Using the SAS® Clinical Standards Toolkit for define.xml creation

Lex Jansen

SAS Institute Inc.

PHARMASUG – 2011

Agenda

- What is the define.xml (CRT-DDS)?
- The CRT-DDS SAS Data Model
- What is SAS Clinical Standards Toolkit (CST)?
- CST Framework
- CST Metadata Files
- Hands-on Part
  - Create CRT-DDS SAS data sets from the SDTM 3.1.2 metadata
  - Validate the CRT-DDS SAS data sets
  - Create the define.xml from the CRT-DDS SAS data sets
  - Extend the CRT-DDS SAS datasets and create the define.xml
  - Import a define.xml file to CRT-DDS SAS data sets

What is the define.xml (CRT-DDS)?

What is the define.xml (CRT-DDS)?

- **July 2004** – FDA adds Study Data Specifications v1.0 to draft eCTD Guidance. This specification references the CDISC SDTM for data tabulation datasets

Electronic Common Technical Document (eCTD)

- Draft Guidance for Industry on Providing Regulatory Submissions in Electronic Format--Human Pharmaceutical Applications and Related Submissions. (Posted 8/28/2003)
  - *Federal Register* Notice [TXT] [PDF]
  - The Draft Guidance

Specifications

- eCTD Backbone Files Specification for Module 1
- eCTD Backbone File Specification for Modules 2 through 5
- eCTD Backbone File Specification for Study Tagging Files
- FDA eCTD Table of Contents Headings and Hierarchy
- Study Data Specifications New!! (Posted 7/21/2004)

What is the define.xml (CRT-DDS)?

- **March 2005** – Study Data Specifications v1.1: Updates Specifications for Data Set Documentation
  - data definitions
  - annotated case report forms (CRFs)
- *"The specification for the data definitions for datasets provided using the CDISC SDTM is included in the Case Report Tabulation Data Definition Specification (**define.xml**) developed by the CDISC define.xml Team"*
- Data Definition for other data sets follows: *Providing Regulatory Submissions in Electronic Format –NDA* (1999), which is the **define.pdf**

What is the define.xml (CRT-DDS)?

January 2010 - Study Data Specifications v1.5.1:

- *"For datasets not prepared using the CDISC SDTM specifications, consult Appendix 2 for information concerning the preparation of a* ***define.pdf*** *data definition file."*
- Appendix 2 specifies a define.pdf specification similar to the 1999 guidance

What is the define.xml (CRT-DDS)?

- As of **January 1, 2008**: follow the eCTD guidance and document submitted data by including data definition tables (define.xml) and annotated case report forms (blankcrf.pdf)

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<?xml-stylesheet type="text/xsl" href="define1-0-0.xsl"?>
<ODM xmlns="http://www.cdisc.org/ns/odm/v1.2"
     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
     xmlns:xlink="http://www.w3.org/1999/xlink"
     xmlns:def="http://www.cdisc.org/ns/def/v1.0"
     xsi:schemaLocation="http://www.cdisc.org/ns/odm/v1.2 define1-0-0.xsd"
     FileOID="Studycdisc01"
     ODMVersion="1.2" FileType="Snapshot"
     CreationDateTime="2007-04-09T12:24:09">
     - <Study OID="cdisc01">
       - <GlobalVariables>
            <StudyName>CDISC01</StudyName>
            <StudyDescription>CDISC01 Test Study.</StudyDescription>
            <ProtocolName>CDISC01</ProtocolName>
       </GlobalVariables>
       <MetaDataVersion  OID="CDISC.SDTM.3.1.1"
            Name="Study CDISC01, Data Definitions"
            Description="Study CDISC01, Data Definitions"
            def:DefineVersion="1.0.0"
            def:StandardName="CDISC SDTM"
            def:StandardVersion="3.1.1">
...
     <ItemGroupDef OID="MH"
            Name="MH" Repeating="Yes" IsReferenceData="No"
            Purpose="Tabulation" def:Label="Medical History"
            def:Structure="One record per medical history event per subject"
            def:DomainKeys="STUDYID, USUBJID, MHCAT, MHTERM, MHSTDTC"
            def:Class="EVENTS"
            def:ArchiveLocationID="Location.MH">
       <ItemRef ItemOID="STUDYID" OrderNumber="1" Mandatory="Yes"
          Role="IDENTIFIER" RoleCodeListOID="RoleCodeList" />
     ...

       <def:leaf ID="Location.MH" xlink:href="mh.xpt">
            <def:title>mh.xpt</def:title>
       </def:leaf>
     </ItemGroupDef>
```

What is the define.xml (CRT-DDS)?

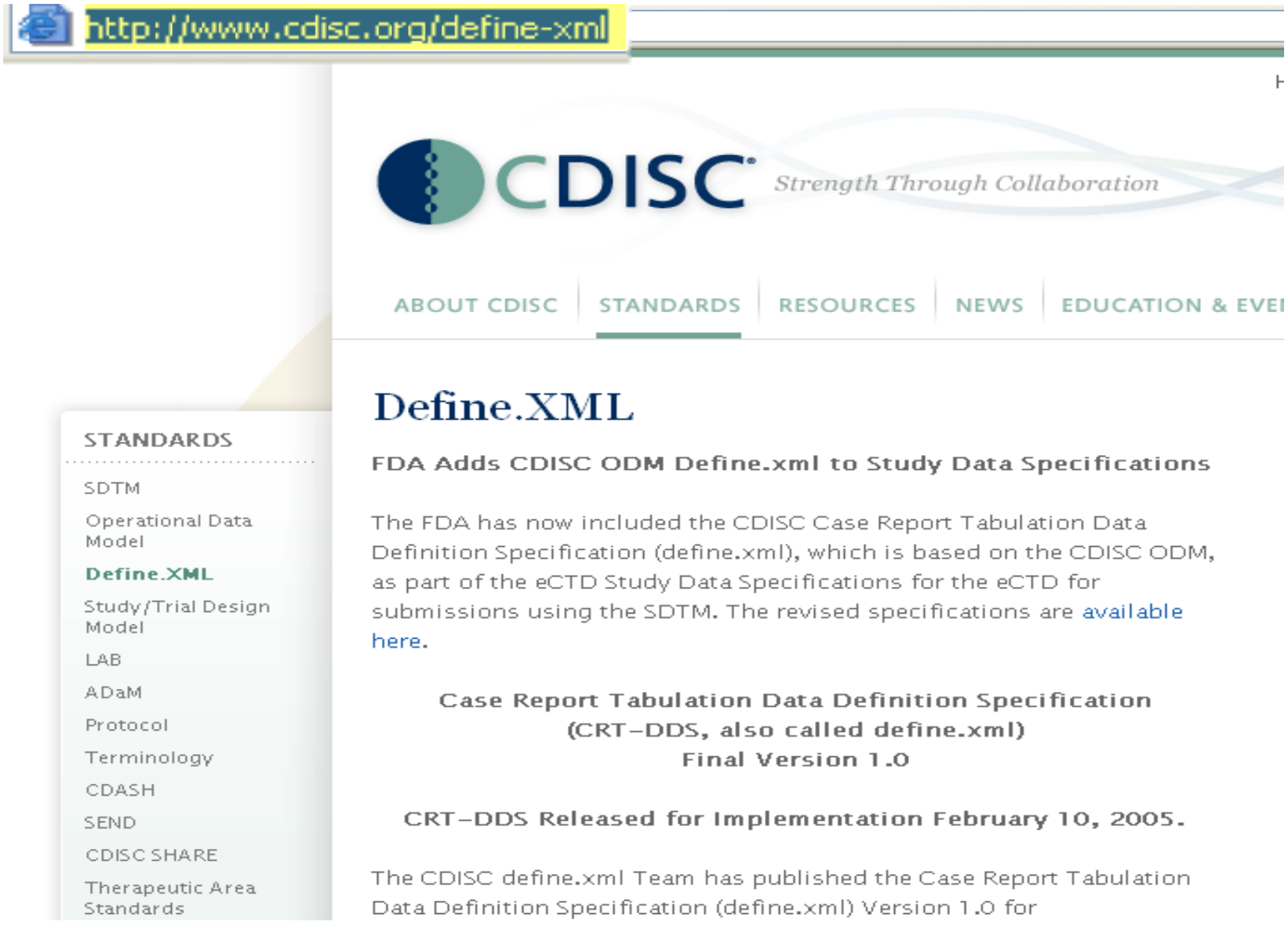
- As of **January 1, 2008**: follow the eCTD guidance and document submitted data by including data definition tables (define.xml) and annotated case report forms (blankcrf.pdf)

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<?xml-stylesheet type="text/xsl" href="define1-0-0.xsl"?>
<ODM xmlns="http://www.cdisc.org/ns/odm/v1.2"
     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
     xmlns:xlink="http://www.w3.org/1999/xlink"
     xmlns:def="http://www.cdisc.org/ns/def/v1.0"
     xsi:schemaLocation="http://www.cdisc.org/ns/odm/v1.2 define1-0-0.xsd"
     FileOID="Studycdisc01"
     ODMVersion="1.2" FileType="Snapshot"
     CreationDateTime="2007-04-09T12:24:09">
  - <Study OID="cdisc01">
    - <GlobalVariables>
        <StudyName>CDISC01</StudyName>
        <StudyDescription>CDISC01 Test Study.</StudyDescription>
        <ProtocolName>CDISC01</ProtocolName>
      </GlobalVariables>
      <MetaDataVersion OID="CDISC.SDTM.3.1.1"
          Name="Study CDISC01, Data Definitions"
          Description="Study CDISC01, Data Definitions"

  <ItemGroupDef OID="MH"
        Name="MH" Repeating="Yes" IsReferenceData="No"
        Purpose="Tabulation" def:Label="Medical History"
        def:Structure="One record per medical history event per subject"
        def:DomainKeys="STUDYID, USUBJID, MHCAT, MHTERM, MHSTDTC"
        def:Class="EVENTS"
        def:ArchiveLocationID="Location.MH">
    <ItemRef ItemOID="STUDYID" OrderNumber="1" Mandatory="Yes"
        Role="IDENTIFIER" RoleCodeListOID="RoleCodeList" />
  ...

    <def:leaf ID="Location.MH" xlink:href="mh.xpt">
            <def:title>mh.xpt</def:title>
    </def:leaf>
  </ItemGroupDef>
</ItemGroupDef>
```

What is the define.xml (CRT-DDS)?

- Case Report Tabulation Data Specification (CRT-DDS, or define.xml): Production version: **1.0.0**
  CRT-DDS **1.0.0** is the only production version right now
- Extension of the CDISC Operational Data Model (**ODM**), an **XML** specification to facilitate the archive and interchange of the metadata and data for clinical research
- Maintained by CDISC's **XML Technologies Team** (formerly known as the ODM team)
- New define.xml version 2 in development with additional metadata support for SDTM and ADaM
  (based on ODM 1.3.1)

What is the define.xml (CRT-DDS)?

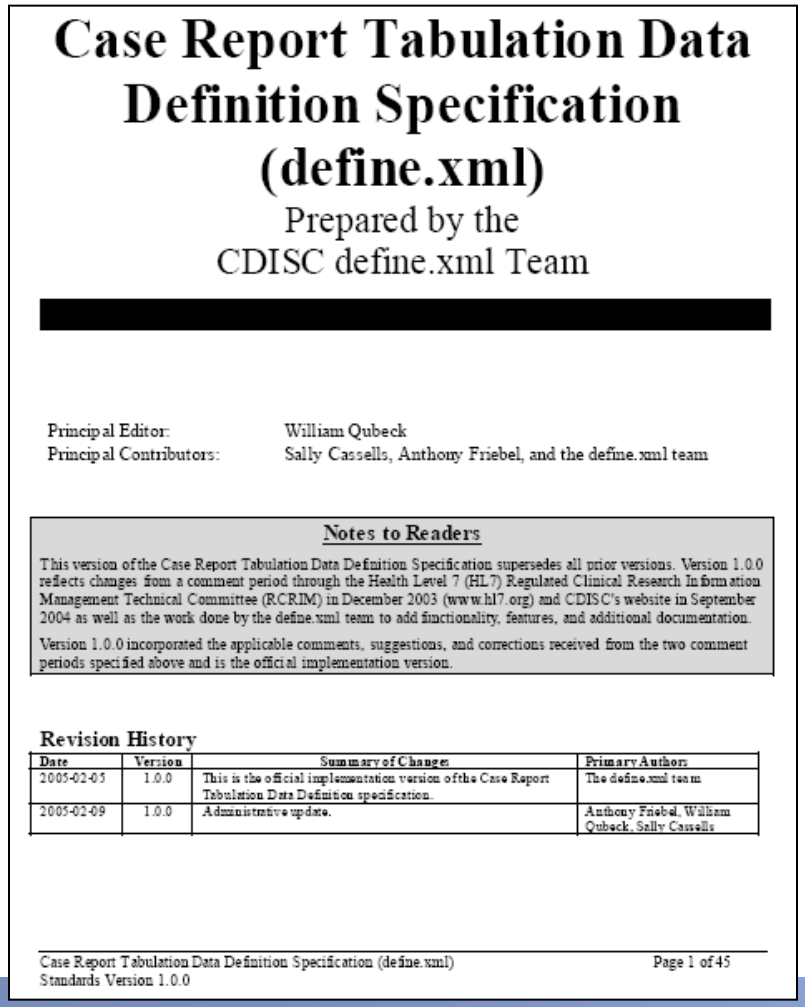


What is the define.xml (CRT-DDS)?

The specifications

http://www.cdisc.org/define-xml

http://www.cdisc.org/odm



Case Report Tabulation Data Definition Specification (define.xml)

Prepared by the CDISC define.xml Team

Principal Editor: William Qubeck
Principal Contributors: Sally Cassells, Anthony Friebel, and the define.xml team

Notes to Readers

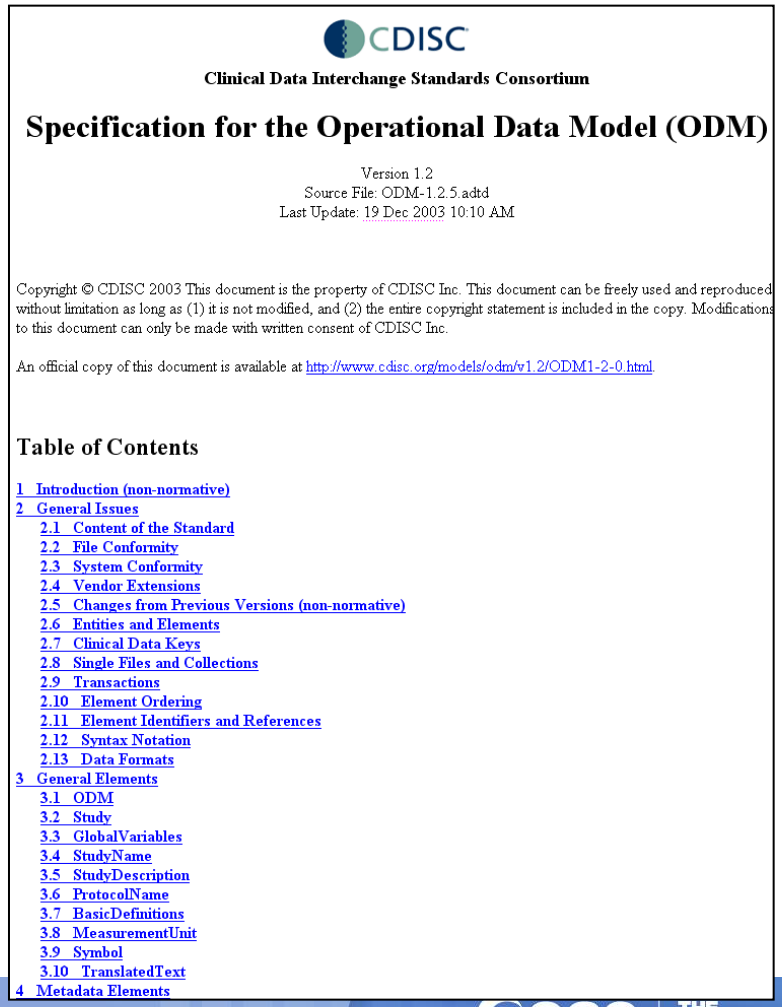
This version of the Case Report Tabulation Data Definition Specification supersedes all prior versions. Version 1.0.0 reflects changes from a comment period through the Health Level 7 (HL7) Regulated Clinical Research Information Management Technical Committee (RCRIM) in December 2003 (www.hl7.org) and CDISC's website in September 2004 as well as the work done by the define.xml team to add functionality, features, and additional documentation.

Version 1.0.0 incorporated the applicable comments, suggestions, and corrections received from the two comment periods specified above and is the official implementation version.

Revision History

Date	Version	Summary of Changes	Primary Authors
2005-02-05	1.0.0	This is the official implementation version of the Case Report Tabulation Data Definition specification.	The define.xml team
2005-02-09	1.0.0	Administrative update.	Anthony Friebel, William Qubeck, Sally Cassells

Case Report Tabulation Data Definition Specification (define.xml) Page 1 of 45
Standards Version 1.0.0



CDISC

Clinical Data Interchange Standards Consortium

Specification for the Operational Data Model (ODM)

Version 1.2
Source File: ODM-1.2.5.adtd
Last Update: 19 Dec 2003 10:10 AM

Copyright © CDISC 2003 This document is the property of CDISC Inc. This document can be freely used and reproduced without limitation as long as (1) it is not modified, and (2) the entire copyright statement is included in the copy. Modifications to this document can only be made with written consent of CDISC Inc.

An official copy of this document is available at http://www.cdisc.org/models/odm/v1.2/ODM1-2-0.html.

Table of Contents

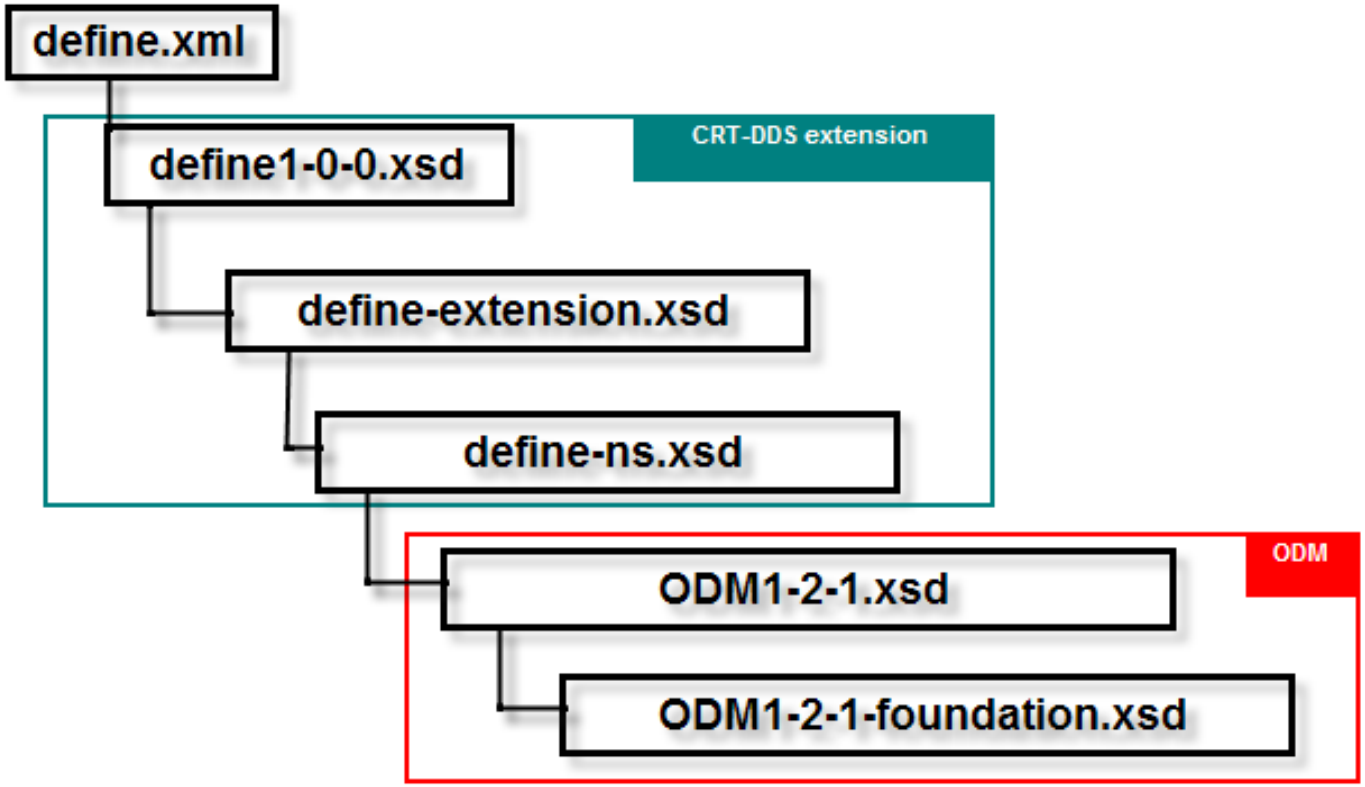
1 Introduction (non-normative)
2 General Issues
- 2.1 Content of the Standard
- 2.2 File Conformity
- 2.3 System Conformity
- 2.4 Vendor Extensions
- 2.5 Changes from Previous Versions (non-normative)
- 2.6 Entities and Elements
- 2.7 Clinical Data Keys
- 2.8 Single Files and Collections
- 2.9 Transactions
- 2.10 Element Ordering
- 2.11 Element Identifiers and References
- 2.12 Syntax Notation
- 2.13 Data Formats

3 General Elements
- 3.1 ODM
- 3.2 Study
- 3.3 GlobalVariables
- 3.4 StudyName
- 3.5 StudyDescription
- 3.6 ProtocolName
- 3.7 BasicDefinitions
- 3.8 MeasurementUnit
- 3.9 Symbol
- 3.10 TranslatedText

4 Metadata Elements

What is the define.xml (CRT-DDS)?

XML schema definitions (XSD) describe the structure of the define.xml



What is the define.xml (CRT-DDS)?

http://www.cdisc.org/define-xml

What is the define.xml (CRT-DDS)?

- define.xml contains **metadata** and is **machine** readable
- define.xml becomes **human** readable with a **stylesheet**

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<?xml-stylesheet type="text/xsl" href="define1-0-0.xsl"?>
<ODM
 xmlns="http://www.cdisc.org/ns/odm/v1.2"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xmlns:xlink="http://www.w3.org/1999/xlink"
 xmlns:def="http://www.cdisc.org/ns/def/v1.0"
 xsi:schemaLocation="http://www.cdisc.org/ns/odm/v1.2 define1-0-0.xsd"
 FileOID="Study1234"
 ODMVersion="1.2"
 FileType="Snapshot"
 CreationDateTime="2004-07-28T12:34:13-06:00">
<Study OID="1234">
 <GlobalVariables>
  <StudyName>1234</StudyName>
  <StudyDescription>1234 Data Definition</StudyDescription>
  <ProtocolName>1234</ProtocolName>
 </GlobalVariables>
 <MetaDataVersion OID="CDISC.SDTM.3.1.0"
  Name="Study 1234, Data Definitions"
  Description="Study 1234, Data Definitions"
  def:DefineVersion="1.0.0"
  def:StandardName="CDISC SDTM"
  def:StandardVersion="3.1.0">
  <def:AnnotatedCRF>
   <def:DocumentRef leafID="blankcrf"/>
  </def:AnnotatedCRF>
  <def:leaf ID="blankcrf" xlink:href="blankcrf.pdf">
   <def:title>Annotated Case Report Form</def:title>
  </def:leaf>
  <def:SupplementalDoc>
   <def:DocumentRef leafID="SupplementalDataDefinitions"/>
  </def:SupplementalDoc>
  <def:leaf ID="SupplementalDataDefinitions"
   xlink:href="supplementaldatadefinitions.pdf">
   <def:title>Supplemental Data Definitions Document</def:title>
  </def:leaf>
```

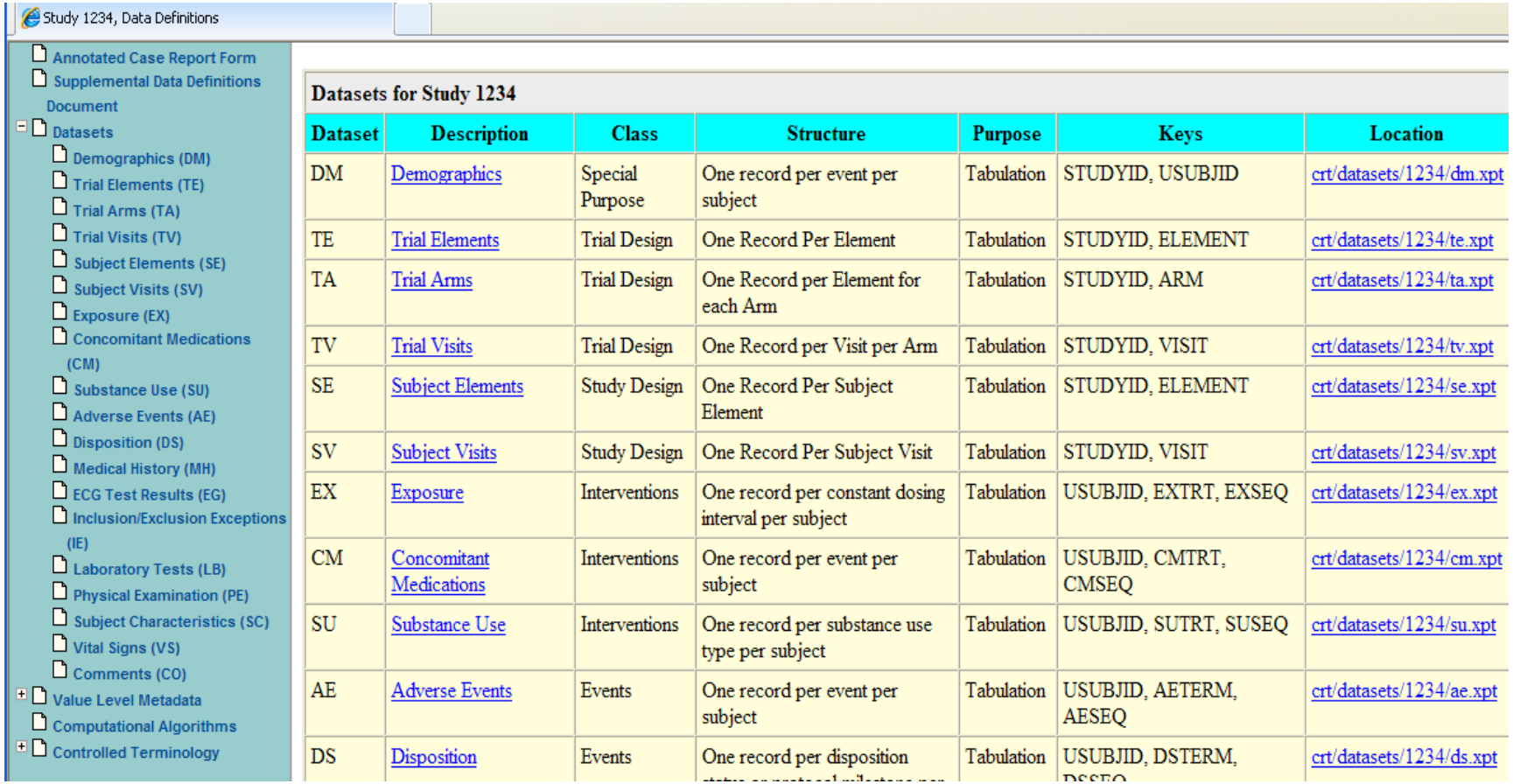
What is the define.xml (CRT-DDS)?

define.xml becomes **human** readable with an **XSL** **stylesheet**

Datasets for Study 1234					
Dataset	**Description**	**Structure**	**Purpose**	**Keys**	**Location**
DM	Demographics	Special Purpose - One record per event per subject	Tabulation	STUDYID, USUBJID	crt/datasets/1234/dm.xpt
TE	Trial Elements	Trial Design - One Record Per Element	Tabulation	STUDYID, ELEMENT	crt/datasets/1234/te.xpt
TA	Trial Arms	Trial Design - One Record per Element for each Arm	Tabulation	STUDYID, ARM	crt/datasets/1234/ta.xpt
TV	Trial Visits	Trial Design - One Record per Visit per Arm	Tabulation	STUDYID, VISIT	crt/datasets/1234/tv.xpt
SE	Subject Elements	Study Design - One Record Per Subject Element	Tabulation	STUDYID, ELEMENT	crt/datasets/1234/se.xpt
SV	Subject Visits	Study Design - One Record Per Subject Visit	Tabulation	STUDYID, VISIT	crt/datasets/1234/sv.xpt
EX	Exposure	Interventions - One record per constant dosing interval per subject	Tabulation	USUBJID, EXTRT, EXSEQ	crt/datasets/1234/ex.xpt
CM	Concomitant Medications	Interventions - One record per event per subject	Tabulation	USUBJID, CMTRT, CMSEQ	crt/datasets/1234/cm.xpt
SU	Substance Use	Interventions - One record per substance use type per subject	Tabulation	USUBJID, SUTRT, SUSEQ	crt/datasets/1234/su.xpt
AE	Adverse Events	Events - One record per event per subject	Tabulation	USUBJID, AETERM, AESEQ	crt/datasets/1234/ae.xpt
DS	Disposition	Events - One record per disposition status or protocol milestone per subject	Tabulation	USUBJID, DSTERM, DSSEQ	crt/datasets/1234/ds.xpt
MH	Medical History	Events - One record per event per subject	Tabulation	USUBJID, MHTERM, MHSEQ	crt/datasets/1234/mh.xpt
EG	ECG Test Results	Findings - One record per event per subject	Tabulation	USUBJID, EGTESTCD, EGSEQ	crt/datasets/1234/eg.xpt
IE	Inclusion/Exclusion Exceptions	Findings - One record per event per subject	Tabulation	USUBJID, IETEST, IESEQ	crt/datasets/1234/ie.xpt
LB	Laboratory Tests	Findings - One record per lab test per subject	Tabulation	USUBJID, LBTESTCD, VISITNUM, TPTNUM, LBSEQ	crt/datasets/1234/lb.xpt

What is the define.xml (CRT-DDS)?

… and looks even fancier with a different **stylesheet**



Datasets for Study 1234

Dataset	Description	Class	Structure	Purpose	Keys	Location
DM	Demographics	Special Purpose	One record per event per subject	Tabulation	STUDYID, USUBJID	crt/datasets/1234/dm.xpt
TE	Trial Elements	Trial Design	One Record Per Element	Tabulation	STUDYID, ELEMENT	crt/datasets/1234/te.xpt
TA	Trial Arms	Trial Design	One Record per Element for each Arm	Tabulation	STUDYID, ARM	crt/datasets/1234/ta.xpt
TV	Trial Visits	Trial Design	One Record per Visit per Arm	Tabulation	STUDYID, VISIT	crt/datasets/1234/tv.xpt
SE	Subject Elements	Study Design	One Record Per Subject Element	Tabulation	STUDYID, ELEMENT	crt/datasets/1234/se.xpt
SV	Subject Visits	Study Design	One Record Per Subject Visit	Tabulation	STUDYID, VISIT	crt/datasets/1234/sv.xpt
EX	Exposure	Interventions	One record per constant dosing interval per subject	Tabulation	USUBJID, EXTRT, EXSEQ	crt/datasets/1234/ex.xpt
CM	Concomitant Medications	Interventions	One record per event per subject	Tabulation	USUBJID, CMTRT, CMSEQ	crt/datasets/1234/cm.xpt
SU	Substance Use	Interventions	One record per substance use type per subject	Tabulation	USUBJID, SUTRT, SUSEQ	crt/datasets/1234/su.xpt
AE	Adverse Events	Events	One record per event per subject	Tabulation	USUBJID, AETERM, AESEQ	crt/datasets/1234/ae.xpt
DS	Disposition	Events	One record per disposition	Tabulation	USUBJID, DSTERM,	crt/datasets/1234/ds.xpt

The CRT-DDS SAS Data Model

The CRT-DDS SAS Data Model

```
1 <?xml version="1.0" encoding="ISO-8859-1" ?>
  <?xml-stylesheet type="text/xsl" href="define1-0-0.xsl"?>

  <ODM xmlns="http://www.cdisc.org/ns/odm/v1.2"
2      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
       xmlns:xlink="http://www.w3.org/1999/xlink"
       xmlns:def="http://www.cdisc.org/ns/def/v1.0"
       xsi:schemaLocation="http://www.cdisc.org/ns/odm/v1.2 define1-0-0.xsd"
       FileOID="Studycdisc01"
       ODMVersion="1.2" FileType="Snapshot"
       CreationDateTime="2007-04-09T12:24:09">
    <Study OID="cdisc01">
3     <GlobalVariables>
4       <StudyName>CDISC01</StudyName>
        <StudyDescription>CDISC01 Test Study.</StudyDescription>
        <ProtocolName>CDISC01</ProtocolName>
      </GlobalVariables>
      <MetaDataVersion OID="CDISC.SDTM.3.1.1"
5                      Name="Study CDISC01, Data Definitions"
                       Description="Study CDISC01, Data Definitions"
                       def:DefineVersion="1.0.0"
                       def:StandardName="CDISC SDTM"
                       def:StandardVersion="3.1.1">
6       < def:AnnotatedCRF >
        < def:SupplementalDoc >
        < def:leaf >
        < def:ComputationMethod >
        < def:ValueListDef >
        < ItemGroupDef >
        < ItemDef >
        < CodeList >
      </MetaDataVersion>
    </Study>
  </ODM>
```

<MetaDataVersion> Elements

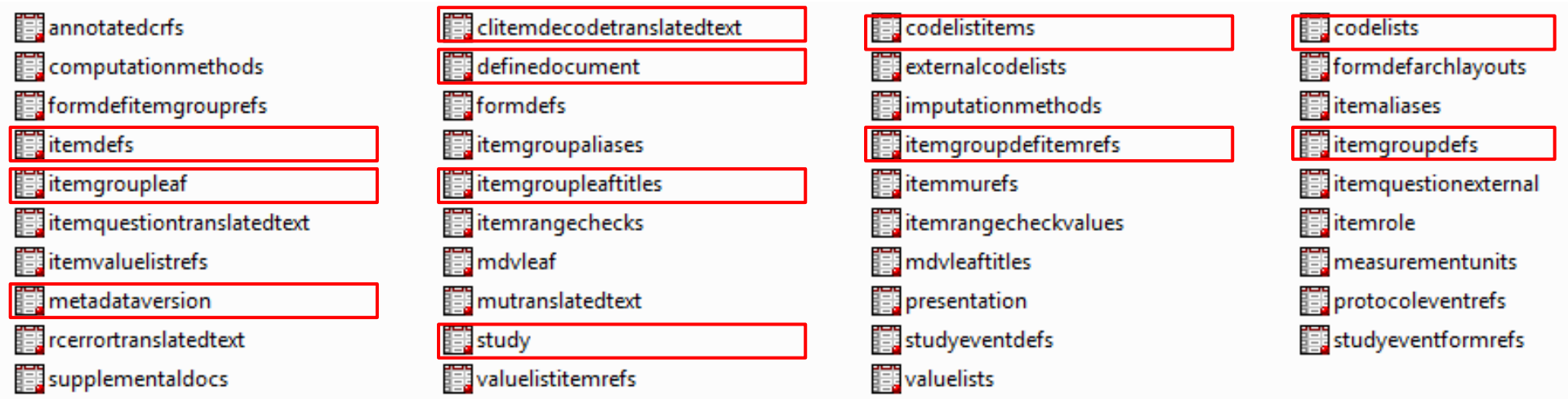
The CRT-DDS SAS Data Model

- define.xml has a deep hierarchy
- define.xml contains many relations

```
<ItemGroupDef OID="MH"
      Name="MH" Repeating="Yes" IsReferenceData="No"
      Purpose="Tabulation" def:Label="Medical History"
      def:Structure="One record per medical history event per subject"
      def:DomainKeys="STUDYID, USUBJID, MHCAT, MHTERM, MHSTDTC"
      def:Class="EVENTS" def:ArchiveLocationID="Location.AE">
  <ItemRef ItemOID="STUDYID" OrderNumber="1" Mandatory="Yes"
     Role="IDENTIFIER" RoleCodeListOID="RoleCodeList" />
...

   <def:leaf ID="Location.MH" xlink:href="mh.xpt">
            <def:title>mh.xpt</def:title>
   </def:leaf>
</ItemGroupDef>

...

<ItemDef OID="STUDYID" Name="STUDYID" DataType="text" Length="7"
         Origin="CRF Page 3" def:Label="Study Identifier" />
```

The CRT-DDS SAS Data Model

- SAS provides data model that represents CRT-DDS Version 1.0 format in 39 SAS data sets
- Patterned to match the XML element and attribute structure of the define.xml file



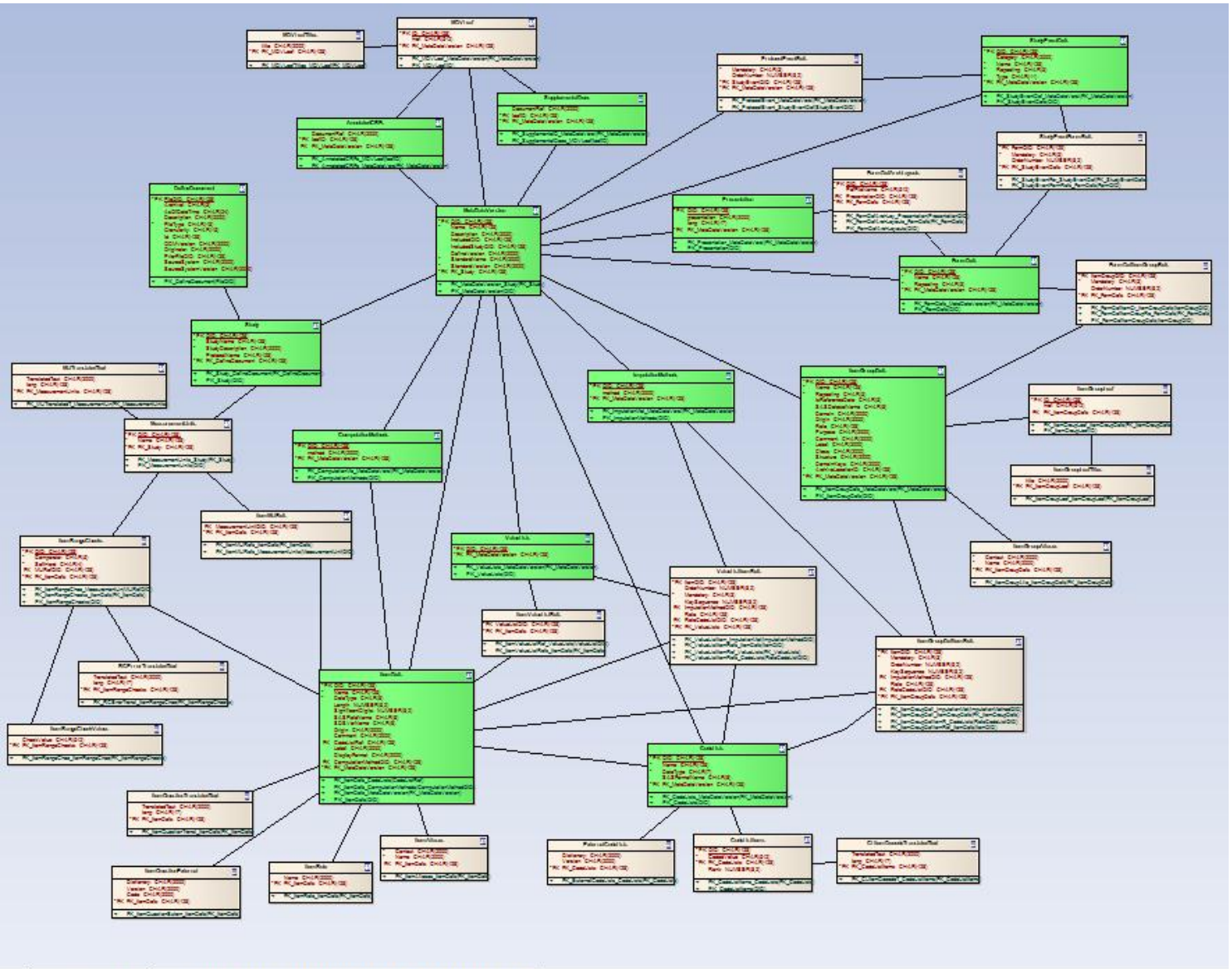
The CRT-DDS SAS Data Model

```
<ItemDef OID="COL10" Name="BRTHDTC" DataType="text" Length="64"
         SASFieldName="BRTHDTC" def:Label="Date/Time of Birth"/>
<ItemDef OID="COL11" Name="AGE" DataType="float"
          Length="8" SASFieldName="AGE" def:Label="Age"/>
<ItemDef OID="COL12" Name="AGEU" DataType="text"
         Length="10" SASFieldName="AGEU" def:Label="Age Units">
         <CodeListRef CodeListOID="AGEU"/>
</ItemDef>
```

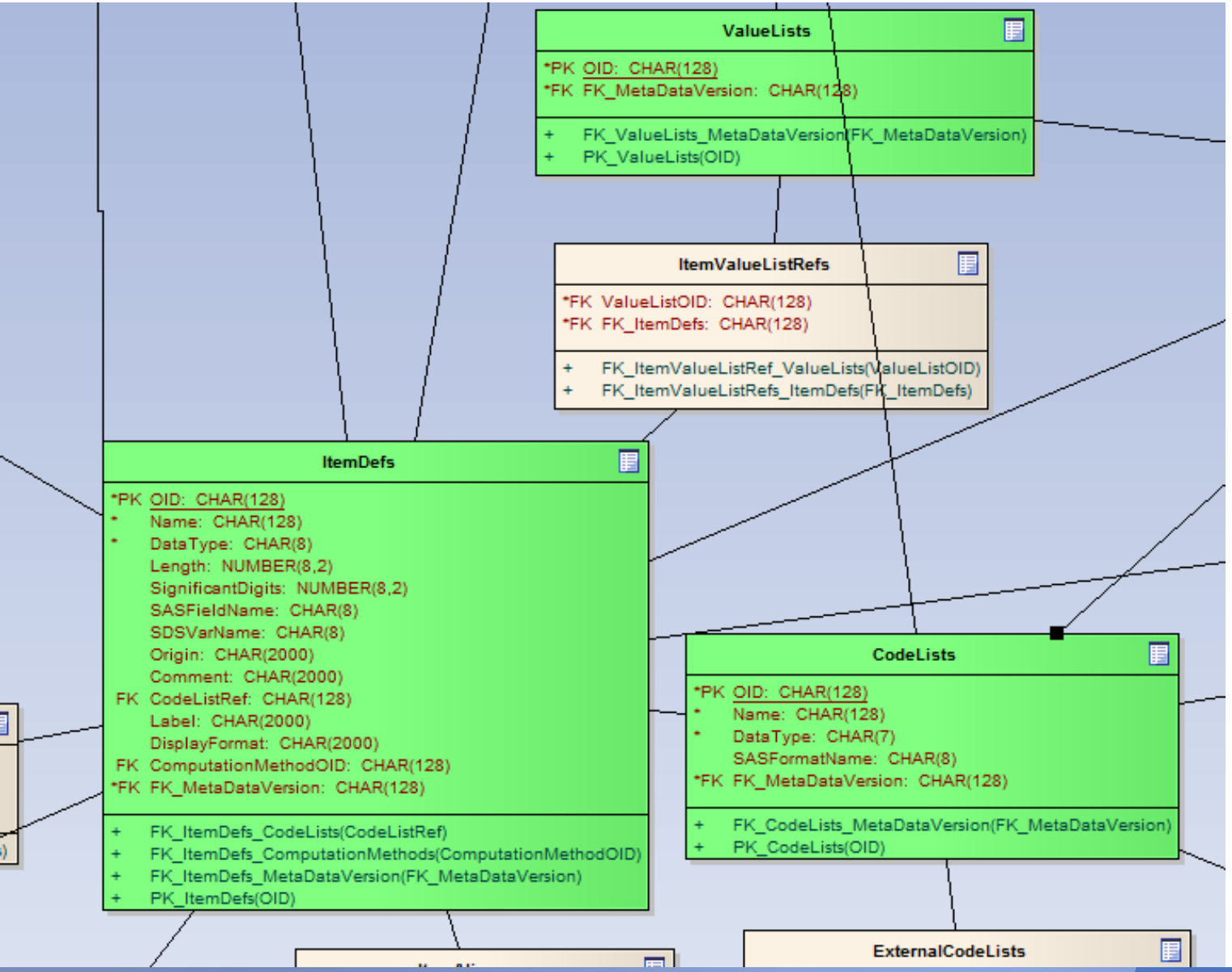
VIEWTABLE: _sywork.Itemdefs

	OID	Name	DataType	Length	CodeListRef	label
10	COL10	BRTHDTC	text	64		Date/Time of Birth
11	COL11	AGE	float	8		Age
12	COL12	AGEU	text	10	AGEU	Age Units

The CRT-DDS SAS Data Model



The CRT-DDS SAS Data Model



What is the SAS Clinical Standards Toolkit (CST)?

What is the SAS Clinical Standards Toolkit?

- Framework to primarily support clinical research activities.
- Initially focusing on standards as defined by CDISC, but not limited to CDISC.
- A collection of “tools”, providing an initial set of standards and functionality that is evolving and growing with updates and releases.
- Designed as an integral part of Clinical Data Integration (CDI), but is available to all SAS users as open source SAS Macros.

What is the SAS Clinical Standards Toolkit?

SAS Clinical Standards Toolkit

- CDISC Models
- Compliance checks
- CRT-DDS (define.xml) creation
- CDISC controlled terminology

SAS Data Integration

- Graphical ETL tool
- Visually create mappings from source to standards
- Access most data systems for source and destination data

Value-Added IP

- Manage and administer CDISC standards
- GUI-based Toolkit use
- Industry transformation processes
- Clinical-trial 'awareness'

SAS Clinical Data Integration

What is the SAS Clinical Standards Toolkit?

- Can modify CST macros, but know what you are doing!
- CDISC models available 3.1.1 and 3.1.2
- Controlled Terminology 2008-08, 2010-03
- CRT-DDS and define.xml capability
- Completely customizable to fit various standards
- Free with 9.1.3 and 9.2

- Toolkit 1.4: ADaM support, ODM 1.3 import/export, more Terminology support

CST Framework

CST Framework

Consists of two distinct pieces (default installation):

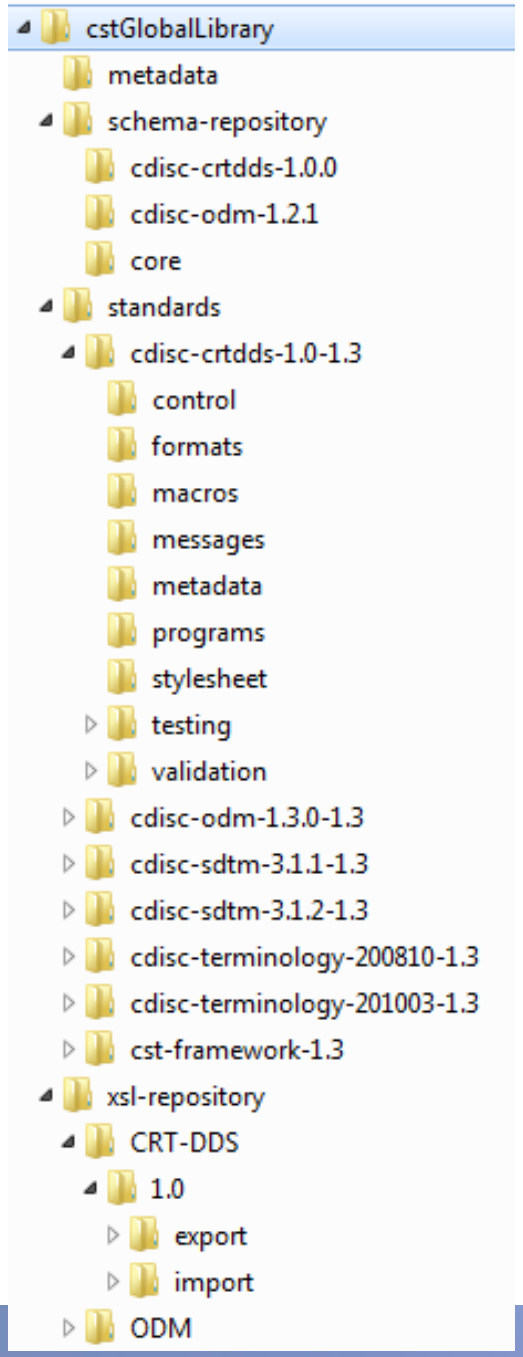
- The components that are installed as part of SAS Foundation and shared files. (SAS Macros, Java JAR files, etc.)
  - » *SASROOT:\ClinicalStandardsToolkitFramework\...*
  - » *SASROOT:\SASClinicalStandardsToolkitSDTM312\...*
  - » *SASROOT:\SASFoundation\9.2\cstframework\sasmacro*

- The global standards library that the framework macros operate on top of.
  - » *C:\cstGlobalLibrary\...*

CST Framework

The Global Standards Library is created during installation. A series of directories are created here:

- **/metadata :** contains data sets that have information about the registered standard versions.
- **/schema-repository :** contains schemas for XML-based standards that are supported
- **/standards :** contains directories for each of the supported standards.
- **/xsl-repository :** contains directories and XSL files used in reading and writing XML files.



CST Framework

The Global Standards Library

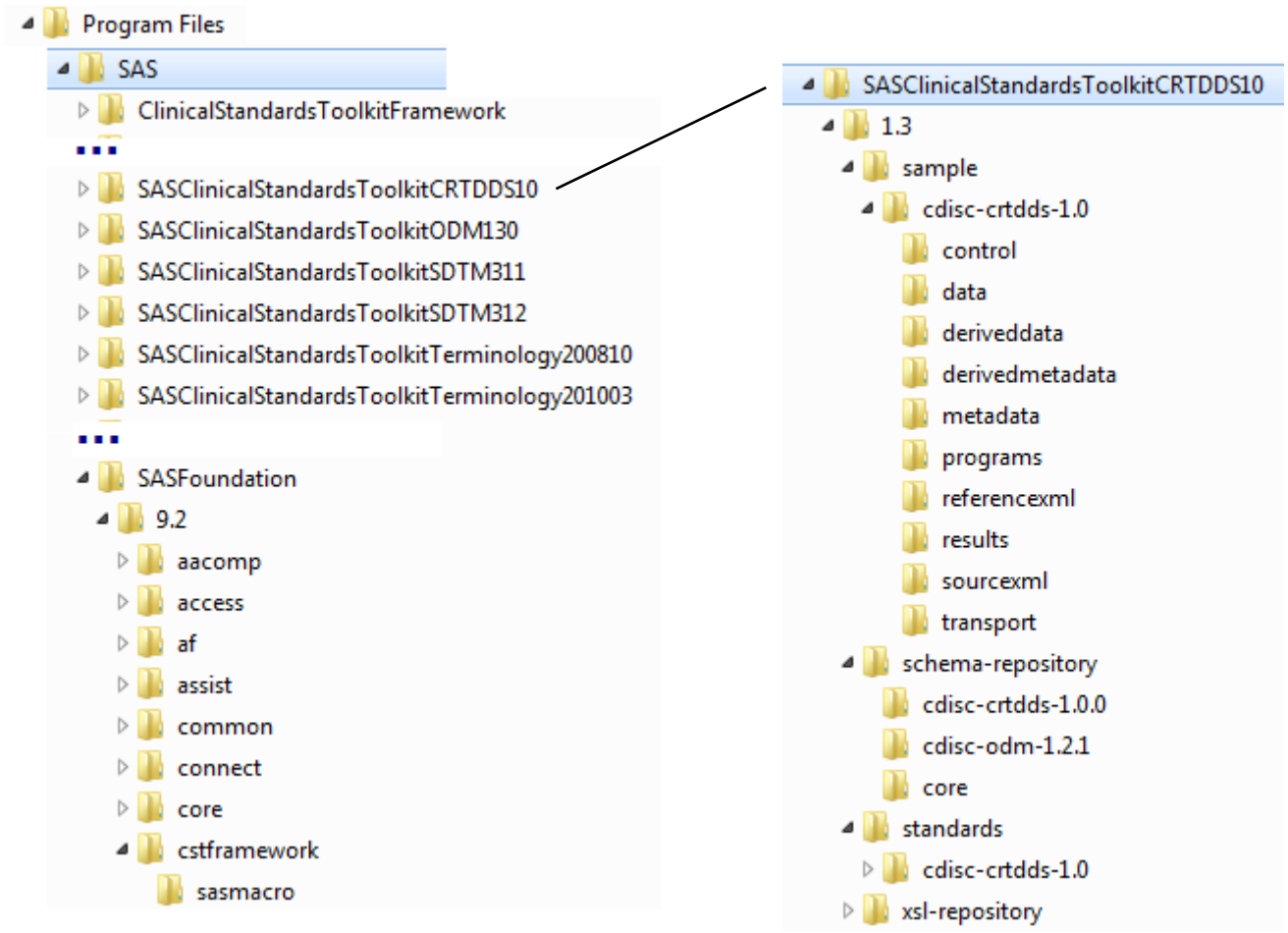
Example of default installation of CST Global Library installed on C: drive.

Custom standards go here.

This is Toolkit 1.3

CST Framework

SASROOT (read-only …)



CST Metadata Files

SAS Clinical Standards Toolkit Metadata

Toolkit provides and uses a series of metadata files to support basic core functions of the toolkit.

Standards

StandardSASReferences

Standardlookup

SASReferences

Properties

Messages

Validation_Master (Validation_Control)

Reference_Tables (Source_Tables)

Reference_Columns (Source_Columns)

Validation_Metrics

Validation_Results

SAS Clinical Standards Toolkit Metadata

The **SASReferences** data set is the “brain center” of the Toolkit. Used to tell Toolkit where things are located. Can be created by the user or generated through the Toolkit. Many users will create their own SASReferences data set.

	standard	standardversion	type	subtype	SASref	reftype
1	CDISC-CRTDDS	1.0	autocall		crtcode	fileref
2	CDISC-CRTDDS	1.0	control	reference	control	libref
3	CDISC-CRTDDS	1.0	externalxml	xml	crtxml	fileref
4	CDISC-CRTDDS	1.0	messages		crtmsg	libref
5	CDISC-CRTDDS	1.0	properties	initialize	inprop	fileref
6	CDISC-CRTDDS	1.0	referencemetadata	column	refmeta	libref
7	CDISC-CRTDDS	1.0	referencemetadata	table	refmeta	libref
8	CDISC-CRTDDS	1.0	referencexml	map	crtmap	fileref

	standard	standardversion	type	path	order	memname
1	CDISC-CRTDDS	1.0	autocall		1	
2	CDISC-CRTDDS	1.0	control	!sasroot/../SASClinicalStandardsToolkitCRTDDS10/1.3/sample/cdis...	.	sasreferences.sas7bdat
3	CDISC-CRTDDS	1.0	externalxml	!sasroot/../SASClinicalStandardsToolkitCRTDDS10/1.3/sample/cdis...	1	define.xml
4	CDISC-CRTDDS	1.0	messages		2	
5	CDISC-CRTDDS	1.0	properties		1	initialize.properties
6	CDISC-CRTDDS	1.0	referencemetadata		.	
7	CDISC-CRTDDS	1.0	referencemetadata		.	
8	CDISC-CRTDDS	1.0	referencexml	!sasroot/../SASClinicalStandardsToolkitCRTDDS10/1.3/sample/cdis...	1	define.map

SAS Clinical Standards Toolkit Metadata

Property files (i.e. initialize.properties) set default preferences for each process. Properties are a series of name-value pairs that are translated into global macro variables available for the duration of a Toolkit process. Invoked by the ***cst_setProperties*** macro.

```
_cstDebug=0
_cstDebugOptions=mprint mlogic symbolgen mautolocdisplay
_cst_rc=0
_cst_MsgID=
_cst_MsgParm1=
_cst_MsgParm2=
_cstResultSeq=0
_cstSeqCnt=0
_cstSrcData=
_cstResultFlag=0
_cstResultsDS=work._cstresults
_cstMessages=work._cstmessages
_cstReallocateSASRefs=0
_cstFMTLibraries=work
_cstMessageOrder=APPEND
_cstSASRefsLoc=
_cstSASRefsName=
_cstSASRefs=work._cstsasrefs
```

SAS Clinical Standards Toolkit Metadata

Messages data sets are used to store information about the framework and standards validation checks. There are framework messages (CST prefix) and standards validation messages (e.g CRT prefix)

	resultid	standardversion	checksource	sourceid	checkseverity
1	CRT0001	***	XMLTRANSFORM	CRT0001	Info
2	CRT0002	***	XMLTRANSFORM	CRT0002	Warning
3	CRT0003	***	XMLTRANSFORM	CRT0003	Error
4	CRT0004	***	SAS	CRT0004	Error
5	CRT0005	***	SAS	CRT0005	Error
6	CRT0006	***	SAS	CRT0006	Error

sourcedescription	messagetext	parameter1
Message not received from XMLTransform	&_cstResultParm1	Message not retrieved from XMLTransform
Message not received from XMLTransform	&_cstResultParm1	Message not retrieved from XMLTransform
Message not received from XMLTransform	&_cstResultParm1	Message not retrieved from XMLTransform
	The following information was missing from the SASReferences file: &_cstParm1	Information Missing
	&_cstParam1 was not assigned for &_cstParam2 in the SASReferences file	
	The parameter &_cstParam1 had an invalid value &_cstParam2	

SAS Clinical Standards Toolkit Metadata

Results data set. Each Toolkit process generates a results data set that can optionally be persisted beyond the SAS session based on SASReferences data set settings. Each results data set captures the outcome of specific process actions, using the ***messages*** data sets to standardize output.

	resultid	checkid	resultseq	seqno	srcdata	message
22	CST0002	CRT0100	9	1	CSTCHECK_NOTUNIQUE	No tables evaluated-check validation control data set
23	CST0002	CRT0100	10	1	CSTCHECK_NOTUNIQUE	No tables evaluated-check validation control data set
24	CST0100	CRT0100	11	1	SRCDATA.ItemGroupDefs	No errors detected in SRCDATA.ItemGroupDefs
25	CST0100	CRT0100	12	1	SRCDATA.ItemGroupLeaf	No errors detected in SRCDATA.ItemGroupLeaf
26	CRT0100	CRT0100	13	1	SRCDATA.ItemDefs	No two values for the source column can be equivalent within the same source ...
27	CST0002	CRT0100	14	1	CSTCHECK_NOTUNIQUE	No tables evaluated-check validation control data set
28	CST0100	CRT0100	15	1	SRCDATA.CodeLists	No errors detected in SRCDATA.CodeLists
29	CST0100	CRT0100	16	1	SRCDATA.CodeListItems	No errors detected in SRCDATA.CodeListItems

22	CST0002	Warning: Check ...	-1	0	tableScope=FormDefs,columnScope=		TableScope should resolve to at least one data set
23	CST0002	Warning: Check ...	-1	0	tableScope=FormDefArchLayouts,columnScope=		TableScope should resolve to at least one data set
24	CST0100	Info	0	0	keys=OID		
25	CST0100	Info	0	0	keys=ID		
26	CRT0100	Error	1	0	keys=OID	OID=COL16	
27	CST0002	Warning: Check ...	-1	0	tableScope=ItemRangeChecks,columnScope=		TableScope should resolve to at least one data set
28	CST0100	Info	0	0	keys=OID		
29	CST0100	Info	0	0	keys=OID		

SAS Clinical Standards Toolkit Metadata

Validation_Master & Validation_Control

Validation_master contains ALL **standard specific** validation information. **Validation_control** contains **study specific** validation information and is created from validation_master. The “brain center” of the Toolkit validation process.

	checkid	standard	standardversion	checksource	sourceid	checkseverity	checktype	codesource	usesourcemetadata
64	CRT0108	CDISC-CRTDDS	***	SAS	CRT0108	Error	Content	cstcheck_column	Y
65	CRT0108	CDISC-CRTDDS	***	SAS	CRT0108	Error	Content	cstcheck_column	Y
66	CRT0109	CDISC-CRTDDS	***	SAS	CRT0109	Error	Content	cstcheck_column	Y
67	CRT0110	CDISC-CRTDDS	***	SAS	CRT0110	Error	Content	cstcheck_recnotfound	Y
68	CRT0110	CDISC-CRTDDS	***	SAS	CRT0110	Error	Content	cstcheck_recnotfound	Y
69	CRT0110	CDISC-CRTDDS	***	SAS	CRT0110	Error	Content	cstcheck_recnotfound	Y
70	CRT0110	CDISC-CRTDDS	***	SAS	CRT0110	Error	Content	cstcheck_recnotfound	Y

	tablescope	columnscope	codelogic
64	ItemGroupDefs	SASDatasetName	if (&_cstColumn ne "" and (length(&_cstColumn) > 32 or anydigit(subst
65	ItemDefs	SASFieldName SDSVarName	if (&_cstColumn ne "" and (length(&_cstColumn) > 32 or anydigit(subst
66	CodeLists	SASFormatName	if &_cstColumn ne "" and length(&_cstColumn) > 32 then cst_error=1;
67	[Study] [DefineDocument]	[Study.FK_DefineDocument] [DefineDocument.FileOID]	proc sql; create table work._cstProblems as select *, 1 as _cstError fr
68	[MeasurementUnits] [Study]	[MeasurementUnits.FK_Study] [Study.OID]	proc sql; create table work._cstProblems as select *, 1 as _cstError fr
69	[MUTranslatedText] [MeasurementU...	[MUTranslatedText.FK_MeasurementUnits] [Measurem...	proc sql; create table work._cstProblems as select *, 1 as _cstError fr
70	[MetaDataVersion] [Study]	[MetaDataVersion.FK_Study] [Study.OID]	proc sql; create table work._cstProblems as select *, 1 as _cstError fr

SAS Clinical Standards Toolkit Metadata

Reference_Tables & Source_Tables

Reference_tables contains ALL **standard specific** table/ domain information. Source_tables contains **study specific** table/domain information and is created from reference_tables. Determines tables to be validated. For CRT-DDS, information is gathered from the CRT-DDS SAS Data Model.

	SASref	table	label	keys	standard	standard version	standard ref	xmlElementName
1	REFDATA	AnnotatedCRFs			CDISC-CRTDDS	1.0		AnnotatedCRFs
2	REFDATA	CLItemDecodeTranslatedText			CDISC-CRTDDS	1.0		CLItemDecodeTranslatedText
3	REFDATA	CodeListItems		OID	CDISC-CRTDDS	1.0		CodeListItems
4	REFDATA	CodeLists		OID	CDISC-CRTDDS	1.0		CodeLists
5	REFDATA	ComputationMethods		OID	CDISC-CRTDDS	1.0		ComputationMethods
6	REFDATA	DefineDocument		FileOID	CDISC-CRTDDS	1.0		DefineDocument
7	REFDATA	ExternalCodeLists			CDISC-CRTDDS	1.0		ExternalCodeLists
8	REFDATA	FormDefArchLayouts		OID	CDISC-CRTDDS	1.0		FormDefArchLayouts
9	REFDATA	FormDefItemGroupRefs			CDISC-CRTDDS	1.0		FormDefItemGroupRefs
10	REFDATA	FormDefs		OID	CDISC-CRTDDS	1.0		FormDefs
11	REFDATA	ImputationMethods		OID	CDISC-CRTDDS	1.0		ImputationMethods
12	REFDATA	ItemAliases			CDISC-CRTDDS	1.0		ItemAliases
13	REFDATA	ItemDefs		OID	CDISC-CRTDDS	1.0		ItemDefs
14	REFDATA	ItemGroupAliases			CDISC-CRTDDS	1.0		ItemGroupAliases
15	REFDATA	ItemGroupDefItemRefs			CDISC-CRTDDS	1.0		ItemGroupDefItemRefs
16	REFDATA	ItemGroupDefs		OID	CDISC-CRTDDS	1.0		ItemGroupDefs
17	REFDATA	ItemGroupLeaf		ID	CDISC-CRTDDS	1.0		ItemGroupLeaf

SAS Clinical Standards Toolkit Metadata

Reference_Columns & Source_Columns

Reference columns contains all **standard specific** table/domain column information. Source_columns contains **study specific** table/domain column information and is created from reference_columns. Determines columns to be validated. For CRT-DDS, information is gathered from the CRT-DDS SAS Data Model.

	SASref	table	column	label	order	type
52	REFDATA	ItemDefs	OID	Unique identifier for this i...	1	C
53	REFDATA	ItemDefs	Name	Item (variable) name	2	C
54	REFDATA	ItemDefs	DataType	Item (variable) data type...	3	C
55	REFDATA	ItemDefs	Length	Item (variable) length	4	N
56	REFDATA	ItemDefs	SignificantDigits	Number of decimal digits...	5	N

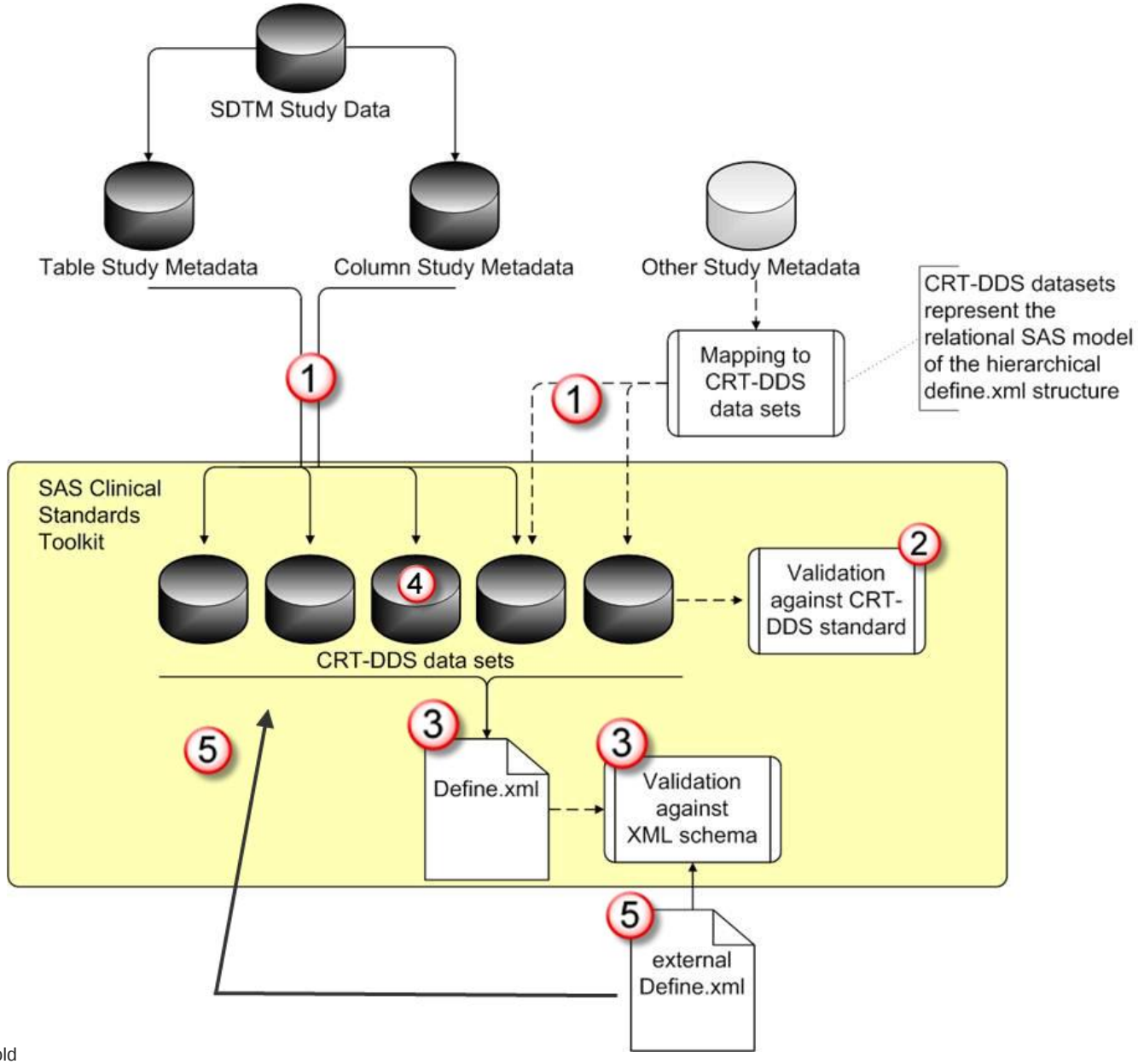
	table	column	length	displayformat	standard	standardversion	core	xmlcodelist
52	ItemDefs	OID	128	$128.	CDISC-CRTDDS	1.0	Req	
53	ItemDefs	Name	128	$128.	CDISC-CRTDDS	1.0	Req	
54	ItemDefs	DataType	8	$8.	CDISC-CRTDDS	1.0	Req	IDTYPE
55	ItemDefs	Length	8	8.	CDISC-CRTDDS	1.0		
56	ItemDefs	SignificantDigits	8	8.	CDISC-CRTDDS	1.0		
57	ItemDefs	SASFieldName	8	$8	CDISC-CRTDDS	1.0		

Hands-on Part

Hands-on

- Create CRT-DDS SAS data sets from the SDTM 3.1.2 metadata
- Validate the CRT-DDS SAS data sets
- Create the define.xml from the CRT-DDS SAS data sets
- Extend the CRT-DDS SAS datasets and create the define.xml
- Import a define.xml file to CRT-DDS SAS data sets

Hands-on



Source: Phuse 2010 presentation Andreas Mangold

Hands-on – CST typical program

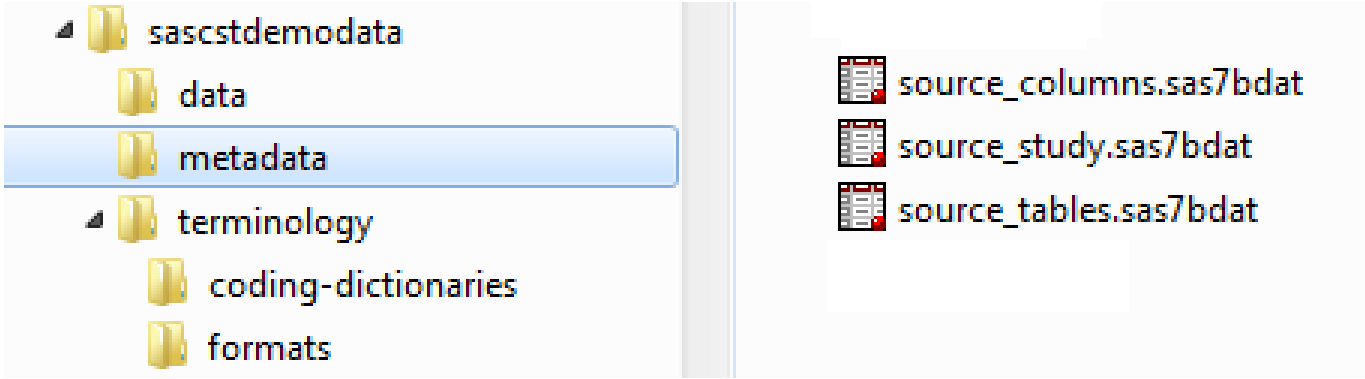
- Define global macro variables ("properties")
  - %LET statements
  - %***cst_setStandardProperties***(_cstStandard=CST-FRAMEWORK,_cstSubType=initialize);
- Define inputs / outputs
  (libnames, filenames, SAS autocall macros, …)
  - 1. Create SASReferences dataset
  - 2. %***cstutil_processsetup***();
    (default: use WORK.SASReferences
- Run process specific macro:
  %***crtdds_sdtm311todefine10***
  %***crtdds_validate***
  %***crtdds_write***
  %***crtdds_xmlvalidate***

Hands-on

01_create_crtdds10_from_sdtm311.sas

- creating the CRTDDS Version 1.0 data sets from the SDTM 3.1.1 model

- INPUT:

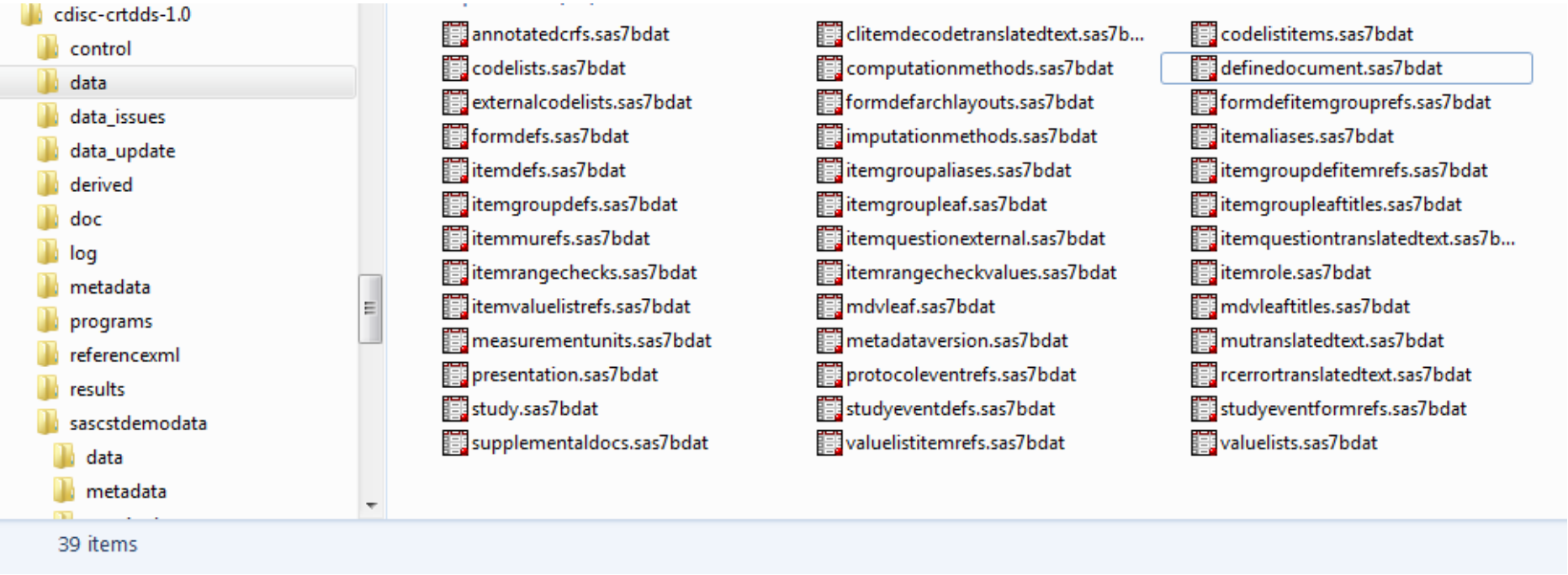


Hands-on

02a_validate_crtdds_data.sas

02b_validate_crtdds_data_issues.sas

- Validation: check compliance of CRT-DDS SAS datasets with CRT-DDS standard



Hands-on

03a_create_crtdds_define.sas

- Create the define.xml from the CRT-DDS SAS data sets

Hands-on

03b_create_crtdds_define_update.sas

- Update the CRT-DDS SAS datasets:
  - » Annotated CRF link
  - » Value level metadata

- Create the define.xml from the CRT-DDS SAS data sets

Hands-on

- Update the CRT-DDS SAS datasets:
  - » Annotated CRF link

- **AnnotatedCRF** – contains document references to the annotated case report form
- **MDVLeaf** – contains the href link for the Annotated CRF
- **MDVLeafTitles** – contains a descriptive title for the Annotated CRF

Hands-on

- Update the CRT-DDS SAS datasets:
  - » Annotated CRF link

```
*Lookup OID for the SDTM 3.1.2 standard in MetaDataVersion;
proc sql noprint;
  select OID into :mdv from srcupd.MetaDataVersion
  where name="CDISC-SDTM 3.1.2";
quit;
*Add records for Annotated CRF;
proc sql;
  insert into srcupd.AnnotatedCRFs
  set DocumentRef = "BlankCRF",
  leafID= "AnnotatedCRF",
  FK_MetaDataVersion = "&mdv";
  insert into srcupd.MDVLeaf
  set ID= "AnnotatedCRF",
  href = "./blankcrf.pdf",
  FK_MetaDataVersion = "&mdv";
  insert into srcupd.MDVLeafTitles
  set title= "Blank Annotated CRF",
  FK_MDVLeaf = "AnnotatedCRF";
quit;
```

Hands-on

- Update the CRT-DDS SAS datasets:
  - » Annotated CRF link

```
<MetaDataVersion OID="1" Name="CDISC-SDTM 3.1.2"
                 Description="CDISC-SDTM 3.1.2"
                 def:DefineVersion="1.2" def:StandardName="CDISC-SDTM"
                 def:StandardVersion="3.1.2">
    <def:AnnotatedCRF>
        <def:DocumentRef leafID="AnnotatedCRF">BlankCRF</def:DocumentRef>
    </def:AnnotatedCRF>
    <def:leaf ID="AnnotatedCRF" xlink:href="./blankcrf.pdf">
        <def:title>Blank Annotated CRF</def:title>
    </def:leaf>
    .....
```

Hands-on

- Update the CRT-DDS SAS datasets:
  - » Value level metadata

SCGRPID	Group ID	text			Identifier	Used to tie together a block of related records in a
SCSPID	Sponsor-Defined Identifier	text			Identifier	Sponsor-defined reference number. Perhaps pre-p database.
SCTESTCD	Subject Characteristic Short Name	text	SCCD		Topic	Short name of the measurement, test, or examinati from a vertical to a horizontal format. The value in (e.g.'1TEST'). SCTESTCD cannot contain charac
SCTEST	Subject Characteristic	text			SynonymQualifier	Verbatim name of the test or examination used to characters. Examples: Subject Initials, Eye Color.
SCCAT	Category for Subject Characteristic	text			GroupingQualifier	Used to define a category of related records.
SCSCAT	Subcategory for Subject	text			GroupingQualifier	A further categorization of the subject characterist

Value Level Metadata

Source Variable	Value	Label	Type	Controlled Terms or Format	Origin	Role	Comment
SCTEST	Initials	Initials	text				Initials
SCTEST	Race	Race	text				Race, Other

Hands-on

- Update the CRT-DDS SAS datasets:
  - » Value level metadata

ValueLists – contains id of value lists

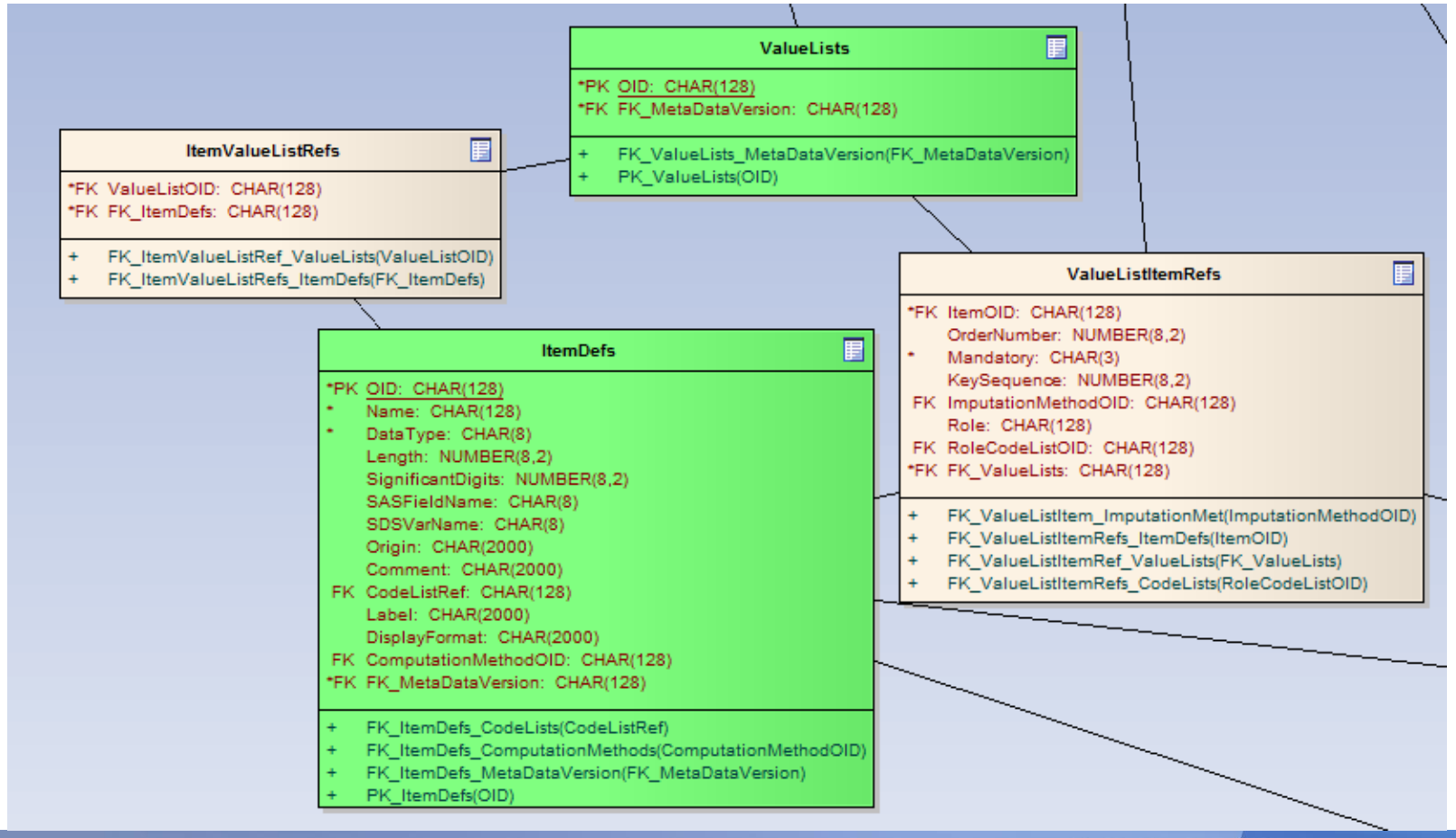
ValueListItemRefs– contains id of each item in a value list

ItemValueListRefs – associates each value list item to a row in the ItemDefs dataset

ItemDefs – contains metadata for each id in ValueListItemRefs

Hands-on

- Update the CRT-DDS SAS datasets:
  - » Value level metadata



Hands-on

- Update the CRT-DDS SAS datasets:
  - » Value level metadata

```
*Lookup OID for the SCTEST column in ItemDefs;
proc sql noprint;
select OID into :srccol from srcupd.ItemDefs
where name='SCTEST';
quit;
*add record for a new valuelist SCTESTVALS;
proc sql ;
insert into srcupd.ValueLists
set OID= "SCTESTVALS",
FK_MetaDataVersion = "&mdv";
*add record associating the value list SCTESTVALS to the OID for SCTEST ItemDefs record;
insert into srcupd.ItemValueListRefs
set ValueListOID= "SCTESTVALS",
FK_ItemDefs = "&srccol";
```

Hands-on

- Update the CRT-DDS SAS datasets:
  - » Value level metadata

```
*add records to the ItemDefs dataset for each value in the SCTESTVAL value list;
insert into srcupd.ItemDefs
set OID= "VAL001",
Name = "Initials",
DataType = "text",
Length = 3,
SASFieldName = "INITIALS",
comment = "Initials",
label="Initials",
FK_MetaDataVersion = "&mdv"
set OID= "VAL002",
Name = "Race",
DataType = "text",
Length = 20,
SASFieldName = "RACEOTH",
comment = "Race, Other",
label="Race",
FK_MetaDataVersion = "&mdv";
```

Hands-on

- Update the CRT-DDS SAS datasets:
  - » Value level metadata

```
*add records associating the value list SCTESTVALS to rows in the ItemDefs dataset;
insert into srcupd.ValueListItemRefs
set ItemOID= "VAL001",
OrderNumber=1,
Mandatory="Yes",
KeySequence=1,
FK_ValueLists = "SCTESTVALS"
set ItemOID= "VAL002",
OrderNumber=2,
Mandatory="Yes",
KeySequence=2,
FK_ValueLists = "SCTESTVALS";
quit;
```

Hands-on

04_create_sascrtdds_fromxml.sas

- Import the define.xml to the CRT-DDS SAS data sets

Questions ?

sas

SAS Institute Inc.
World Headquarters
SAS Campus Drive
Cary, NC 27513 USA
Tel: +1 919 677 8000
Fax: +1 919 677 4444
www.sas.com

Lex Jansen
Software Developer
Health and Life Sciences
Tel: +1 919 531 9860 Ext 19860

E-mail: Lex.Jansen@sas.com

THE
POWER
TO KNOW.

www.sas.com