "" "1" "" "")
;
quit;

%cstutil_processsetup();
%odm_xmlvalidate();
%odm_read();
```


Running this code will result in 76 SAS data sets which contain the SAS representation of the ODM 1.3.1 XML file. The inputs and outputs of this macro are specified via a *SASReferences* file. For the complete code the reader is referred to the sample driver program in the sample study that ships with the SAS Clinical Standards Toolkit. The *odm_read* macro has several optional macro parameters that can be specified.

Some of these parameters:

<code>_cstBuildSrcMetadata</code>	Create the source metadata files (e.g. <code>source_tables</code> and <code>source_columns</code>) as a part of the read operation. Default=Y (yes)
<code>_cstBuildFmtCat</code>	Build format catalog(s), representing language-specific codelist TranslatedText, as a part of the read operation. Default=Y (yes)

SAS® CLINICAL STANDARDS TOOLKIT AND ODM – VALIDATING ODM

What does 'validating' an ODM XML file really mean? We can think of 3 different levels of validation:

- Is the XML file well-formed?
- Is the XML file valid according to an XML schema?
- Does the content of the XML file conform to the specification rules?

We will take a look at these 3 levels of validation.

Well-formed XML

Well-formedness rules specify constraints such as "Every start-tag must have a matching end-tag", and "Attribute values must be quoted". Every XML document, without exception, must be well-formed. This means it must adhere to a number of rules, including the following:

- Proper element nesting is strictly enforced, and every open tag must be matched by a corresponding close tag.
- Elements may nest but may not overlap.
- There must be exactly one root element.
- Attribute values must be quoted.
- An element may not have two attributes with the same name.
- Comments and processing instructions may not appear inside tags.
- No unescaped < or & signs may occur in the character data of an element or attribute.

This is not an exhaustive list. There are many, many ways a document can be malformed [3].

Whether the error is small or large, likely or unlikely, an XML parser reading a document is required to report it. It may or may not report multiple well-formedness errors it detects in the document. However, the parser is not allowed to try to fix the document and make a best-faith effort of providing what it thinks the author really meant. It can't fill in missing quotes around attribute values, insert an omitted end-tag, or ignore the comment that's inside a start-tag. The parser is required to return an error. The SAS Clinical Standards Toolkit will report errors when trying to import an XML document that is not well-formed.

As an example, the following XML fragment is not well-formed, since the open tag (`ItemDef`) is not matched by the close tag (`Itemdef`):

```
<ItemDef OID="ItemDef.OID.DM.SEX" Name="SEX" DataType="string" Length="2"
  SASFieldName="SEX" Comment="Sex of the subject. ">
  <CodeListRef CodeListOID="CodeLists.OID.SEX"/>
</Itemdef>
```

When validating this ODM file with the *odm_xmlvalidate* macro, the SAS Clinical Standards Toolkit reports the following in the Results SAS data set:

VIEWTABLE: Results.Read_results				
	resultid	srcdata	message	resultseverity
32	ODM0003	XML VALIDATION	(Line 443/Column 15) The element type "ItemDef" must be terminated by the matching end-tag "</ItemDef>".	Error
33	ODM0003	XML TRANSFORMER	The element type "ItemDef" must be terminated by the matching end-tag "</ItemDef>".	Error

A subsequent call to the *odm_read* macro would not import this malformed ODM XML file, and the SAS Clinical Standards Toolkit would report in the SAS Log:

```
MPRINT(ODM_READ): * set logging to INFO;
MPRINT(ODM_READ): prefs.callvoidmethod('setLogLevelString','INFO');
MPRINT(ODM_READ): dcl javaobj
transformer("com/sas/ptc/transform/xml/StandardXMLImporter",
prefs);
MPRINT(ODM_READ): transformer.exceptiondescribe(1);
MPRINT(ODM_READ): transformer.callvoidmethod('exec');
MPRINT(ODM_READ): * check the return values here and get results path;
MPRINT(ODM_READ): transformer.delete();
MPRINT(ODM_READ): prefs.delete();
MPRINT(ODM_READ): run;

ERROR: 'The element type "ItemDef" must be terminated by the matching end-tag
"</ItemDef>".'
```

```
ERROR: 'com.sun.org.apache.xml.internal.utils.WrappedRuntimeException: The element type
"ItemDef" must be terminated by the matching end-tag "</ItemDef>".'
```

Validating ODM – XML Schema Validation

XML schema validation of an ODM XML file involves verification that the XML file is valid structurally and syntactically according to the XML schema for the standard describing the ODM file. In our case this is the ODM 1.3.1 XML schema. Every XML file that validates according to an XML schema is well-formed, but not every well-formed XML file is valid.

The SAS Clinical Standards Toolkit includes a call to the *odm_xmlvalidate* macro immediately before the call to the *odm_read* macro in the *create_sasodm_fromxml.sas* sample driver program.

As an example, the following XML fragment is well-formed, but not valid since the element should have been `<ItemDef>` instead of `<Itemdef>` for the Item with Name="SEX":

```
<ItemDef OID="ItemDef.OID.DM.RFSTDTC" Name="RFSTDTC" DataType="string" Length="64"
SASFieldName="RFSTDTC"
Comment="Reference Start Date/time for the subject in ISO 8601 character format.
Usually equivalent to date/time when subject was first exposed to study treatment. Required
for all randomized subjects; will be null for all subjects who did not meet the milestone the
date requires, such as screen failures or unassigned subjects."/>
<Itemdef OID="ItemDef.OID.DM.SEX" Name="SEX" DataType="string" Length="2"
SASFieldName="SEX"
Comment="Sex of the subject.">
<CodeListRef CodeListOID="CodeLists.OID.SEX"/>
</Itemdef>
<ItemDef OID="ItemDef.OID.DM.SITEID" Name="SITEID" DataType="string" Length="40"
SASFieldName="SITEID"
Comment="Unique identifier for a site within a study."/>
```

When validating this ODM file with the *odm_xmlvalidate* macro, the SAS Clinical Standards Toolkit reports the following in the Results SAS data set:

VIEWTABLE: Results.Read_results				
	resultid	srcdata	message	resultseverity
32	ODM0003	XML VALIDATION	(Line 441/Column 136) cvc-complex-type 2.4.a: Invalid content was found starting with element 'Itemdef'. One of '{ http://www.cdisc.org/ns/odm/v1.3 ":ItemDef, http://www.cdisc.org/ns/odm/v1.3 ":CodeList, http://www.cdisc.org/ns/odm/v1.3 ":ImputationMethod, http://www.cdisc.org/ns/odm/v1.3 ":Presentation, http://www.cdisc.org/ns/odm/v1.3 ":ConditionDef, http://www.cdisc.org/ns/odm/v1.3 ":MethodDef}' is expected.	Error
33	ODM0002	XML TRANSFORMER	Document validation failed	Warning
34	ODM0011	ODM_XMLVALIDATE	Errors were reported in the generation of the ODM file.	Error

Since the ODM XML file is well-formed, the *odm_read* macro will still import this ODM file, ignoring the `<Itemdef>` element with Name="SEX".

VIEWTABLE: Srcdata.Itemdefs						
	OID	Name	Data Type	Length	Significant Digits	SASFieldName
94	ItemDef.OID.DM.RACE	RACE	string	40		. RACE
95	ItemDef.OID.DM.RFENDTC	RFENDTC	string	64		. RFENDTC
96	ItemDef.OID.DM.RFSTDTC	RFSTDTC	string	64		. RFSTDTC
97	ItemDef.OID.DM.SITEID	SITEID	string	40		. SITEID
98	ItemDef.OID.DM.STUDYID	STUDYID	string	40		. STUDYID

This example shows that the *odm_read* macro will attempt to import an invalid ODM XML file. However, importing an invalid ODM XML file may result in an incomplete import.

Validating ODM – Validation of the SAS® Data Set Representation

Validation of the SAS Representation of the ODM XML file relies on the definition of a master set of validation checks that are specific to the table and column metadata that define a set of data, and checks that are specific to the data itself. In the SAS Clinical Standards Toolkit, CDISC ODM validation uses the same types of metadata and the same workflow process that is common to validation of all data standards (e.g. SDTM, ADaM and CRT-DDS). SAS provides a set of validation checks for CDISC ODM that are designed to verify the metadata definitions and values of the default 76 data sets that comprise the SAS representation of the ODM 1.3.1 model. These checks were created by SAS and are based on the CDISC ODM specification [2]. Metadata about each check is provided in the Validation Master data set in the folder <global standards library directory>/standards/cdisc-odm-1.3.1-1.5/validation/control. The *odm_validate* macro controls the validation workflow for ODM. As each check is processed from the run-time validation check data set, the check determines the source of the table and column metadata to use. The *reference_tables* and *reference_columns* data sets contain the metadata for the default 76 data sets that comprise the SAS representation for CDISC ODM 1.3.1. Unless you make customizations or run-time modifications, the source metadata *source_tables* and *source_columns* data sets contain the same content as the reference metadata *reference_tables* and *reference_columns* data sets. If all 76 ODM tables contribute information to the ODM XML file, then the validation process can run directly against the reference tables and columns data sets. In this case, the USESOURCEMETADATA data flag in the validation check data set needs to be set to N. However, you can elect to run validation against a subset of the 76 tables. In this case, a *source_tables* data set that contains the subset needs to be created from the *reference_tables* data set. And, a corresponding *source_columns* data set needs to be created from the *reference_columns* data set. The run-time validation check data set can contain all of the checks, and USESOURCEMETADATA can be left set to Y, which is the default value.

Table 2 enumerates the types of checks to be done on ODM data. Each check type is assumed to operate on data that exists in a source column within a source data set. A check type may reference one or more parameters that will be used to validate the source column data. These parameters may be character strings, or may be a representation of some column other than the source column, against which the source column data must be compared. All character comparisons are case sensitive. Character data is assumed to have been ‘trimmed’ of any leading or trailing whitespace.

Table 2 CDISC ODM Validation Checks

Check Type	Category	Description
Unique in data set	Structural	No two values for the source column can be equivalent within the same source data set
Required character value	Data	The 'trimmed' (whitespace removed) value of the character data must consist of one or more characters.
Required numeric value	Data	The numeric value of the column cannot be 'missing'.
Enumeration (s0,s1, ...)	Data	If character data exists, its value must match one of the given enumerated character strings. All string comparisons are case-sensitive.
Foreign key (targetColumn)	Structural	Each existing value in this column must have an equivalent value in the given target column.
Foreign key required (targetColumn)	Structural	A value is required for this column in every row, and each value must have an equivalent value in the given target column. This check is the equivalent of running the 'Required character value' check, and failing if that check fails. If 'required character value' passes, the 'foreign key()' check is run.
Character format: language	Data	The character data must consist of 1-8 alphabetical characters of either case, followed optionally by a hyphen (-) character and any sequence of 1-8 alphabetical characters of either case, or numeric digits, after that hyphen. For example, 'e' is a legal value, as is 'en-us' and 'english' and 'english-d842'. Illegal values include '1en', 'mumblespeak' and 'en_us'. The hyphen-character sequence can be repeated any number of times, also making a value such as 'english-mumbly-growly-47' a legal value. Regular expression: "[a-zA-Z]{1,8}(-[a-zA-Z0-9]{1,8})?"
Character format: fileName	Data	The character data must not contain any characters other than upper- and lower-case letters of the alphabet, numeric digits, the underscore (_) character or the period (.) character. Regular expression: [A-Za-z0-9_.]+
Character format: sasFormat	Data	The first character must be either a lower- or upper-case letter, an underscore (_), or the dollar sign (\$). Any subsequent character must be either an upper- or lowercase letter, a numeric digit, the underscore (_), or a period (.). Regular expression: [A-Za-z_\$][A-Za-z0-9_.]*
Character format: sasName	Data	The first character must be either a lower- or upper-case letter, or an underscore (_). Any subsequent character must be either an upper- or lowercase letter, a numeric digit, or the underscore (_). Regular expression: [A-Za-z_][A-Za-z0-9_]*
Unique across data sets (targetcolumn0, ...)	Structural	No value in this column can be equal to any value in any of the given data set column(s)
Primary key	Data	Must satisfy the 'Unique in data set' check type as well as the 'Required character value' check type.
Must Have Corresponding Value (targetColumn)	Structural	For each distinct value in this column, there must be at least one equivalent value in the supplied target column.
No Duplicates Per Unique Value (targetColumn)	Structural	For each distinct value in the target column, each value in the source column must be unique. That is, the same value cannot appear more than once in the source column for each distinct value in the target column.

Let's have a look at an example.

```

<ItemGroupDef OID="ItemGroupDefs.OID.DM" Name="DM" Repeating="No"
    IsReferenceData="N" SASDatasetName="DM" Domain="DM"
    Purpose="Tabulation">
    <ItemRef ItemOID="ItemDef.DM.AGEU" Mandatory="No" OrderNumber="11" Role="RecordQualifier"/>
    <ItemRef ItemOID="ItemDef.DM.AGEU" Mandatory="No" OrderNumber="12" Role="VariableQualifier"/>
...
</ItemGroupDef>
...
<ItemDef OID="ItemDef.DM.AGEU" Name="AGEU" DataType="string" Length="10"
    SASFieldName="AGEU" Comment="Units associated with AGE. ">
    <CodeListRef CodeListOID="CodeLists.OID.AGEU"/>
</ItemDef>

<CodeList OID="CodeLists.OID.AGEUNIT" Name="AGEU" DataType="string"
    SASFormatName="AGEU">
    <EnumeratedItem CodedValue="YEARS" />
</CodeList>

```

Although this ODM file is valid according to the XML schema, there are several issues:

- There is a duplicate "ItemDef.OID.DM.AGEU" ItemRef element
- The CodeList referenced (CodeLists.OID.AGU) does not exist
- The SASFormatName should start with a "\$" character

When running the *odm_validate* macro on this ODM file, several issues are reported in the Results SAS data set:

VIEWTABLE: Results.Validation_results								
	resultid	checkid	srcdata	message	resultseverity	resultflag	_cst_rc	actual
41	ODM0100	ODM0100	SRCDATA.ITEMGROUPDEFITEMREFS	No two values for the source column can be equivalent within the same source data set	Error	1	0	keys=FK_ITEMGROUPDEFS ITEMROID
201	ODM0110	ODM0110	SRCDATA.ITEMDEFS (SRCDATA.CODELISTS)	The foreign key OID does not have a corresponding value in the target data set SRCDATA.ITEMDEFS	Error	1	0	CODELISTREF=CodeLists.OID.AGEU
323	ODM0217	ODM0217	SRCDATA.CODELISTS	A CodeList contains a SASFormatName value that does not have the correct format: CodeList datatype is text or string and the FormatName does not start with \$.	Error	1	0	DATATYPE=string,SASFORMATNAME=AGEU

SAS® CLINICAL STANDARDS TOOLKIT AND ODM – CLINICAL DATA AND REFERENCE DATA IN ODM

Starting in ODM 1.3.0, there are two forms of the ItemData elements -- the element used by the ODM for transmitting clinical data item values and reference data item values. These are the untyped and typed forms.

An example of a typed ItemData element and an untyped ItemData element:

```
<ItemDataFloat ItemOID="ID.VS.VSSTRESN">76.1</ItemDataFloat>
<ItemData ItemOID="ID.AE.AETERM" Value="HEADACHE" />
```

Both of these data values will be stored in the Value variable in the ItemData SAS data set. In the case of typed data the ItemDataType variable in the ItemData SAS data set will have the data type (e.g. Float). In the case of untyped data the ItemDataType variable in the ItemData SAS data set will be null.

Normally, Typed and untyped data transmission should not be mixed within a single ODM file. However, in the example provided by the Toolkit both transmission types are part of the same example for demonstration purposes.

SAS® CLINICAL STANDARDS TOOLKIT AND ODM – EXTRACTING CLINICAL DATA AND REFERENCE DATA IN ODM

The *odm_extractdomaindata* macro supports the extraction of clinical data or reference data from the SAS data sets that were created by the *odm_read* macro.

The macro has the following assumptions:

- An ODM XML file is available that contains sufficient metadata and content for extractable clinical and reference data
- A full SAS representation of an ODM file is available (e.g. the *odm_read* macro has been run against the XML file).
- The SAS representation of an ODM file contains both metadata and data. By default, the driver assumes all source data files reside in the sample derived/data folder typically populated running the *odm_read* macro. However, the source data files and the source metadata files can be in different folders.
- Any codelists defined in the XML file and associated with extracted data set columns are available as a part of the *odm_read* output

ODM integer and float data types are converted to SAS numeric data; all other ODM data types are converted to SAS character data. In case an integer or float value cannot be converted, there will be a warning in the LOG and Results data set.

Here is a partial listing of the metadata in a sample ODM XML file:

```
<ItemGroupDef OID="ItemGroupDefs.OID.AE" Repeating="Yes"
  SASDatasetName="AE" Name="Adverse Events" Domain="AE"
  Comment="Some adverse events from this trial">
  <ItemRef ItemOID="ID.TAREA" OrderNumber="1" Mandatory="No" />
  <ItemRef ItemOID="ID.PNO" OrderNumber="2" Mandatory="No" />
  <ItemRef ItemOID="ID.SCTRY" OrderNumber="3" Mandatory="No" />
  <ItemRef ItemOID="ID.F_STATUS" OrderNumber="4" Mandatory="No" />
  <ItemRef ItemOID="ID.LINE_NO" OrderNumber="5" Mandatory="No" />
  <ItemRef ItemOID="ID.AETERM" OrderNumber="6" Mandatory="No" />
  <ItemRef ItemOID="ID.AESTMON" OrderNumber="7" Mandatory="No" />
  <ItemRef ItemOID="ID.AESTDAY" OrderNumber="8" Mandatory="No" />
  <ItemRef ItemOID="ID.AESTYR" OrderNumber="9" Mandatory="No" />
```

```

<ItemRef ItemOID="ID.AESTDT"      OrderNumber="10" Mandatory="No" />
<ItemRef ItemOID="ID.AEENMON"    OrderNumber="11" Mandatory="No" />
<ItemRef ItemOID="ID.AEENDAY"    OrderNumber="12" Mandatory="No" />
<ItemRef ItemOID="ID.AEENYR"     OrderNumber="13" Mandatory="No" />
<ItemRef ItemOID="ID.AEENDT"     OrderNumber="14" Mandatory="No" />
<ItemRef ItemOID="ID.AESEV"      OrderNumber="15" Mandatory="No" />
<ItemRef ItemOID="ID.AEREL"      OrderNumber="16" Mandatory="No" />
<ItemRef ItemOID="ID.AEOUT"      OrderNumber="17" Mandatory="No" />
<ItemRef ItemOID="ID.AEACTTRT"   OrderNumber="18" Mandatory="No" />
<ItemRef ItemOID="ID.AECONTRT"   OrderNumber="19" Mandatory="No" />
</ItemGroupDef>

<ItemDef OID="ID.AESTDT" SASFieldName="AESTDT"
  Name="Derived Start Date" DataType="date"/>
<ItemDef OID="ID.AEENMON" SASFieldName="AEENMON"
  Name="Stop Month - Enter Two Digits 01-12" DataType="integer" Length="2" />
<ItemDef OID="ID.AEENDAY" SASFieldName="AEENDAY"
  Name="Stop Day - Enter Two Digits 01-31" DataType="integer" Length="2" />
<ItemDef OID="ID.AEENYR" SASFieldName="AEENYR"
  Name="Stop Year - Enter Four Digit Year" DataType="integer" Length="4" />
<ItemDef OID="ID.AEENDT" SASFieldName="AEENDT"
  Name="Derived Stop Date" DataType="date"/>
<ItemDef OID="ID.AESEV" SASFieldName="AESEV"
  Name="Severity" DataType="text" Length="1">
  <CodeListRef CodeListOID="CL.$AESEV" />
</ItemDef>
<ItemDef OID="ID.AEREL" SASFieldName="AEREL"
  Name="Relationship to study drug" DataType="text" Length="1">
  <CodeListRef CodeListOID="CL.$AEREL" />
</ItemDef>

```

Here is a partial listing of the data from the same sample ODM file:

```

<ClinicalData StudyOID="Study.OID" MetaDataVersionOID="MetaDataVersion.OID.1">
  <SubjectData SubjectKey="S001P011" TransactionType="Insert">
    <StudyEventData StudyEventOID="StudyEventDefs.OID.6.AdverseEvent"
      StudyEventRepeatKey="1">
      <FormData FormOID="FormDefs.OID.AE" FormRepeatKey="1">
        <ItemGroupData ItemGroupOID="ItemGroupDefs.OID.AE"
          ItemGroupRepeatKey="1">
          <ItemData ItemOID="ID.TAREA" Value="ONC" />
          <ItemData ItemOID="ID.PNO" Value="143-02" />
          <ItemData ItemOID="ID.SCTRY" Value="USA" />
          <ItemData ItemOID="ID.F STATUS" Value="V" />
          <ItemData ItemOID="ID.LINE_NO" Value="1" />
          <ItemData ItemOID="ID.AETERM" Value="HEADACHE" />
          <ItemData ItemOID="ID.AESTMON" Value="06" />
          <ItemData ItemOID="ID.AESTDAY" Value="10" />
          <ItemData ItemOID="ID.AESTYR" Value="1999" />
          <ItemData ItemOID="ID.AESTDT" Value="1999-06-10" />
          <ItemData ItemOID="ID.AEENMON" Value="06" />
          <ItemData ItemOID="ID.AEENDAY" Value="14" />
          <ItemData ItemOID="ID.AEENYR" Value="1999" />
          <ItemData ItemOID="ID.AEENDT" Value="1999-06-14" />
          <ItemData ItemOID="ID.AESEV" Value="1" />
          <ItemData ItemOID="ID.AEREL" Value="0" />
          <ItemData ItemOID="ID.AEOUT" Value="1" />
          <ItemData ItemOID="ID.AEACTTRT" Value="0" />
          <ItemData ItemOID="ID.AECONTRT" Value="1" />
        </ItemGroupData>
      </FormData>
    </StudyEventData>
  </SubjectData>
</ClinicalData>

```

The `odm_extractdomaindata` macro will create the data set as depicted in displays 7 and 8. The first 12 columns in this data set, whose name starts with two underscores, are the so-called data set keys. The macro parameter `_cstODMMinimumKeyset` determines whether these keys will be part of the extracted data set.

Display 7 The AE SAS data set (unformatted) as created by the *odm_extractdomaindata* macro

Obs	__StudyOID	__MetaDataVersionOID	__SubjectKey	__StudyEventOID	__StudyEventRepeatKey	__FormOID
1	Study.OID	MetaDataVersion.OID.1	S001P011	StudyEventDefs.OID.6.AdverseEvent	1	FormDefs.OID.AE
2	Study.OID	MetaDataVersion.OID.1	S001P011	StudyEventDefs.OID.6.AdverseEvent	1	FormDefs.OID.AE

__FormRepeatKey	__ItemGroupOID	__ItemGroupRepeatKey	__TransactionType	__LocationOID	__UserOID
1	ItemGroupDefs.OID.AE	1	Insert	Location.OID.S001	User.OID.I008
1	ItemGroupDefs.OID.AE	2	Insert	Location.OID.S001	User.OID.I008

TAREA	PNO	SCTRY	F_STATUS	LINE_NO	AETERM	AESTMON	AESTDAY	AESTYR	AESTDT
ONC	143-02	USA	V	1	HEADACHE	6	10	1999	1999-06-10
ONC	143-02	USA	V	2	CONGESTION	6	11	1999	1999-06-11

AEENMON	AEENDAY	AEENYR	AEENDT	AESEV	AEREL	AEOUT	AEACTTRT	AECONTRT
6	14	1999	1999-06-14	1	0	1	0	1
.	.	.	1999	1	0	2	0	1

Display 8 The AE SAS data set (formatted) as created by the *odm_extractdomaindata* macro

Obs	__StudyOID	__MetaDataVersionOID	__SubjectKey	__StudyEventOID	__StudyEventRepeatKey	__FormOID
1	Study.OID	MetaDataVersion.OID.1	S001P011	StudyEventDefs.OID.6.AdverseEvent	1	FormDefs.OID.AE
2	Study.OID	MetaDataVersion.OID.1	S001P011	StudyEventDefs.OID.6.AdverseEvent	1	FormDefs.OID.AE

__FormRepeatKey	__ItemGroupOID	__ItemGroupRepeatKey	__TransactionType	__LocationOID	__UserOID
1	ItemGroupDefs.OID.AE	1	Insert	Location.OID.S001	User.OID.I008
1	ItemGroupDefs.OID.AE	2	Insert	Location.OID.S001	User.OID.I008

TAREA	PNO	SCTRY	F_STATUS	LINE_NO	AETERM	AESTMON	AESTDAY	AESTYR	AESTDT
Oncology	143-02	United States	Source verified, queried	1	HEADACHE	6	10	1999	1999-06-10
Oncology	143-02	United States	Source verified, queried	2	CONGESTION	6	11	1999	1999-06-11

AEENMON	AEENDAY	AEENYR	AEENDT	AESEV	AEREL	AEOUT	AEACTTRT	AECONTRT
6	14	1999	1999-06-14	Mild	None	Resolved, no residual effects	None	Medication required
.	.	.	1999	Mild	None	Continuing	None	Medication required

The *odm_extractdomaindata* macro has the following signature:

```
%macro odm_extractdomaindata (
    cstSourceMetadata=,
    cstSourceData=,
    cstIsReferenceData=No,
    cstSelectAttribute=Name,
    cstSelectAttributeValue=,
    cstLang=en,
    cstMaxLabelLength=256,
    cstAttachFormats=Yes,
    cstODMMinimumKeyset=No,
    cstOutputLibrary=,
    cstOutputDS=
);
```


The **_cstSourceMetadata** and **_cstSourceData** parameters contain the SAS libref for the SAS ODM metadata representation data SAS ODM representation. If not specified, the code will look first for type=sourcedata from SASReferences. If this is not provided, data set(s) source will be assumed to be in the WORK SAS library.

The **_cstIsReferenceData** parameter tells whether the data to be extracted is reference data (e.g. laboratory reference ranges, trial design data) or clinical data.

The **_cstSelectAttribute** contains the ItemGroup attribute that identifies which ItemGroup will be extracted. Valid values are OID, Name, SASDatasetName or Domain.

The **_cstSelectAttributeValue** parameter contains the value of the attribute as defined by the **_cstSelectAttribute** parameter that identifies the ItemGroup to be extracted.

The **_cstLang** parameter specifies a language identifier for the language tag attribute (xml:lang) in the ODM TranslatedText elements.

The **_cstMaxLabelLength** parameter determines the maximum value of labels to be created. If not provided, 256 is assumed.

Formats will be attached to the data set variables in case the parameter **_cstAttachFormats** has a value of 'Yes'.

The parameter **_cstODMMinimumKeyset** determines the creation of data set keys. If not provided, No is assumed.

The **_cstOutputLibrary** parameter defines the SAS library where where extracted data set(s) are to be written. If not specified, the code will look first for type=targetdata from SASReferences. If this is not provided, data set(s) will be written to WORK. The **_cstOutputDS** parameter contains the name of the extracted data set. If this is an invalid SAS data set name, an error will be generated. If the name is not provided, the code will look first for type=targetdata from SASReferences.

Two sample driver modules for each ODM version (1.3.0 and 1.3.1) are provided by SAS to demonstrate the use of the *odm_extractdomaindata* macro:

<sample study library directory>/cdisc-odm-1.3.0-1.5/programs/extract_domaindata_all.sas
<sample study library directory>/cdisc-odm-1.3.0-1.5/programs/extract_domaindata.sas

<sample study library directory>/cdisc-odm-1.3.1-1.5/programs/extract_domaindata_all.sas
<sample study library directory>/cdisc-odm-1.3.1-1.5/programs/extract_domaindata.sas

The *extract_domain_all.sas* sample driver modules demonstrate how all data sets can be extracted at once. The following shows a code fragment:

```
filename incCode CATALOG "work._cstCode.domains.source" lrecl=255;

data _null_;
  set srcdata.itemgroupdefs(keep=OID Name IsReferenceData SASDatasetName Domain);
  file incCode;
  length macrocall $400 _cstOutputName $100;

  _cstOutputName=SASDatasetName;
  * If we have to use the Name, Only use letters and digits;
  if missing( _cstOutputName) then _cstOutputName=cats(compress(Name, 'adk'));
  * If first character a digit, prepend an underscore;
  if anydigit(_cstOutputName)=1 then _cstOutputName=cats('_', _cstOutputName);
  * Cut long names;
  if length(_cstOutputName) > 32 then _cstOutputName=substr(_cstOutputName, 1, 32);

  macrocall=cats('%odm_extractdomaindata(_cstSelectAttribute=OID',
    ', _cstSelectAttributeValue=', OID,
    ', _cstIsReferenceData=', IsReferenceData,
    ', _cstMaxLabelLength=256',
    ', _cstAttachFormats=Yes',
    ', _cstODMMinimumKeyset=No',
    ', _cstLang=en',
    ', _cstOutputDS=', _cstOutputName, ');');

  put macrocall;
run;

%include incCode;
filename incCode clear;
```


SAS® CLINICAL STANDARDS TOOLKIT AND ODM – IMPORTING CDISC ODM CONTROLLED TERMINOLOGY XML FILES

In order to read an ODM XML Controlled Terminology file as published quarterly by NCI [4], a specialized macro named *ct_read* is available in SAS Clinical Standards Toolkit. A sample driver program that uses this macro is located at:

<sample study library directory>/cdisc-ct-1.0-1.5/programs

The *ct_read* macro will convert an ODM XML Controlled Terminology file into the SAS representation of this standard as shown in Display 9.

Display 9 Data sets created by the *ct_read* macro

c:\cstsamplelibrary\cdisc-ct-1.0-1.5\data\sdtm\201212

Library

Name	Obs	Vars
CLITEMDECODETRANSLATEDTEXT	0	3
CODELISTALIASES	0	3
CODELISTITEMALIASES	0	3
CODELISTITEMS	0	7
CODELISTITEMSYNONYM	0	2
CODELISTS	136	9
CODELISTSYNONYM	136	2
CODELISTTRANSLATEDTEXT	136	3
ENUMERATEDITEMALIASES	0	3
ENUMERATEDITEMS	6902	7
ENUMERATEDITEMSYNONYM	7277	2
EXTERNALCODELISTS	0	5
METADATAVERSION	1	6
ODM	1	13
STUDY	1	5

LibraryCODELISTS

FreezeHideShow...FormatFilter...Font...Find

TableView

	OID	Name	Data Type	ExtCodeID	CodeListExtensible	CDISCSubmissionV	PreferredTerm
1	CLC66767.ACN	Action Taken with Stu...	text	C66767	No	ACN	CDISC SDTM Ac...
2	CLC66768.OUT	Outcome of Event	text	C66768	No	OUT	CDISC SDTM Ad...
3	CLC66780.AGESPAN	Age Span	text	C66780	Yes	AGESPAN	CDISC SDTM Ag...
4	CLC66781.AGEU	Age Unit	text	C66781	No	AGEU	CDISC SDTM Ag...
5	CLC66797.IECAT	Category for Inclusion...	text	C66797	No	IECAT	CDISC SDTM Ca...
6	CLC66784.TOXGRV3	Common Terminology ...	text	C66784	No	TOXGRV3	CDISC SDTM Co...
7	CLC87162.TOXGRV4	Common Terminology ...	text	C87162	No	TOXGRV4	CDISC SDTM Co...
8	CLC66785.TCNTRL	Control Type	text	C66785	Yes	TCNTRL	CDISC SDTM Co...
9	CLC66786.COUNTRY	Country	text	C66786	No	COUNTRY	CDISC SDTM Co...

LibraryENUMERATEDITEMS

FreezeHideShow...FormatFilter...Font...Find

TableView

	OID	CodedValue	ExtCodeID	CDISCDefinition	PreferredTerm	FK_CodeLists
1	N65587	DOSE INCREASED	C49503	An indication that...	Dose Increased	CLC66767.ACN
2	N65598	DOSE NOT CHANGED	C49504	An indication that...	Dose Not Changed	CLC66767.ACN
3	N65609	DOSE REDUCED	C49505	An indication that...	Dose Reduced	CLC66767.ACN
4	N65620	DRUG INTERRUPTED	C49501	An indication that...	Drug Interrupted	CLC66767.ACN
5	N65631	DRUG WITHDRAWN	C49502	An indication that...	Drug Withdrawn	CLC66767.ACN
6	N65642	NOT APPLICABLE	C48660	Determination of ...	Not Applicable	CLC66767.ACN
7	N65659	UNKNOWN	C17998	Not known, not o...	Unknown	CLC66767.ACN
8	N65700	FATAL	C48275	The termination o...	Death Related to Adv...	CLC66768.OUT
9	N65720	NOT RECOVERED/N...	C49494	One of the possib...	Not Recovered or Not...	CLC66768.OUT
10	N65731	RECOVERED/RESOL...	C49498	One of the possib...	Recovered or Resolved	CLC66768.OUT
11	N65742	RECOVERED/RESOL...	C49495	One of the possib...	Recovered or Resolved	CLC66768.OUT

To be able to utilize the NCI CDISC Controlled Terminology in a SAS Clinical Standards Toolkit process, the SAS data sets that were created by the *ct_read* macro need to be converted to a SAS format catalog. Also, to enable SAS Clinical Data Integration to import Controlled Terminology the SAS data set representation, as created by the *ct_read* macro, needs to be combined into one SAS data set.

The following display shows the Action Taken with Study Treatment CodeList as an example of Controlled Terminology in ODM XML.

```

<CodeList OID="CL.C66767.ACN" Name="Action Taken with Study Treatment"
  DataType="text" nciodm:ExtCodeID="C66767" nciodm:CodeListExtensible="No">
  <Description>
    <TranslatedText xml:lang="en">Action Taken with Study Treatment</TranslatedText>
  </Description>
  <EnumeratedItem CodedValue="DOSE INCREASED" nciodm:ExtCodeID="C49503">
    <nciodm:CDISCDefinition>An indication that a medication schedule was modified by addition;
      either by changing the frequency, strength or amount. (NCI)</nciodm:CDISCDefinition>
    <nciodm:PreferredTerm>Dose Increased</nciodm:PreferredTerm>
  </EnumeratedItem>
  <EnumeratedItem CodedValue="DOSE NOT CHANGED" nciodm:ExtCodeID="C49504"> [3 lines]
  <EnumeratedItem CodedValue="DOSE REDUCED" nciodm:ExtCodeID="C49505"> [3 lines]
  <EnumeratedItem CodedValue="DRUG INTERRUPTED" nciodm:ExtCodeID="C49501"> [3 lines]
  <EnumeratedItem CodedValue="DRUG WITHDRAWN" nciodm:ExtCodeID="C49502"> [3 lines]
  <EnumeratedItem CodedValue="NOT APPLICABLE" nciodm:ExtCodeID="C48660"> [5 lines]
  <EnumeratedItem CodedValue="UNKNOWN" nciodm:ExtCodeID="C17998">
    <nciodm:CDISCSynonym>U</nciodm:CDISCSynonym>
    <nciodm:CDISCSynonym>Unknown</nciodm:CDISCSynonym>
    <nciodm:CDISCDefinition>Not known, not observed, not recorded, or refused. (NCI)</nciodm:CDISCDefinition>
    <nciodm:PreferredTerm>Unknown</nciodm:PreferredTerm>
  </EnumeratedItem>
  <nciodm:CDISCSubmissionValue>ACN</nciodm:CDISCSubmissionValue>
  <nciodm:CDISCSynonym>Action Taken with Study Treatment</nciodm:CDISCSynonym>
  <nciodm:PreferredTerm>CDISC SDTM Action Taken with Study Treatment Terminology</nciodm:PreferredTerm>
</CodeList>

```

The *ct_createformats* macro will create the data set as (partially) depicted in display 10. The macro will use this data set to create the SAS format as depicted in display 11.

Display 10 The CTERMS data set as created by the *ct_createformats* macro

VIEWTABLE: Trgdata.Cterms ("CDISC SDTM Controlled Terminology, 2012-12-21")											
	description	codelist	codelist_code	codelist_name	codelist_ext	datatype	type	fmtname	code	cdisc_submission_value	cdisc_synonym
1	CDISC SDTM Controlled Terminology, 2012-12-21	ACN	C66767	Action Taken with Study Treatment	No	text	C	ACN	C49503	DOSE INCREASED	
2	CDISC SDTM Controlled Terminology, 2012-12-21	ACN	C66767	Action Taken with Study Treatment	No	text	C	ACN	C49504	DOSE NOT CHANGED	
3	CDISC SDTM Controlled Terminology, 2012-12-21	ACN	C66767	Action Taken with Study Treatment	No	text	C	ACN	C49505	DOSE REDUCED	
4	CDISC SDTM Controlled Terminology, 2012-12-21	ACN	C66767	Action Taken with Study Treatment	No	text	C	ACN	C49501	DRUG INTERRUPTED	
5	CDISC SDTM Controlled Terminology, 2012-12-21	ACN	C66767	Action Taken with Study Treatment	No	text	C	ACN	C49502	DRUG WITHDRAWN	
6	CDISC SDTM Controlled Terminology, 2012-12-21	ACN	C66767	Action Taken with Study Treatment	No	text	C	ACN	C48660	NOT APPLICABLE	NA; Not Applicable
7	CDISC SDTM Controlled Terminology, 2012-12-21	ACN	C66767	Action Taken with Study Treatment	No	text	C	ACN	C17998	UNKNOWN	U; Unknown

Display 11 The \$ACN SAS format as created by the *ct_createformats* macro

#	Name	Type	Description
1	ACN	FORMATC	Action Taken with Study Treatment (ACN - C66767)

FORMAT NAME: \$ACN LENGTH: 16 NUMBER OF VALUES: 7 MIN LENGTH: 1 MAX LENGTH: 40 DEFAULT LENGTH: 16 FUZZ: 0		
START	END	LABEL (VER. V7 V8 11JAN2013:17:17:33)
DOSE INCREASED	DOSE INCREASED	DOSE INCREASED
DOSE NOT CHANGED	DOSE NOT CHANGED	DOSE NOT CHANGED
DOSE REDUCED	DOSE REDUCED	DOSE REDUCED
DRUG INTERRUPTED	DRUG INTERRUPTED	DRUG INTERRUPTED
DRUG WITHDRAWN	DRUG WITHDRAWN	DRUG WITHDRAWN
NOT APPLICABLE	NOT APPLICABLE	NOT APPLICABLE
UNKNOWN	UNKNOWN	UNKNOWN

The `ct_createformats` macro has the following signature:

```
%macro ct_createformats(
  cstLang=en,          /* Language tag in TranslatedText to use      */
  _cstCreateCatalog=1, /* Create format catalog                          */
  _cstKillCatFirst=0,  /* Empty catalog first                             */
  _cstUseExpression=,  /* Expression to create the SAS format name       */
  _cstAppendChar=F,    /* Letter to append in case SAS format name       */
                      /* ends with digit                                */
  _cstDeleteEmptyColumns=1, /* Delete columns in output data set that are    */
                      /* completely missing                             */
  _cstTrimCharacterData=1 /* Truncate character data in output data set    */
                      /* to the minimum value needed.                  */
);
```

An effort is made by the `ct_createformats` macro to map the `CodeList/nciodm:CDISCSubmissionValue` in the **codelist** variable to the **fmtname** variable that has to contain a valid SAS format name.

The `ct_createformats` macro uses the following steps to create a valid SAS format name:

- Apply a user defined expression to create the **fmtname** variable
- In case **fmtname** is still empty, use the `CodeList/SASFormatName` attribute (typically empty in NCI EVS ODM XML files)
- In case **fmtname** is still empty, use the `CodeList/nciodm:CDISCSubmissionValue` in the **codelist** variable.
- In case the **fmtname** ends with a digit, add the character as specifies by the `_cstAppendChar` macro parameter (default=F)

After the application of these steps the contents of the **fmtname** variable will be validated against the following regular expression:

```
'm/^(?={1,32}$)([a-zA-Z_][a-zA-Z0-9_]*[a-zA-Z_])$/'
```

In case the contents of the **fmtname** variable does not validate against this regular expression the **fmtname** variable does not contain a valid SAS formats name and the content will be set to missing. The **codelist** will then not be used in to create a SAS format.

Two sample driver modules are provided by SAS to demonstrate the use of the `ct_createformats` macro:

```
<sample study library directory>/cdisc-ct-1.0-1.5/programs/create_ctformats.sas
<sample study library directory>/cdisc-ct-1.0-1.5/programs/create_ctformats_qs.sas
```

Both of these sample driver modules demonstrate how the CDISC Submission Value can be mapped to a valid SAS format name.

```
%let _cstCTCat=;
%let _cstCTLibrary=;
%let _cstSASrefLib=;
%let _cstTrgDataLibrary=;
%let _cstCTData=;

%let _cstStandard=CDISC-CT;
%let _cstStandardVersion=1.0.0;

%cst_setStandardProperties(_cstStandard=CST-FRAMEWORK,_cstSubType=initialize);
%cst_getRegisteredStandards(_cstOutputDS=work._cstStandards);

%cst_createdsfromtemplate(_cstStandard=CST-FRAMEWORK,
  _cstType=control,_cstSubType=reference,_cstOutputDS=work.sasreferences);

proc sql;
  insert into work.sasreferences
  values ("CST-FRAMEWORK" "1.2" "messages" "" "messages"
    "libref" "input" "dataset" "N" "" "" 1 "" "")
  values ("% cstStandard" "& cstStandardVersion" "messages" "" "odmmmsg"
    "libref" "input" "dataset" "N" "" "" 2 "" "")
  values ("% cstStandard" "& cstStandardVersion" "autocall" "" "autol"
    "fileref" "input" "folder" "N" "" "" 1 "" "")
  values ("% cstStandard" "& cstStandardVersion" "control" "reference" "control"
    "libref" "both" "dataset" "N" "" "&workpath" . "sasreferences" "")
  values ("% cstStandard" "& cstStandardVersion" "results" "results" "results"
    "libref" "output" "dataset" "Y" "" "&studyOutputPath/results" .
    "cterms_results_&_cstCTStandard._&_cstCTDate..sas7bdat" "");
```

```

values ("%_cstStandard" "%_cstStandardVersion" "sourcedata" "" "srcdata"
       "libref" "input" "folder" "N" ""
       "&studyRootPath/data/&_cstCTStandard/&_cstCTDate" . "" "")
values ("%_cstStandard" "%_cstStandardVersion" "fmtsearch" "" "cstfmt"
       "libref" "output" "catalog" "N" ""
       "&studyOutputPath/data/&_cstCTStandard/&_cstCTDate/formats" 1 "cterms.sas7bcat" "")
values ("%_cstStandard" "%_cstStandardVersion" "targetdata" "" "trgdata"
       "libref" "output" "dataset" "Y" ""
       "&studyOutputPath/data/&_cstCTStandard/&_cstCTDate/formats" . "cterms.sas7bdat" "")
values ("%_cstStandard" "%_cstStandardVersion" "properties" "initialize" "inprop"
       "fileref" "input" "file" "N" "" "" 1 "" "")
;
quit;

%cstutil_processsetup();

/* An F will be appended to CodeList Submission Values that end with a digit;
/* This will create valid formats from CodeLists like TOXGRV3;

%ct_createformats(
    _cstLang=en,
    _cstCreateCatalog=1,
    _cstKillCatFirst=1,
    _cstUseExpression=,
    _cstAppendChar=F
);

```

Alternatively a custom format can be used to map CodeList Submission Values to valid SAS format names;

```

proc format lib=work.formats;
    value $ tox
        "TOXGRV3" = "TOXGRV3F"
        "TOXGRV4" = "TOXGRV4F"
    ;
quit;

%ct_createformats(
    _cstLang=en,
    _cstKillCatFirst=1,
    _cstUseExpression=%str(strip(put(cdiscsubmissionvalue, $_tox32.)))
);

```

Or, in the case of Questionnaire CodeLists:

```

proc format;
    value $ qs
        "ADAS-Cog CDISC Version TEST"    "ADASCOG "
        "ADAS-Cog CDISC Version TESTCD"  "ADASCOGC "
        "ADCS-CGIC TEST"                  "ADCSGIC "
        "ADCS-CGIC TESTCD"                 "ADCSGICC"
        "AIMS TEST"                        "AIMS "
        "AIMS TESTCD"                       "AIMSC "
        "AVLT TEST"                         "AVLT "
        "AVLT TESTCD"                       "AVLTC "
        "BARS TEST"                         "BARS "
        "BARS TESTCD"                       "BARSC "
        "BPI Short Form TEST"               "BPI "
        "BPI Short Form TESTCD"             "BPIC "
        "BPI TEST"                          "BPISH "
        "BPI TESTCD"                        "BPISHC "
        "BPRS-A TEST"                       "BPRSA "
        ...
        "WPAI-SHP TEST"                     "WPAISHP "
        "WPAI-SHP TESTCD"                    "WPAISHPC "
        "YMRS TEST"                          "YMRS "
        "YMRS TESTCD"                        "YMRSC "
    ;
quit;

%ct_createformats(
    _cstLang=en,
    _cstCreateCatalog=1,
    _cstKillCatFirst=1,
    _cstUseExpression=%str(strip(put(cdiscsubmissionvalue, $_qs32.))),
    _cstAppendChar=F
);

```

CONCLUSION

With the production releases of SAS Clinical Standards Toolkit 1.5, the CDISC ODM 1.3.1 reference standard will support reading and representing in SAS a complete ODM XML file, building an ODM XML file, and validating the structure and content of the SAS representation of an ODM XML file. In addition, it will validate the structural integrity of the ODM XML file. Also, clinical data and reference data can be extracted from a CDISC ODM 1.3.1 file. The SAS Clinical Standard Toolkit 1.5 has also implemented the import of NCI CDISC Controlled Terminology files in ODM XML format.

REFERENCES

1. SAS Institute, Inc.: SAS® Clinical Standards Toolkit 1.5: User Guide. Cary, NC, USA: SAS Institute Inc. 2013. Available from: <http://support.sas.com/rnd/base/cdisc/cst/index.html>
2. CDISC Operational Data Model (ODM), Version 1.3.1, February 11, 2010 (<http://www.cdisc.org/odm>)
3. Eliotte Rusty Harold & W. Scott Means, 2004, [*XML in a Nutshell*](#), *A Desktop Quick Reference* (O'Reilly and Associates)
4. Terminology Resources: NCI Enterprise Vocabulary Services (EVS), Dictionaries, FedMed, FDA, CDISC, and NCPDP Terminology. (<http://www.cancer.gov/cancertopics/cancerlibrary/terminologyresources/cdisc>)
FTP site: <ftp://ftp1.nci.nih.gov/pub/cacore/EVS/CDISC/>

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Lex Jansen
SAS Institute Inc.
SAS Campus Drive
Cary, North Carolina 27513 USA
Phone: 919-531-9860
Email: lex.jansen@sas.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are registered trademarks or trademarks of their respective companies.

Appendix 1: ODM 1.3.1 SAS data sets

Below is the complete listing of the SAS data sets, with member columns for each, which will comprise the ODM 1.3.1 data in the SAS Clinical Standards Toolkit v.1.4.

No data set has more than one variable that acts as the key/index for that table. The names of key variables are prepended with two asterisks (**). Some tables do not have any key at all. Foreign key variables' names are prepended with two carat characters (^) and reference (in [brackets]) the name of the data set for which it is a foreign key. Required fields (fields for which, in any observation for that data set, a non-nil, and non-whitespace-only value must be supplied) are marked with an 'X' between brackets: [X].

Per the standard, which is very flexible, only the DefineDocument data set, containing valid values for the FileOID and FileType variables, are needed in order to create a minimal, but valid ODM-compliant XML document.

All table and column names are case-sensitive, and must be specified exactly as shown.

In the SAS implementation of the relational data model, the keys were extended to define a unique record in every SAS data set. For example, a unique record in the EnumeratedItems SAS data set is defined by the variables FK_CODELISTS and CODEDVALUE. These SAS data set keys can be found in the SAS *reference_tables* SAS data set, which contains the metadata for the tables that define the SAS representation of the ODM 1.3.1 standard:

<global standards library directory>\standards\cdisc-odm-1.3.1-1.5\metadata\reference_tables.sas7bdat

SAS Data Set Name	Variable Name	SAS Data Type	Length (if char)
ODM			
	**FileOID	character	64 (oid)
	Archival	character	3
	AsOfDateTime	character	32
	CreationDateTime	character	32
	Description	character	2000 (text)
	FileType	character	13
	Granularity	character	15
	Id	character	64 (oid)
	ODMVersion	character	2000 (text)
	Originator	character	2000 (text)
	PriorFileOID	character	64 (oid)
	SourceSystem	character	2000 (text)
	SourceSystemVersion	character	2000 (text)
Study			
	**OID	character	64 (oid)
	StudyName	character	128 (name)
	StudyDescription	character	2000 (text)
	ProtocolName	character	128 (name)
	^FK_ODM [ODM]	character	64 (oid)
MeasurementUnits			
	**OID	character	64 (oid)
	Name	character	128 (name)
	^FK_Study [Study]	character	64 (oid)
MUTranslatedText			
	TranslatedText	character	2000 (text)
	lang	character	128 (name)
	^FK_MeasurementUnits [MeasurementUnits]	character	64 (oid)
MUAliases			
	Context	character	2000 (text)
	Name	character	2000 (text)
	^FK_MeasurementUnits [MeasurementUnits]	character	64 (oid)
MetaDataVersion			
	**OID	character	64 (oid)
	Name	character	128 (name)
	Description	character	2000 (text)
	IncludedOID	character	64 (oid)
	IncludedStudyOID	character	64 (oid)
	^FK_Study [Study]	character	64 (oid)
ProtocolEventRefs			
	Mandatory	character	3
	OrderNumber	numeric	8
	^StudyEventOID [StudyEventDefs]	character	64 (oid)
	^FK_MetaDataVersion [MetaDataVersion]	character	64 (oid)
	^CollectionExceptionConditionOID [ConditionDefs]	character	64 (oid)
ProtocolTranslatedText			

SAS Data Set Name	Variable Name	SAS Data Type	Length (if char)
	TranslatedText	character	2000 (text)
	lang	character	17
	^FK_MetaDataVersion [MetaDataVersion]	character	64 (oid)
ProtocolAliases			
	Context	character	2000 (text)
	Name	character	2000 (text)
	^FK_MetaDataVersion [MetaDataVersion]	character	64 (oid)
StudyEventDefs			
	**OID	character	64 (oid)
	Category	character	2000 (text)
	Name	character	128 (name)
	Repeating	character	3
	Type	character	11
	^FK_MetaDataVersion [MetaDataVersion]	character	64 (oid)
StudyEventDefTranslatedText			
	TranslatedText	character	2000 (text)
	lang	character	17
	^FK_StudyEventDefs [StudyEventDefs]	character	64 (oid)
StudyEventFormRefs			
	^FormOID [FormDefs]	character	64 (oid)
	Mandatory	character	3
	OrderNumber	numeric	8
	^FK_StudyEventDefs [StudyEventDefs]	character	64 (oid)
	^CollectionExceptionConditionOID [ConditionDefs]	character	64 (oid)
StudyEventAliases			
	Context	character	2000 (text)
	Name	character	2000 (text)
	^FK_StudyEventDefs [StudyEventDefs]	character	64 (oid)
FormDefs			
	**OID	character	64 (oid)
	Name	character	128 (name)
	Repeating	character	3
	^FK_MetaDataVersion [MetaDataVersion]	character	64 (oid)
FormDefTranslatedText			
	TranslatedText	character	2000 (text)
	lang	character	17
	^FK_FormDefs [FormDefs]	character	64 (oid)
FormDefItemGroupRefs			
	^ItemGroupOID [ItemGroupDefs]	character	64 (oid)
	Mandatory	character	3
	OrderNumber	numeric	8
	^FK_FormDefs [FormDefs]	character	64 (oid)
	^CollectionExceptionConditionOID [ConditionDefs]	character	64 (oid)
FormDefArchLayouts			
	**OID	character	64 (oid)
	PdfFileName	character	512 (path)
	^PresentationOID [Presentation]	character	64 (oid)
	^FK_FormDefs [FormDefs]	character	64 (oid)
FormAliases			
	Context	character	2000 (text)
	Name	character	2000 (text)
	^FK_FormDefs [FormDefs]	character	64 (oid)
ItemGroupDefs			
	**OID	character	64 (oid)
	Name	character	128 (name)
	Repeating	character	3
	IsReferenceData	character	3
	SASDatasetName	character	8
	Domain	character	2000 (text)
	Origin	character	2000 (text)
	Role	character	128 (name)
	Purpose	character	2000 (text)
	Comment	character	2000 (text)
	^FK_MetaDataVersion [MetaDataVersion]	character	64 (oid)
ItemGroupDefTranslatedText			
	TranslatedText	character	2000 (text)
	lang	character	17
	^FK_ItemGroupDefs [ItemGroupDefs]	character	64 (oid)
ItemGroupDefItemRefs			

SAS Data Set Name	Variable Name	SAS Data Type	Length (if char)
	^ItemOID [ItemDefs]	character	64 (oid)
	Mandatory	character	3
	OrderNumber	numeric	8
	KeySequence	numeric	8
	^ImputationMethodOID [ImputationMethods]	character	64 (oid)
	^MethodOID [MethodDefs]	character	64 (oid)
	Role	character	128 (name)
	^RoleCodeListOID [CodeLists]	character	64 (oid)
	^FK_ItemGroupDefs [ItemGroupDefs]	character	64 (oid)
	^CollectionExceptionConditionOID [ConditionDefs]	character	64 (oid)
ItemGroupAliases			
	Context	character	2000 (text)
	Name	character	2000 (text)
	^FK_ItemGroupDefs [ItemGroupDefs]	character	64 (oid)
ItemDefs			
	**OID	character	64 (oid)
	Name	character	128 (name)
	DataType	character	18
	Length	numeric	8
	SignificantDigits	numeric	8
	SASFieldName	character	8
	SDSVarName	character	8
	Origin	character	2000 (text)
	Comment	character	2000 (text)
	^CodeListRef [CodeLists]	character	64 (oid)
	^FK_MetaDataVersion [MetaDataVersion]	character	64 (oid)
ItemDefTranslatedText			
	TranslatedText	character	2000 (text)
	lang	character	17
	^FK_ItemDefs [ItemDefs]	character	64 (oid)
ItemQuestionTranslatedText			
	TranslatedText	character	2000 (text)
	lang	character	17
	^FK_ItemDefs [ItemDefs]	character	64 (oid)
ItemQuestionExternal			
	Dictionary	character	2000 (text)
	Version	character	2000 (text)
	Code	character	2000 (text)
	^FK_ItemDefs [ItemDefs]	character	64 (oid)
ItemMURefs			
	^MeasurementUnitOID [MeasurementUnits]	character	64 (oid)
	^FK_ItemDefs [ItemDefs]	character	64 (oid)
ItemRangeChecks			
	**OID	character	64 (oid)
	Comparator	character	5
	SoftHard	character	4
	^MURefOID [MeasurementUnits]	character	64 (oid)
	^FK_ItemDefs [ItemDefs]	character	64 (oid)
ItemRangeCheckValues			
	CheckValue	character	512 (value)
	^FK_ItemRangeChecks [ItemRangeChecks]	character	64 (oid)
RCErrorsTranslatedText			
	TranslatedText	character	2000 (text)
	lang	character	17
	^FK_ItemRangeChecks [ItemRangeChecks]	character	64 (oid)
ItemRCFormalExpression			
	Context	character	2000 (text)
	Expression	character	2000 (text)
	^FK_ItemRangeChecks [ItemRangeChecks]	character	64 (oid)
ItemRole			
	Name	character	2000 (text)
	^FK_ItemDefs [ItemDefs]	character	64 (oid)
ItemAliases			
	Context	character	2000 (text)
	Name	character	2000 (text)
	^FK_ItemDefs [ItemDefs]	character	64 (oid)
CodeLists			
	**OID	character	64 (oid)
	Name	character	128 (name)

SAS Data Set Name	Variable Name	SAS Data Type	Length (if char)
	DataType	character	7
	SASFormatName	character	8
	^FK_MetaDataVersion [MetaDataVersion]	character	64 (oid)
CodeListTranslatedText			
	TranslatedText	character	2000 (text)
	lang	character	17
	^FK_CodeLists [CodeLists]	character	64 (oid)
CodeListItems			
	##OID	character	64 (oid)
	CodedValue	character	512 (value)
	Rank	numeric	8
	^FK_CodeLists [CodeLists]	character	64 (oid)
CLItemDecodeTranslatedText			
	TranslatedText	character	2000 (text)
	lang	character	17
	^FK_CodeListItems [CodeListItems]	character	64 (oid)
CodeListItemAliases			
	Context	character	2000 (text)
	Name	character	2000 (text)
	^FK_CodeListItems [CodeListItems]	character	64 (oid)
ExternalCodeLists			
	Dictionary	character	2000 (text)
	Version	character	2000 (text)
	ref	character	2000 (text)
	href	character	2000 (text)
	^FK_CodeLists [CodeLists]	character	64 (oid)
EnumeratedItems			
	##OID	character	64 (oid)
	CodedValue	character	512 (value)
	Rank	numeric	8
	^FK_CodeLists [CodeLists]	character	64 (oid)
EnumeratedItemAliases			
	Context	character	2000 (text)
	Name	character	2000 (text)
	^FK_EnumeratedItems [EnumeratedItems]	character	64 (oid)
CodeListAliases			
	Context	character	2000 (text)
	Name	character	2000 (text)
	^FK_CodeLists [CodeLists]	character	64 (oid)
ImputationMethods			
	**OID	character	64 (oid)
	method	character	2000 (text)
	^FK_MetaDataVersion [MetaDataVersion]	character	64 (oid)
Presentation			
	**OID	character	64 (oid)
	presentation	character	2000 (text)
	lang	character	17
	^FK_MetaDataVersion [MetaDataVersion]	character	64 (oid)
MethodDefs			
	**OID	character	64 (oid)
	Name	character	128 (name)
	Type	character	11
	^FK_MetaDataVersion [MetaDataVersion]	character	64 (oid)
MethodDefTranslatedText			
	TranslatedText	character	2000 (text)
	lang	character	17
	^FK_MethodDefs [MethodDefs]	character	64 (oid)
MethodDefFormalExpression			
	Context	character	2000 (text)
	Expression	character	2000 (text)
	^FK_MethodDefs [MethodDefs]	character	64 (oid)
MethodAliases			
	Context	character	2000 (text)
	Name	character	2000 (text)
	^FK_MethodDefs [MethodDefs]	character	64 (oid)
ConditionDefs			
	**OID	character	64 (oid)
	Name	character	128 (name)
	^FK_MetaDataVersion [MetaDataVersion]	character	64 (oid)

SAS Data Set Name	Variable Name	SAS Data Type	Length (if char)
ConditionDefTranslatedText			
	TranslatedText	character	2000 (text)
	lang	character	17
	^FK_ConditionDefs [ConditionDefs]	character	64 (oid)
ConditionDefFormalExpression			
	Context	character	2000 (text)
	Expression	character	2000 (text)
	^FK_ConditionDefs [ConditionDefs]	character	64 (oid)
ConditionAliases			
	Context	character	2000 (text)
	Name	character	2000 (text)
	^FK_ConditionDefs [ConditionDefs]	character	64 (oid)
ClinicalData			
	##OID	character	64 (oid)
	^StudyOID [Study]	character	64 (oid)
	^MetaDataVersionOID [MetaDataVersion]	character	64 (oid)
	^FK_ODM [ODM]	character	64 (oid)
SubjectData			
	##OID	character	64 (oid)
	SubjectKey	character	2000 (text)
	TransactionType	character	7
	^InvestigatorRefOID [User]	character	64 (oid)
	^SiteRefOID [Location]	character	64 (oid)
	^FK_ClinicalData [ClinicalData]	character	64 (oid)
StudyEventData			
	##OID	character	64 (oid)
	^StudyEventOID [StudyEventDefs]	character	64 (oid)
	StudyEventRepeatKey	character	2000 (text)
	TransactionType	character	7
	^FK_SubjectData [SubjectData]	character	64 (oid)
FormData			
	##OID	character	64 (oid)
	^FormOID [FormDefs]	character	64 (oid)
	FormRepeatKey	character	2000 (text)
	TransactionType	character	7
	^ArchiveLayoutRefOID [FormDefArchLayouts]	character	64 (oid)
	^FK_StudyEventData [StudyEventData]	character	64 (oid)
ItemGroupData			
	##OID	character	64 (oid)
	^ItemGroupOID [ItemGroupDefs]	character	64 (oid)
	ItemGroupRepeatKey	character	2000 (text)
	TransactionType	character	7
	^FK_FormData [FormData]	character	64 (oid)
	^FK_ReferenceData [ReferenceData]	character	64 (oid)
ItemData			
	##OID	character	64 (oid)
	^ItemOID [ItemDefs]	character	64 (oid)
	TransactionType	character	7
	TransactionOrder	numeric	8
	Value	character	2000 (text)
	IsNull	character	3
	ItemDataType	character	18
	^FK_ItemGroupData [ItemGroupData]	character	64 (oid)
	AuditRecordID [AuditRecord]	character	64 (oid)
	SignatureID [Signature]	character	64 (oid)
	AnnotationID [Annotation]	character	64 (oid)
	MeasurementUnitOID [MeasurementUnits]	character	64 (oid)
ReferenceData			
	##GeneratedID	character	64 (oid)
	^StudyOID [Study]	character	64 (oid)
	^MetaDataVersionOID [MetaDataVersion]	character	64 (oid)
	^FK_ODM [ODM]	character	64 (oid)
AdminData			
	##GeneratedID	character	64 (oid)
	StudyOID	character	64 (oid)
	^FK_ODM [ODM]	character	64 (oid)
User			
	**OID	character	64 (oid)
	UserType	character	12 (enum)

SAS Data Set Name	Variable Name	SAS Data Type	Length (if char)
	LoginName	character	2000 (text)
	DisplayName	character	2000 (text)
	FullName	character	2000 (text)
	FirstName	character	2000 (text)
	LastName	character	2000 (text)
	Organization	character	2000 (text)
	PictureImageType	character	2000 (text)
	PictureFileName	character	2000 (text)
	Pager	character	2000 (text)
	^FK_AdminData [AdminData]	character	64 (oid)
UserAddress			
	##GeneratedID	character	64 (oid)
	City	character	2000 (text)
	StateProv	character	2000 (text)
	Country	character	2 (text)
	PostalCode	character	2000 (text)
	OtherText	character	2000 (text)
	^FK_User [User]	character	64 (oid)
UserAddressStreetName			
	StreetName	character	2000 (text)
	^FK_UserAddress [UserAddress]	character	64 (oid)
UserEmail			
	Email	character	2000 (text)
	^FK_User [User]	character	64 (oid)
UserFax			
	Fax	character	2000 (text)
	^FK_User [User]	character	64 (oid)
UserPhone			
	Phone	character	2000 (text)
	^FK_User [User]	character	64 (oid)
UserLocationRef			
	^LocationOID [Location]	character	64 (oid)
	^FK_User [User]	character	64 (oid)
Location			
	**OID	character	64 (oid)
	Name	character	2000 (text)
	LocationType	character	7 (enum)
	^FK_AdminData [AdminData]	character	64 (oid)
LocationVersion			
	^StudyOID [Study]	character	64 (oid)
	^MetaDataVersionOID [MetaDataVersion]	character	64 (oid)
	EffectiveDate	character	10 (date)
	^FK_Location [Location]	character	64 (oid)
SignatureDef			
	**OID	character	64 (oid)
	Methodology	character	10 (enum)
	Meaning	character	2000 (text)
	LegalReason	character	2000 (text)
	^FK_AdminData [AdminData]	character	64 (oid)
AuditRecord			
	ID	character	2000 (text)
	EditPoint	character	14 (enum)
	UsedImputationMethod	character	3 (Yes/No)
	^UserOID [User]	character	64 (oid)
	^LocationOID [Location]	character	64 (oid)
	DateTimeStamp	character	25 (datetime)
	ReasonForChange	character	2000 (text)
	SourceID	character	2000 (text)
	ParentType	character	14 (datanode)
	ParentKey	character	64 (oid)
	GrandParentType	character	14 (datanode)
	GrandParentKey	character	64 (oid)
Signature			
	ID	character	2000 (text)
	^UserOID [User]	character	64 (oid)
	^LocationOID [Location]	character	64 (oid)
	^SignatureDefOID [SignatureDef]	character	64 (oid)
	DateTimeStamp	character	25 (datetime)
	ParentType	character	14 (datanode)

SAS Data Set Name	Variable Name	SAS Data Type	Length (if char)
	ParentKey	character	64 (oid)
	GrandParentType	character	14 (datanode)
	GrandParentKey	character	64 (oid)
Annotation			
	##GeneratedID	character	64 (oid)
	ID	character	2000 (text)
	SeqNum	numeric	8 (integer)
	TransactionType	character	7 (enum)
	CommentSponsorOrSite	character	7 (enum)
	Comment	character	2000 (text)
	ParentType	character	14 (datanode)
	ParentKey	character	64 (oid)
	GrandParentType	character	14 (datanode)
	GrandParentKey	character	64 (oid)
AnnotationFlag			
	FlagValue	character	2000 (text)
	^FlagValueCodeListOID [CodeLists]	character	64 (oid)
	FlagType	character	128 (name)
	^FlagTypeCodeListOID [CodeLists]	character	64 (oid)
	^FK_Annotation [Annotation]	character	64 (oid)
Association			
	##GeneratedID	character	64 (oid)
	^StudyOID [Study]	character	64 (oid)
	^MetaDataVersionOID [MetaDataVersion]	character	64 (oid)
	^FK_ODM [ODM]	character	64 (oid)
KeySet			
	**OID	character	64 (oid)
	^StudyOID [Study]	character	64 (oid)
	SubjectKey	character	2000 (text)
	^StudyEventOID [StudyEventDefs]	character	64 (oid)
	StudyEventRepeatKey	character	2000 (text)
	^FormOID [FormDefs]	character	64 (oid)
	FormRepeatKey	character	2000 (text)
	^ItemGroupOID [ItemGroupDefs]	character	64 (oid)
	ItemGroupRepeatKey	character	2000 (text)
	^ItemOID [ItemDefs]	character	64 (oid)
	^FK_Association [Association]	character	64 (oid)