

# Define.xml - Tips and Techniques for Creating CRT - DDS

---

Julie Maddox  
Mark Lambrecht

SAS Institute Inc.



THE  
POWER  
TO KNOW®

# Agenda

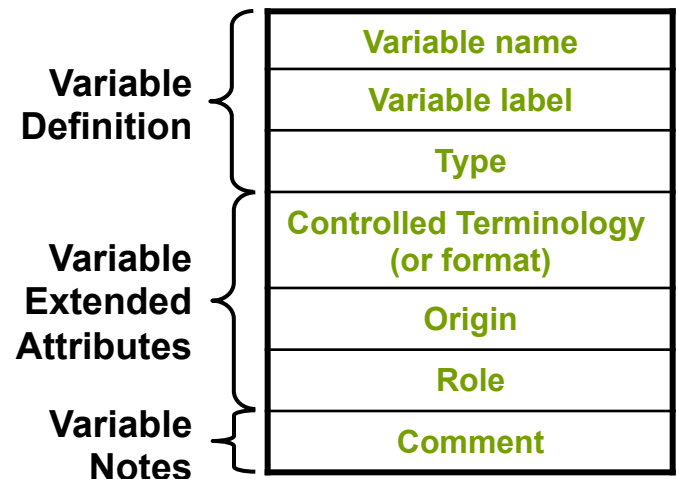
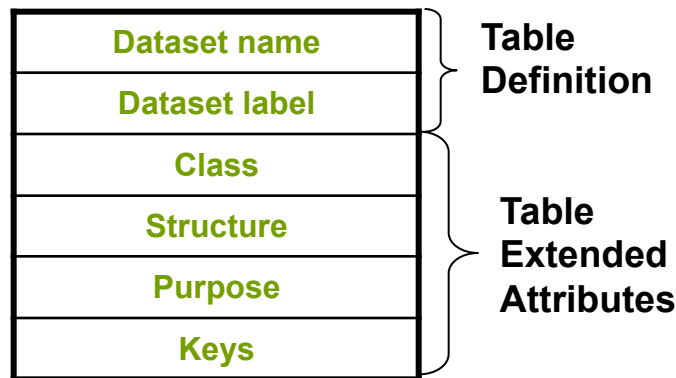
- CDISC standards in SAS : introduction to SAS Clinical Data Integration
- CDISC data and metadata collection in SAS Clinical Data Integration
- Augmenting the define.xml file via the SAS CRT-DDS tables

# Clinical Data Integration functionalities

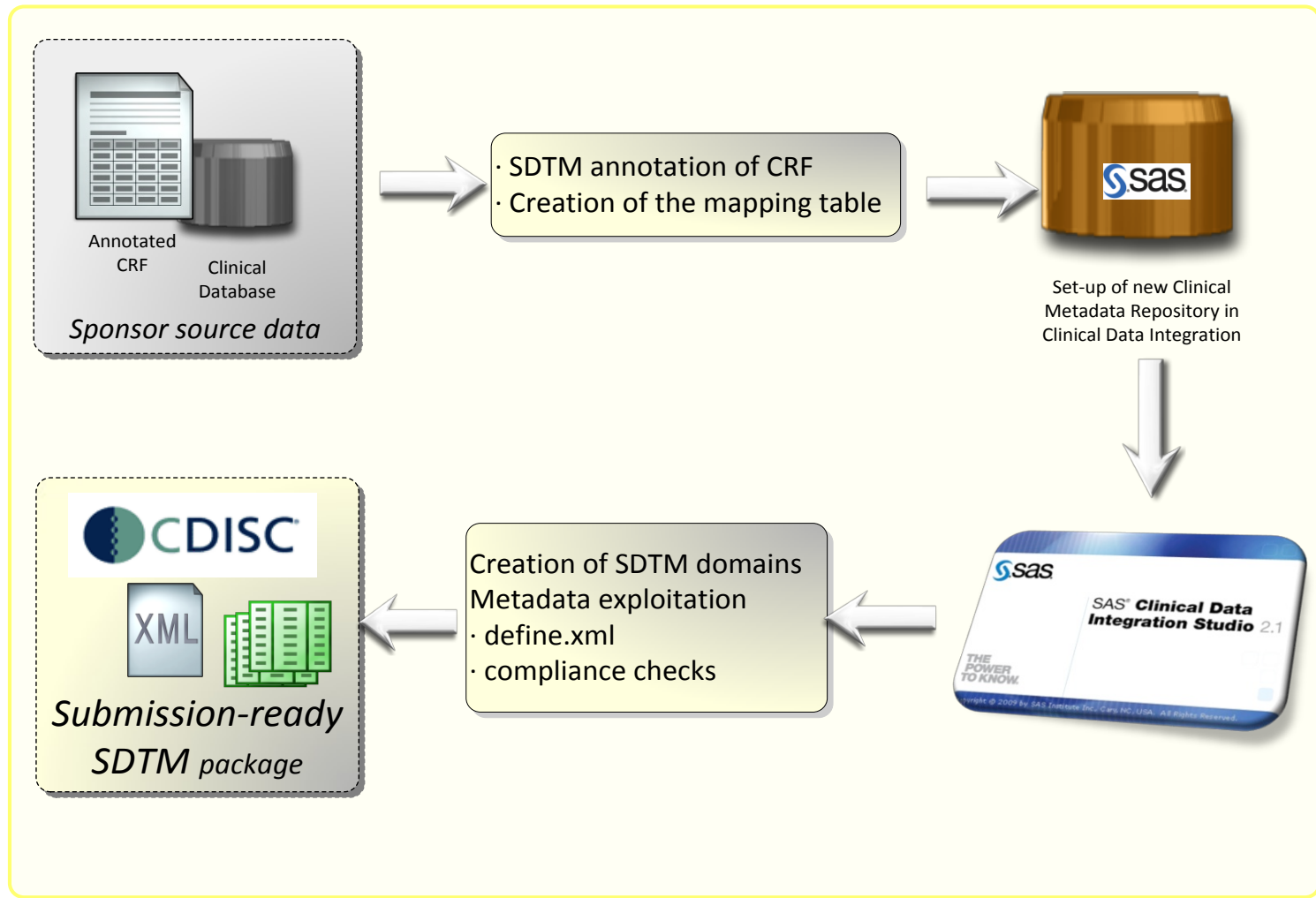
- SAS code-based solution
- Centralized robust metadata repository
- Robust generation of processes that transform and integrate data
- Extensive transformation library to process data
- Included SDTM 3.1.2 IG data model, other standards
- Producing CRT-DDS (define.xml) throughout the process
- Extensible with user-written SAS code transformations

# Types of metadata : CDISC-related and Study-Level metadata

- Structural metadata : data elements
  - Lengths and formats
  - Description and labels of datasets / columns
  - Primary keys
- Codelists
- Value-level metadata
- Computational algorithms



# Process overview



# Metadata Collection, Domains, Columns

**DM Properties**

General | Columns | Indexes | Keys | Parameters | Physical Storage | Options  
Notes | Extended Attributes | Authorization | Clinical Domain

Domain Name:

Data Standard:

Clinical Component:

Properties:

| Name              | Value                   |
|-------------------|-------------------------|
| Archive           | ../transport/dm.xpt     |
| Archive Title     | DM Transport File       |
| Classification    | Special Purpose Domains |
| Domain Identifier | DM                      |
| Is Custom         | false                   |
| Purpose           | Tabulation              |
| Structure         | One record per subject  |

**COUNTRY Properties**

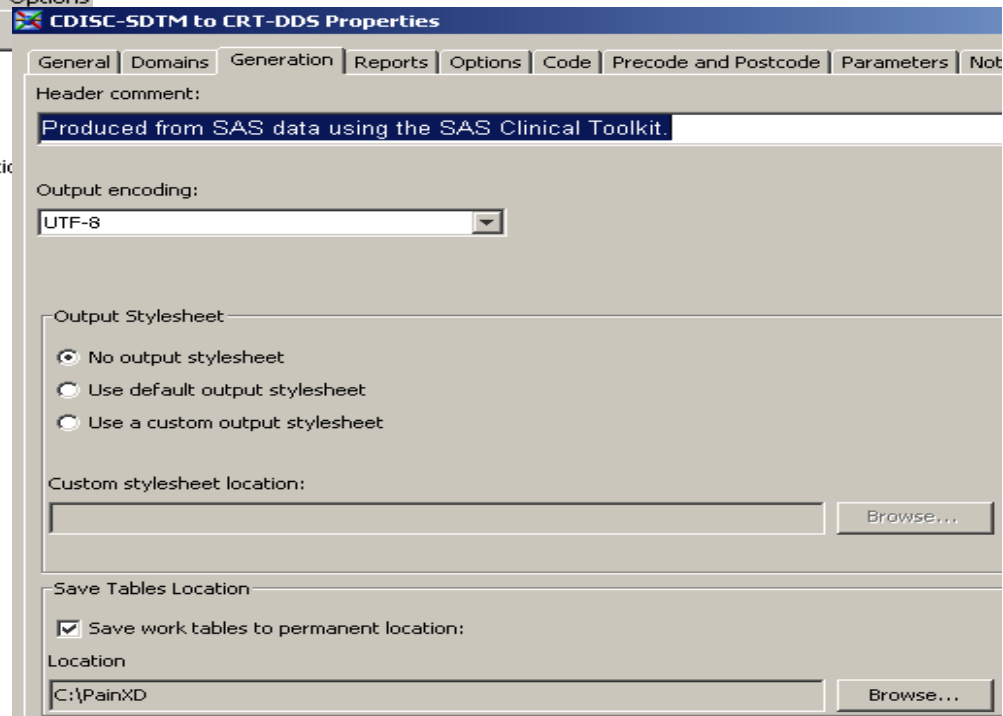
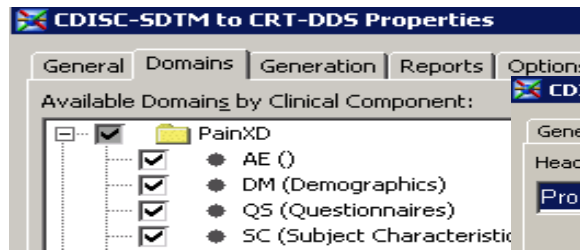
Notes | Extended Attributes | Clinical Column

Label:

Properties:

| Name                 | Value              |
|----------------------|--------------------|
| Contributes to Key   | false              |
| Core                 | Req                |
| Display Format       |                    |
| Method               |                    |
| Origin               |                    |
| Prefix by Identifier | false              |
| Qualifiers           | UPPERCASE          |
| Role                 | RecordQualifier    |
| Term                 | (COUNTRY) ISO 3166 |
| XML Codelist         | COUNTRY            |
| XML Type             | text               |

# Define.xml file transformation process

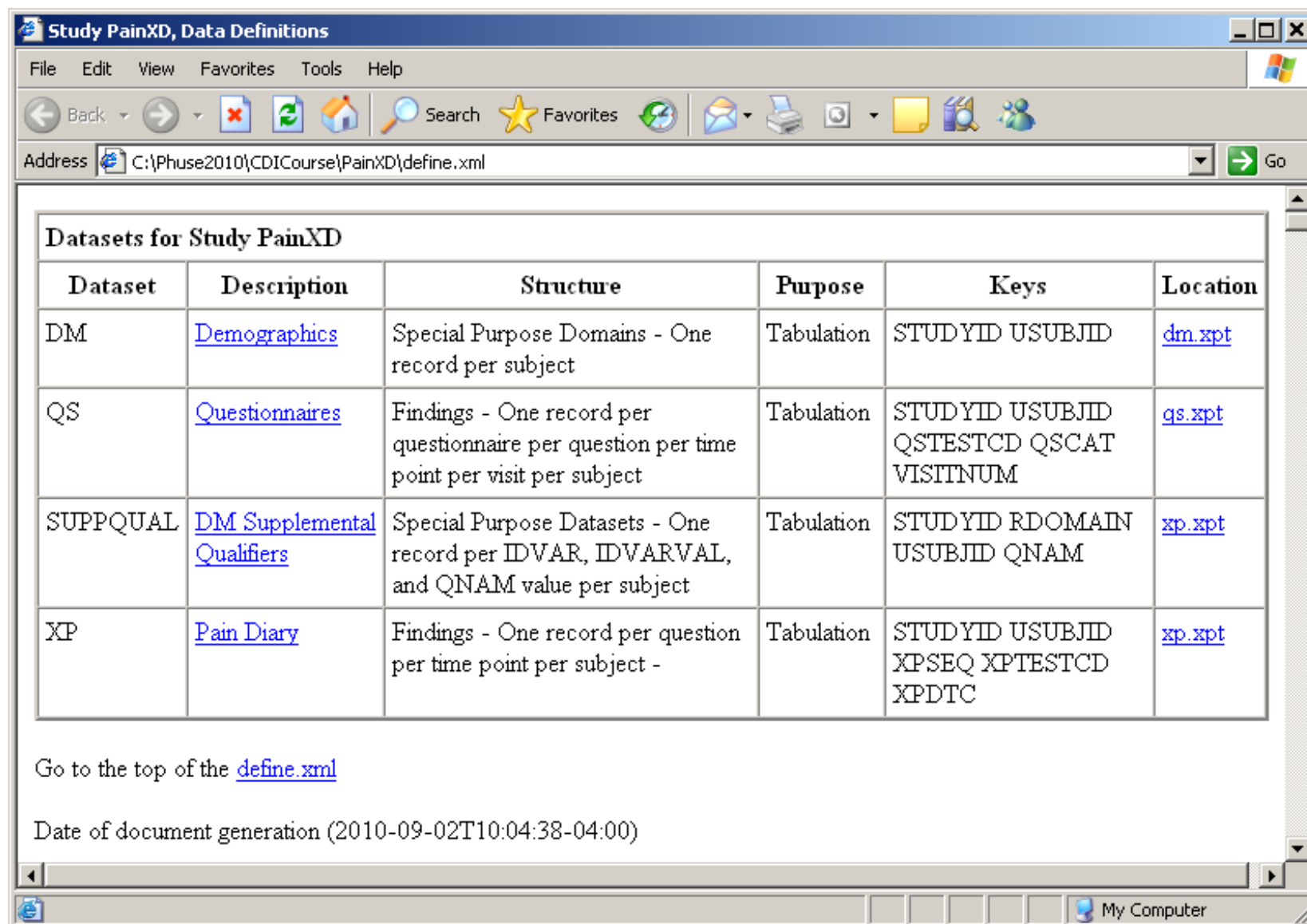


- Select Domains to include
- Select define.xml generation options
  - Customize style sheet
  - Output Encoding
  - Location for CRT-DDS tables

# Clinical Standards Toolkit Code is Generated

```
92  /* Create cstsasreferences table */
93
94  %let _cstSASRefs=work.sasreferences;
95  %let _cstSASRefsLoc=&workpath;
96  %let _cstSASRefsName=sasreferences;
97  %cst_setStandardProperties(_cstStandard=CST-FRAMEWORK, _cstSubType=i
98  %cst_createds(_cstStandard=CST-FRAMEWORK, _cstType=control, _cstSubTyp
99  proc sql;
100  INSERT INTO work.sasreferences
101  136  /* Create Study Staging CRT-DDS Datasets */
102  137  %cst_setStandardProperties(_cstStandard=CDISC-SDTM, _cstStandardVersior
103  138  %cst_createds(_cstStandard=CDISC-SDTM, _cstStandardVersion=3.1.1, _cstTy
104  139  PROC SQL NOPRINT;
140  INSERT INTO work.source_tables
141  VALUES ("1", "AE", "Adverse Events", "Events", "..lae.xpt", "", "One record per
142  433  %let _cstSASRefs=work.sasreferences;
143  434  %let _cstSASRefsLoc=&workpath;
144  435  %let _cstSASRefsName=sasreferences;
145  436  /* Transform CDISC-SDTM domains to CRT-DDS file; */
146  437  %crtdds_sdtm311todefne10(_cstOutLib=work, _cstSourceTables=work.source_tables, _
147  438  %crtdds_write(_cstCreateDisplayStyleSheet=1, _cstResultsOverrideDS=crtdds_results,
148  439  _cstOutputEncoding=UTF-8, _cstHeaderComment=Produced from SAS data using the SA
149  440
150  441
```

# Define.xml



The screenshot shows a SAS web interface window titled "Study PainXD, Data Definitions". The address bar indicates the file path: C:\Phuse2010\CDICourse\PainXD\define.xml. The main content area displays a table titled "Datasets for Study PainXD" with the following data:

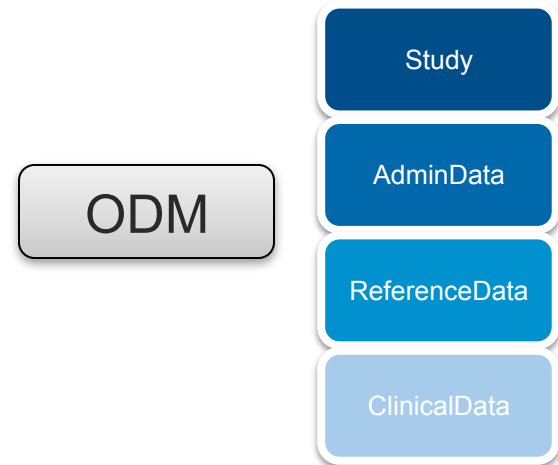
| Dataset  | Description                                | Structure                                                                                 | Purpose    | Keys                                          | Location               |
|----------|--------------------------------------------|-------------------------------------------------------------------------------------------|------------|-----------------------------------------------|------------------------|
| DM       | <a href="#">Demographics</a>               | Special Purpose Domains - One record per subject                                          | Tabulation | STUDYID USUBJID                               | <a href="#">dm.xpt</a> |
| QS       | <a href="#">Questionnaires</a>             | Findings - One record per questionnaire per question per time point per visit per subject | Tabulation | STUDYID USUBJID<br>QSTESTCD QSCAT<br>VISITNUM | <a href="#">qs.xpt</a> |
| SUPPQUAL | <a href="#">DM Supplemental Qualifiers</a> | Special Purpose Datasets - One record per IDVAR, IDVARVAL, and QNAM value per subject     | Tabulation | STUDYID RDOMAIN<br>USUBJID QNAM               | <a href="#">xp.xpt</a> |
| XP       | <a href="#">Pain Diary</a>                 | Findings - One record per question per time point per subject -                           | Tabulation | STUDYID USUBJID<br>XPSEQ XPTESTCD<br>XPDTCT   | <a href="#">xp.xpt</a> |

Go to the top of the [define.xml](#)

Date of document generation (2010-09-02T10:04:38-04:00)

# CRT-DDS SAS DATA MODEL

- SAS provided data model that represents CRT-DDS Version 1.0 format
- Patterned to match the XML element and attribute structure of the define.xml file
  - StudyEvent : patient visit
  - Form : data-entry form
  - ItemGroup : panel or SAS dataset
  - Item : a dataset variable or SAS field
  - Codelist : SAS format or lookup table



# ItemDef, Domain Column metadata

```
<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: <u>sywork.Itemdefs</u> |       |         |          |        |             |                    |
|-----------------------------------|-------|---------|----------|--------|-------------|--------------------|
|                                   | 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          |

# Augmenting the define.xml file

- Additional metadata can be added to the SAS CRT-DDS data sets
- SAS Clinical Standards Toolkit will incorporate the additional data into the define.xml file.

# SAS CRT-DDS tables for Annotated CRF

- **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

# SAS Code to add data for Annotated CRF

```
*Lookup OID for the SDTM 3.1.2 standard in MetaDataVersion;
proc sql noprint;
    select OID into :mdv from _svWork.MetaDataVersion
    where name="CDISC-SDTM 3.1.2";
quit;

*Add records for Annotated CRF;
proc sql;
    insert into _svWork.AnnotatedCRFs
    set DocumentRef = "BlankCRF",
        leafID = "AnnotatedCRF",
        FK_MetaDataVersion = "&mdv";
    insert into _svWork.MDVLeaf
    set ID = "AnnotatedCRF",
        href = "./blankcrf.pdf",
        FK_MetaDataVersion = "&mdv";
    insert into _svWork.MDVLeafTitles
    set title = "Blank Annotated CRF",
        FK_MDVLeaf = "AnnotatedCRF";
quit;
```

# Define.xml with Annotated CRF metadata

```
<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>
  . . . . .
```

# Define.xml with Annotated CRF metadata

## ■ Annotated Case Report Form

| Datasets for Study PainXD |                                         |                                                                                                            |            |                 |                                   |
|---------------------------|-----------------------------------------|------------------------------------------------------------------------------------------------------------|------------|-----------------|-----------------------------------|
| Dataset                   | Description                             | Structure                                                                                                  | Purpose    | Keys            | Location                          |
| DM                        | <a href="#">Demographics</a>            | Special Purpose Domains - One record per subject                                                           | Tabulation | STUDYID USUBJID | <a href="#">DM transport file</a> |
| QS                        | <a href="#">Questionnaires</a>          | Findings - One record per questionnaire per question per time point per visit per subject                  | Tab        |                 |                                   |
| SUPPDM                    | <a href="#">Supplemental Qualifiers</a> | Special Purpose Datasets - One record per IDVAR, IDVARVAL, and QNAM value per subject                      | Tab        |                 |                                   |
| XP                        | <a href="#">Pain Diary</a>              | FINDINGS - One record per finding per object per time point per time point reference per visit per subject | Tab        |                 |                                   |

Blank Annotated CRF ([./blankcrf.pdf](#))

# SAS CRT-DDS tables for 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

# Value Level metadata

| Column | Value  | Label            | Type | Length | Comment                    |
|--------|--------|------------------|------|--------|----------------------------|
| SCTEST | HEIGHT | Height in inches | Text | 10     | Height taken barefoot      |
| SCTEST | WEIGHT | Weight in pounds | Text | 10     | Weight taken without shoes |

- Look up the SCTEST column id so we can associate the value level metadata to the source column

```
proc sql noprint;  
    select OID into :srccol from _syWork.ItemDefs  
    where name='SCTEST';  
quit;
```

# SAS Code for Value Level metadata - ValueLists, ItemValueListRefs

- Add record for new value list SCTESTVALS and associate it to the SCTEST source column

Proc SQL;

```
insert into _svWork.ValueLists
  set OID= "SCTESTVALS",
  FK_MetaDataVersion = "&mdv";
insert into _svWork.ItemValueListRefs
  set ValueListOID= "SCTESTVALS",
  FK_ItemDefs = "&srccol";
```

quit;

# SAS Code for Value Level metadata - ItemDefs

- Add records for each value in Value list to the ItemDefs table

Proc SQL;

insert into \_svWork.ItemDefs

set OID= "VAL001",

Name = "Height",

DataType = "text",

Length = 3,

SASFieldName = "HEIGHT",

comment = "Height taken barefoot",

label="Height in inches",

FK\_MetaDataVersion = "&mdv"

set OID= "VAL002",

Name = "Weight",

DataType = "text",

Length = 4,

SASFieldName = "WEIGHT",

comment = "Weight without shoes",

label="Weight in pounds",

FK\_MetaDataVersion = "&mdv";

# SAS Code for Value Level metadata – ValueListItemRefs

- Add records associating the SCTESTVALS value list to each value item in the ItemDefs table

Proc SQL;

```
insert into _svWork.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";
```

# Define.xml with Value Level Metadata

```
<ItemDef OID="COL58" Name="SCTEST" DataType="text" Length="40" SASFieldName="SCTEST"
    def:Label="Subject Characteristic">
    <def:ValueListRef ValueListOID="SCTESTVALS"/>
</ItemDef>
<ItemDef OID="VAL001" Name="Height" DataType="text" Length="3" SASFieldName="HEIGHT"
    Comment="Height taken barefoot" def:Label="Height in inches"/>
<ItemDef OID="VAL002" Name="Weight" DataType="text" Length="4" SASFieldName="WEIGHT"
    Comment="Weight without shoes" def:Label="Weight in pounds"/>

.....

<def:ValueListDef OID="SCTESTVALS">
    <ItemRef ItemOID="VAL001" OrderNumber="1" Mandatory="Yes" KeySequence="1"/>
    <ItemRef ItemOID="VAL002" OrderNumber="2" Mandatory="Yes" KeySequence="2"/>
</def:ValueListDef>
```

# Define.xml with Value Level Metadata

- Value Level metadata

| Subject Characteristics Dataset (SC) |                            |                      |                            |                  |        |                            |         |        |      |                       |
|--------------------------------------|----------------------------|----------------------|----------------------------|------------------|--------|----------------------------|---------|--------|------|-----------------------|
| Variable                             | Label                      | Type                 | Controlled Terms or Format |                  | Origin | Role                       | Comment |        |      |                       |
| STUDYID                              | Study Identifier           | text                 |                            |                  |        | Identifier                 |         |        |      |                       |
| DOMAIN                               | Domain Abbreviation        | text                 |                            |                  |        | Identifier                 |         |        |      |                       |
| USUBJID                              | Unique Subject Identifier  | text                 |                            |                  |        | Identifier                 |         |        |      |                       |
| SCSEQ                                | Sequence Number            | integer              |                            |                  |        | Identifier                 |         |        |      |                       |
| SCGRPID                              | Group ID                   | text                 |                            |                  |        | Identifier                 |         |        |      |                       |
| SCSPID                               | Sponsor-Defined Identifier | text                 |                            |                  |        | Identifier                 |         |        |      |                       |
| SCTESTCD                             | Subject Name               | Value Level Metadata |                            |                  |        |                            |         |        |      |                       |
| <a href="#">SCTEST</a>               | Subject                    |                      |                            |                  |        |                            |         |        |      |                       |
|                                      |                            | Source Variable      | Value                      | Label            | Type   | Controlled Terms or Format |         | Origin | Role | Comment               |
|                                      |                            | SCTEST               | Height                     | Height in inches | text   |                            |         |        |      | Height taken barefoot |
|                                      |                            | SCTEST               | Weight                     | Weight in pounds | text   |                            |         |        |      | Weight without shoes  |

# Summary

- SAS Clinical Data Integration provides industry-standard capabilities to convert clinical data into CDISC data and produce the CRT-DDS file
- Ongoing central metadata management is key to this built-in process
- The CRT-DDS can be customized at a detailed level using the SAS CRT-DDS tables



Questions ???

[www.sas.com](http://www.sas.com)