

---

# Data Convert

CDISCDEUN CDISC DACH User Network

21 January 2022

**Hanspeter Schnitzer**, Dipl.-Ing.



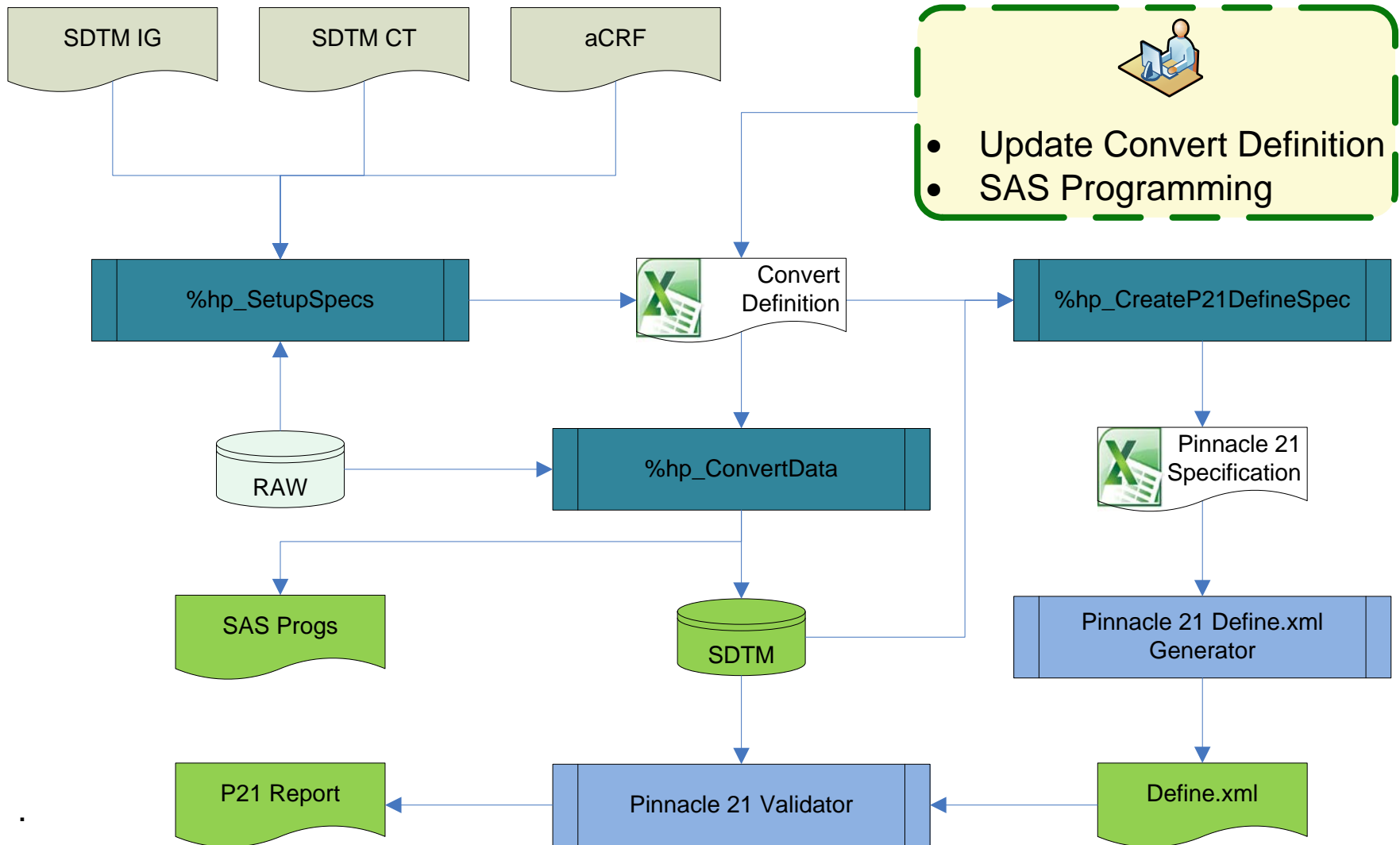
[www.hpschnitzer.de](http://www.hpschnitzer.de)

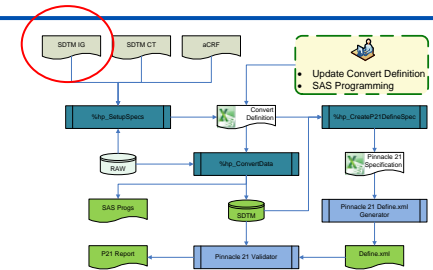


[www.estimondo.com](http://www.estimondo.com)

---

# Overview Raw to SDTM





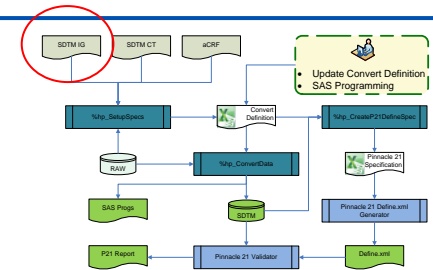
## SDTM IG - Domains Summary:

	A	B	C	D	E	F
1	<b>Dataset</b>	<b>Description</b>	<b>Class</b>	<b>Structure</b>	<b>Purpose</b>	<b>Keys</b>
	AE	Adverse Events	Events	One record per adverse event per subject	Tabulation	STUDYID, USUBJID, AEDECOD, AESTDTC

Define.xml:

### Tabulation Datasets for Study CDISC01 (SDTM-IG 3.1.2)

Dataset	Description	Class	Structure	Purpose	Keys	Location	Documentation
AE	<a href="#">Adverse Events</a>	EVENTS	One record per adverse event per subject	Tabulation	STUDYID, USUBJID, AEDECOD, AESTDTC	<a href="#">ae.xpt</a>	



## SDTM IG - Domains Definition:

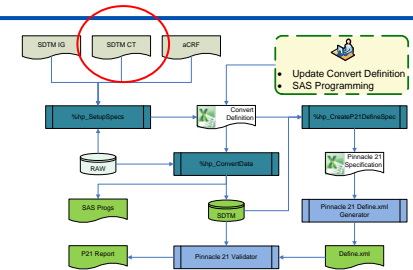
	A	B	C	D	E	F	G
	Variable Name	Variable Label	Type	Controlled Terms, Codelist or Format	Role	CDISC Notes	Core
1							
2	STUDYID	Study Identifier	Char		Identifier	Unique identifier for a study.	Req
3	DOMAIN	Domain Abbreviation	Char	AE	Identifier	Two-character abbreviation for the domain.	Req

## Define.xml:

### Adverse Events (AE) [Location: [ae.xpt](#)]

Variable	Label	Key	Type	Length	Controlled Terms or Format	Origin	Derivation/Comment
STUDYID	Study Identifier	1	text	7		Protocol	
DOMAIN	Domain Abbreviation		text	2	["AE" = "Adverse Events"] <Domain Abbreviation (AE)>	Assigned	
USUBJID	Unique Subject Identifier	2	text	14		Derived	Concatenation of STUDYID and SUBJID

# Controlled Terminology



## Controlled Terminology:

Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value	CDISC Synonym(s)	CDISC Definition	NCI Preferred Term
C66741		Yes	Vital Signs Test Code	VSTESTCD	Vital Signs Test Code	The test code given to the test that analyzes a particular set of vital signs including temperature, respiratory rate, heart rate, and blood pressure.	CDISC SDTM Vital Sign Test Code Terminology
C16358	C66741		Vital Signs Test Code	BMI	Body Mass Index	A general indicator of the body fat an individual is carrying based upon the ratio of weight to height. (NCI)	Body Mass Index
C25299	C66741		Vital Signs Test Code	DIABP	Diastolic Blood Pressure	The minimum blood pressure in the systemic arterial circulation during the cardiac cycle.	Diastolic Blood Pressure
C49680	C66741		Vital Signs Test Code	FRMSIZE	Body Frame Size	The categorization of a person's body frame into small, medium and large based on the measurement of wrist circumference or the breadth of the elbow. (NCI)	Body Frame Size

Define.xml:

Vital Signs Test Code [C66741] [CDISC/NCI SDTM 2011-12-09]	
Permitted Value (Code)	Display Value (Decode)
BMI [C16358]	Body Mass Index
DIABP [C25299]	Diastolic Blood Pressure
FRMSIZE [C49680]	Body Frame Size
HEIGHT [C25347]	Height
PULSE [C49676]	Pulse Rate
SYSBP [C25298]	Systolic Blood Pressure
WEIGHT [C25208]	Weight

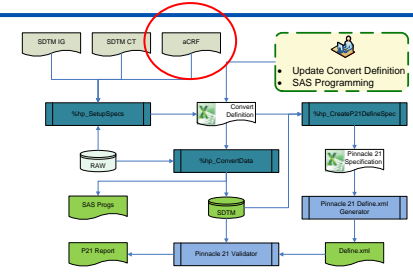
# Annotated CRF

aCRF.pdf

Sex:  Male  Female **SEX**

Ethnic origin:  Hispanic or Latino  Not Hispanic or Latino **ETHNIC**

Race:  American Indian or Alaska Native  Asian  Black or African American  Native Hawaiian or Other Pacific Islander  White  Other **RACE**



Tools Fill & Sign Comment

- Annotations
- Drawing Markups
- Review
- Comments List (771)
  - Find
  - Expand All
  - Collapse All
  - Undock Comment List
  - Print with Comment Summary
  - Create Comment Summary
  - Import Data File...
  - Export Selected to Data File...
  - Export All to Data File...
  - Export to Word...
  - Commenting Preferences...

.fdf file

```

1 0 obj
<</C[1.0 1.0 1.0] /Contents (SEX) /CreationDate (D:20200709110505+02'00') /DA(0 0 1 rg /Cour 12 Tf) /DS(font: Courier, monospace 12.0pt; text-align:left; color:#0000FF) /F 4 /M(D:20200709111006+02'00') /NM(e3502936-9a94-481f-973f-81da4edf1beb) /Page 3 /RC(<?xml version="1.0"?><body xmlns="http://www.w3.org/1999/xhtml" xmlns:xfa="http://www.xfa.org/schema/xfa-data/1.0/" xfa:APIVersion="Acrobat:20.9.0" xfa:spec="2.0.2" style="font-size:12.0pt;text-align:left;color:#00FFFF;font-weight:normal;font-style:normal;font-family:Courier,monospace;font-stretch:normal"><p dir="ltr"><span style="color:#0000FF;font-family:Courier">SEX</span></p></body>) /Rect[357.871 249.744 465.871 268.143] /Subj (Textfeld) /Subtype/FreeText/T (NNNNNN) /Type/Annot>>
endobj
    
```

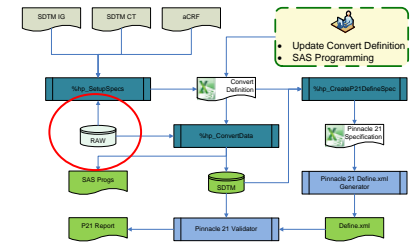
## Demographics (DM) [Location: [dm.xpt](#)]

Variable	Label	Key	Type	Length	Controlled Terms or Format	Origin	Derivation/Comment
SEX	Sex		text	1	["F" = "Female", "M" = "Male", "U" = "Unknown"] <Sex>	CRF Page 4	

# Input Data RAW \*)

Metadata

Column Name	Type	Length	Format
SEX	Text	2000	
BRTHDAT	Text	2000	
ETHNIC	Text	2000	
ETHNIOTH	Text	2000	



Data

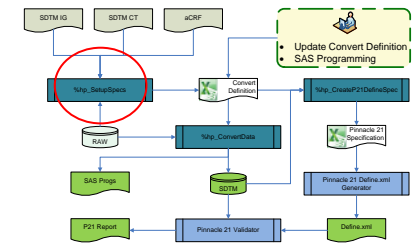
	SEX	BRTHDAT	ETHNIC	ETHNIOTH
1		19591229	7	
2		19820411	10	WHITE-CAUCASIAN/HISPANIC



Convert to Controlled Terminology

\*) Other input formats (CDASH, SDTM, ...) are possible!

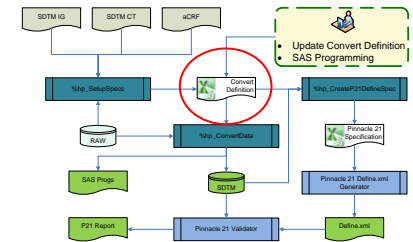
# Create Convert/Define.xml Definition



```
%hp_SetupSpecs(  
  iLibIn           = LIBRAW  
  ,iStandardDS    = AE DM DS EX SUPPAE SUPPDM TA TE TI TS TV  
  ,iCustomDS      = Events - XA XB | Findings - XC XD | Interventions - XE XF  
  ,iSUPPVars      = AE - ACNMOD AETRTEM | DM - RACEOTH  
  ,iFmtVars       = AERAW - AEOUTRAW | DMRAW - ETHNICRAW | CMRAW - CMDOSURAW CMDOSFRQRAW  
  ,iVisitVars     = SVRAW - VISITNUMBER VISITNAME  
  ,iLibMeta       = LIBMETA  
  ,iDSCT          = Ctsdtm20200925  
  ,iDSIG          = Sdtmig32  
  ,iDSCRFFdf     = acrf.fdf  
  ,iTmpDoc        = Specs_Template.xls  
  ,iLibOut        = LIBMETA  
  ,iDocOut        = SDTMSpecs.xls  
);
```

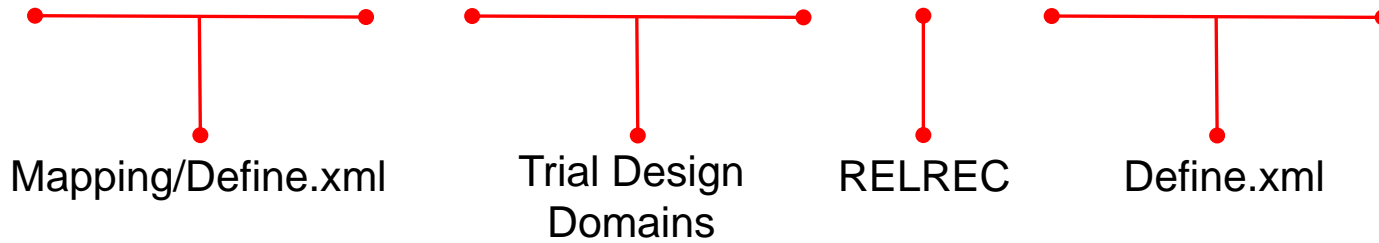


# Convert/Define.xml Definition



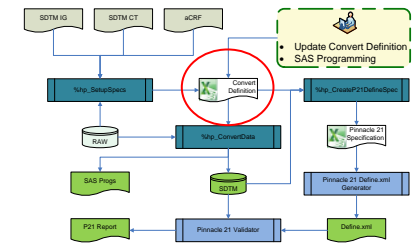
## Convert.xls

Datasets	Variables	Formats	TA	TD	TE	TI	TS	TV	RELREC	Study	Dictionaries	Documents
----------	-----------	---------	----	----	----	----	----	----	--------	-------	--------------	-----------



# Convert/Define.xml Definition (Datasets)

1 Analysis Dataset Metadata									
Dataset	Description	Class	Structure	Purpose	Key variables	Location	Documentation	Source	



After running %hp\_SetupSpecs

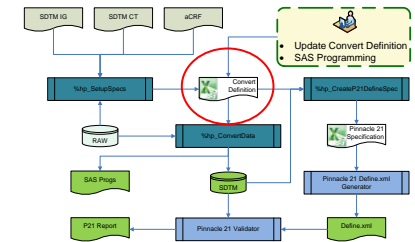
1 Analysis Dataset Metadata									
Dataset	Description	Class	Structure	Purpose	Key variables	Location	Documentation	Source	
AE	Adverse Events	Events	One record per adverse event per subject	Tabulation	STUDYID, USUBJID, AEDECOD, AESTDTC	AE.XPT			
DM	Demographics	Special Purpose	One record per subject	Tabulation	STUDYID, USUBJID	DM.XPT			
DS	Disposition	Events	One record per disposition status or protocol milestone per subject	Tabulation	STUDYID, USUBJID, DSDECOD, DSSTDTC	DS.XPT			
EX	Exposure	Interventions	One record per constant dosing interval per subject	Tabulation	STUDYID, USUBJID, EXTRT, EXSTDTC	EX.XPT			
SUPPAE	Supplemental Qualifiers for AE	Special Purpose	One record per IDVAR, IDVARVAL, and QNAM value per subject	Tabulation	STUDYID, RDOMAIN, USUBJID, IDVAR, IDVARVAL, QNAM	SUPPAE.XPT			
SUPPDM	Supplemental Qualifiers for DM	Special Purpose	One record per IDVAR, IDVARVAL, and QNAM value per subject	Tabulation	STUDYID, RDOMAIN, USUBJID, IDVAR, IDVARVAL, QNAM	SUPPDM.XPT			
TA	Trial Arms	Trial Design	One record per planned Element per Arm	Tabulation	STUDYID, ARMCD, TAETORD	TA.XPT			
TE	Trial Elements	Trial Design	One record per planned Element	Tabulation	STUDYID, ETCOD	TE.XPT			
TI	Trial Inclusion/ Exclusion Criteria	Trial Design	One record per I/E criterion	Tabulation	STUDYID, IETESTCD	TI.XPT			
TS	Trial Summary	Trial Design	One record per trial summary parameter value	Tabulation	STUDYID, TSPARMCD, TSSEQ	TS.XPT			
TV	Trial Visits	Trial Design	One record per planned Visit per Arm	Tabulation	STUDYID, VISITNUM, ARMCD	TV.XPT			

After manually update and before calling %hp\_ConvertData

1 Analysis Dataset Metadata									
Dataset	Description	Class	Structure	Purpose	Key variables	Location	Documentation	Source	
AE	Adverse Events	Events	One record per adverse event per subject	Tabulation	STUDYID, USUBJID, AEDECOD, AESTDTC	AE.XPT		AE2AE_DATA_VIEW_coded_2020-06-12.xlsx AEV5AE_DATA_VIEW_coded_2020-06-12.xlsx AE2AE_DATA_VIEW_coded_2020-06-12.xlsx AEV5AE_DATA_VIEW_coded_2020-06-12.xlsx DM.xlsx DMv5.xlsx DS.xlsx DSv5.xlsx SV.xlsx TA.xlsx TAv5.xlsx	
DM	Demographics	Special Purpose	One record per subject	Tabulation	STUDYID, USUBJID	DM.XPT		DS.xlsx DSv5.xlsx SV.xlsx TA.xlsx TAv5.xlsx EX.xlsx EXV5.xlsx SV.xlsx	
DS	Disposition	Events	One record per disposition status or protocol milestone per subject	Tabulation	STUDYID, USUBJID, DSDECOD, DSSTDTC	DS.XPT		AE2AE_DATA_VIEW_coded_2020-06-12.xlsx AEV5AE_DATA_VIEW_coded_2020-06-12.xlsx DM.xlsx DMv5.xlsx	
EX	Exposure	Interventions	One record per constant dosing interval per subject	Tabulation	STUDYID, USUBJID, EXTRT, EXSTDTC	EX.XPT			
SUPPAE	Supplemental Qualifiers for AE	Special Purpose	One record per IDVAR, IDVARVAL, and QNAM value per subject	Tabulation	STUDYID, RDOMAIN, USUBJID, IDVAR, IDVARVAL, QNAM	SUPPAE.XPT			
SUPPDM	Supplemental Qualifiers for DM	Special Purpose	One record per IDVAR, IDVARVAL, and QNAM value per subject	Tabulation	STUDYID, RDOMAIN, USUBJID, IDVAR, IDVARVAL, QNAM	SUPPDM.XPT			
TA	Trial Arms	Trial Design	One record per planned Element per Arm	Tabulation	STUDYID, ARMCD, TAETORD	TA.XPT			
TE	Trial Elements	Trial Design	One record per planned Element	Tabulation	STUDYID, ETCOD	TE.XPT			
TI	Trial Inclusion/ Exclusion Criteria	Trial Design	One record per I/E criterion	Tabulation	STUDYID, IETESTCD	TI.XPT			
TS	Trial Summary	Trial Design	One record per trial summary parameter value	Tabulation	STUDYID, TSPARMCD, TSSEQ	TS.XPT			
TV	Trial Visits	Trial Design	One record per planned Visit per Arm	Tabulation	STUDYID, VISITNUM, ARMCD	TV.XPT			

# Convert/Define.xml Definition (Variables)

1	Analysis Variable Metadata														
2	Dataset	Variable Name	Variable Label	Type	Controlled Terms	Format	Role	Source / Computational Method	Comment	Origin	Core	DSIn	Data-step	DSFunc	DSOpt
58															
59															
60															
61															
62															
63															
64															
65															
66															



## After running %hp\_SetupSpecs

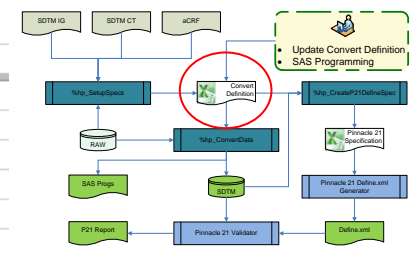
1	Analysis Variable Metadata														
2	Dataset	Variable Name	Variable Label	Type	Controlled Terms	Format	Role	Source / Computational Method	Comment	Origin	Core	DSIn	Data-step	DSFunc	DSOpt
58	DM	STUDYID	Study Identifier	Char			Identifier	= STUDYID		Protocol	Req	_all_			
59	DM	DOMAIN	Domain Abbreviation	Char			Identifier	= "DM"		Assigned	Req	_all_			
60	DM	USUBJID	Unique Subject Identifier	Char			Identifier	= USUBJID		Derived	Req	_all_			
61	DM	SUBJID	Subject Identifier for the Study	Char			Topic	= SUBJID		CRF 1	Req	DM	1	SET	
62	DM	RFSTDTC	Subject Reference Start Date/Time	Char	ISO 8601		Record Qualifier	= RFSTDTC		Assigned	Exp	DM	1		
63	DM	RFENDTC	Subject Reference End Date/Time	Char	ISO 8601		Record Qualifier	= RFENDTC		CRF 4	Exp	DM	1		
64	DM	RFXSTDTC	Date/Time of First Study Treatment	Char	ISO 8601		Record Qualifier	= RFXSTDTC		Assigned	Exp	DM	1		
65	DM	RFXENDTC	Date/Time of Last Study Treatment	Char	ISO 8601		Record Qualifier	= RFXENDTC		Assigned	Exp	DM	1		
66	DM	RFICDTC	Date/Time of Informed Consent	Char	ISO 8601		Record Qualifier	= RFICDTC		Assigned	Exp	DM	1		

## After manually update and before calling %hp\_ConvertData

1	Analysis Variable Metadata														
2	Dataset	Variable Name	Variable Label	Type	Controlled Terms	Format	Role	Source / Computational Method	Comment	Origin	Core	DSIn	Data-step	DSFunc	DSOpt
58	DM	STUDYID	Study Identifier	Char			Identifier	= STUDY		Protocol	Req	_all_			
59	DM	DOMAIN	Domain Abbreviation	Char			Identifier	= "DM"		Assigned	Req	_all_			
60	DM	USUBJID	Unique Subject Identifier	Char			Identifier	= catx("-", STUDY, PT)	Concatenation of	Derived	Req	_all_			
61	DM	SUBJID	Subject Identifier for the Study	Char			Topic	= PT		CRF 1	Req	DM	1	SET	drop = sex
62	DM	RFSTDTC	Subject Reference Start Date/Time	Char	ISO 8601		Record Qualifier	%CreateISO8601(VISDAT, RFSTDTC, 1)	Derived as date of visit 2	Derived	Exp	SV	4	MERGE	where = (VISIT_NUMBER eq "20")
63	DM	RFENDTC	Subject Reference End Date/Time	Char	ISO 8601		Record Qualifier	%LastVisDat	Derived as latest	Derived	Exp	MACRO	9		
64	DM	RFXSTDTC	Date/Time of First Study Treatment	Char	ISO 8601		Record Qualifier	= RFSTDTC	Derived as date of	Derived	Exp	SV	4		
65	DM	RFXENDTC	Date/Time of Last Study Treatment	Char	ISO 8601		Record Qualifier	%LastVisDat(RFXENDTC)	Derived as latest	Derived	Exp	MACRO	7		
66	DM	RFICDTC	Date/Time of Informed Consent	Char	ISO 8601		Record Qualifier	%CreateISO8601(DSSSTDAT, RFICDTC_1)	Derived as date of visit 1	Derived	Exp	DS	5	MERGE	where = (VISIT_NUMBER eq "10")

# Convert/Define.xml Definition (Formats)

1	Analysis Variables Mapping Formats			
2	Format Name	Format Type	Code	Decode
4				
5				
6				
7				
8				
9				
10				
11				



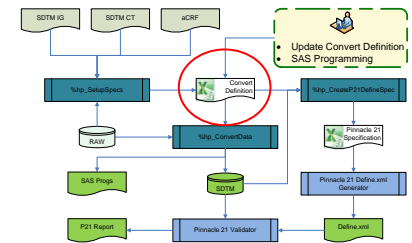
After running %hp\_SetupSpecs

1	Analysis Variables Mapping Formats			
2	Format Name	Format Type	Code	Decode
4	AEOUT_DLV	C	Death	Death
5	AEOUT_DLV	C	Persist	Persist
6	AEOUT_DLV	C	Recovered w/ Sequelae	Recovered w/ Sequelae
7	AEOUT_DLV	C	Recovered w/o Sequelae	Recovered w/o Sequelae
8	AEOUT_DLV	C	Unknown	Unknown
9	ETHNIC_DLV	C	Asia- East Asian Heritage	Asia- East Asian Heritage
10	ETHNIC_DLV	C	Other	Other
11	ETHNIC_DLV	C	Caucasian/ European Heritage	White- Caucasian/ European Heritage

After manually update and before calling %hp\_ConvertData

1	Analysis Variables Mapping Formats			
2	Format Name	Format Type	Code	Decode
4	AEOUT	C	Death	FATAL
5	AEOUT	C	Persist	NOT RECOVERED/NOT RESOLVED
6	AEOUT	C	Recovered w/ Sequelae	RECOVERED/RESOLVED WITH SEQUELAE
7	AEOUT	C	Recovered w/o Sequelae	RECOVERED/RESOLVED
8	AEOUT	C	Unknown	UNKNOWN
9	ETHNIC	C	Asia- East Asian Heritage	ASIAN
10	ETHNIC	C	Other	OTHER
11	ETHNIC	C	Caucasian/ European Heritage	WHITE

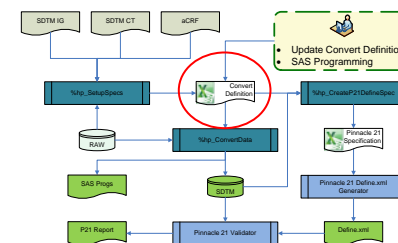
# Convert/Define.xml Definition (T-Domains - TA) estimando The world of estimands



After running %hp\_SetupSpecs

1	Trial Arms (TA)										
2	STUDYID	DOMAIN	ARMCD	ARM	TAETORD	ETCD	ELEMENT	TABRANCH	TATRANS	EPOCH	
4	BS01	TA								BASELINE	
5	BS01	TA								BLINDED TREATMENT	
6	BS01	TA								CONTINUATION TREATMENT	
7	BS01	TA								FOLLOW-UP	
8	BS01	TA								INDUCTION TREATMENT	
9	BS01	TA								LONG-TERM FOLLOW-UP	
10	BS01	TA								OBSERVATION	
11	BS01	TA								OPEN LABEL TREATMENT	
12	BS01	TA								RUN-IN	
13	BS01	TA								SCREENING	
14	BS01	TA								TREATMENT	
15	BS01	TA								WASHOUT	

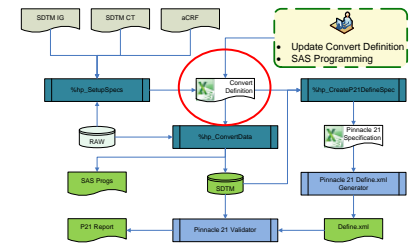
# Convert/Define.xml Definition (T-Domains - TS) estimondo The world of estimands



After running %hp\_SetupSpecs

1	Trial Summary (TS)											
2	STUDYID	DOMAIN	TSSEQ	TSGRPID	TSPARMCD	TSPARM	TSVAL	TSVA1	TSVALNF	TSVALCD	TSVCDREF	TSVCDVER
4	BS01	TS	1		ACTSUB	Actual Number of Subjects						
5	BS01	TS	1		ADAPT	Adaptive Design						
6	BS01	TS	1		ADDON	Added on to Existing Treatments						
7	BS01	TS	1		AGEMAX	Planned Maximum Age of Subjects						

...

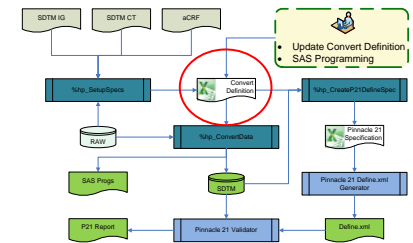


After running %hp\_SetupSpecs

1	Trial Visits (TV)								
2	STUDYID	DOMAIN	VISITNUM	VISIT	VISITDY	ARMCD	ARM	TVSTRL	TVENRL
4	BS01	TV	10	VISIT 1					
5	BS01	TV	20	VISIT 2					
6	BS01	TV	30	VISIT 3					
7	BS01	TV	40	VISIT 4					

...

# Convert/Define.xml Definition (RELREC)



After running %hp\_SetupSpecs

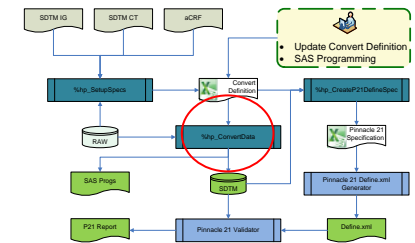
1	Related Records (RELREC)						
2	STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
4	EFC1234	AE	123456	AESEQ	5		1

For RELREC on USUBJID level a SAS macro can be developed and included!





# Convert Data



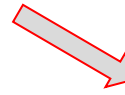
```
%hp_ConvertData(  
  iLibIn      = LIBRAW  
  ,iSpecsDir = LIBMETA  
  ,iSpecsIn  = SDTMSpecs.xls  
  ,iLibOut   = LIBSDTM  
);
```

# Output - Data SDTM \*)

## Data

SEX	RACE	ETHNIC
F	BLACK OR AFRICAN AMERICAN	NOT HISPANIC OR LATINO
M	ASIAN	NOT HISPANIC OR LATINO
M	WHITE	HISPANIC OR LATINO

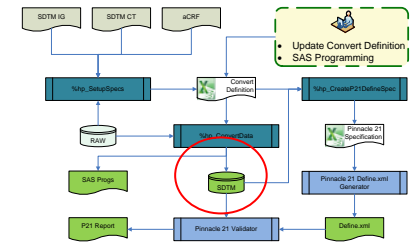
VSTESTCD	VSTEST	VSORRESU
PULSE	Pulse Rate	BEATS/MIN
SYSBP	Systolic Blood Pressure	mmHg
SYSBP	Systolic Blood Pressure	mmHg
SYSBP	Systolic Blood Pressure	mmHg
SYSBP	Systolic Blood Pressure	mmHg
WEIGHT	Weight	LB



### Race [CL.RACE, C74457]

Permitted Value (Code)
WHITE [C41261]
AMERICAN INDIAN OR ALASKA NATIVE [C41259]
BLACK OR AFRICAN AMERICAN [C16352]
ASIAN [C41260]
NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER [C41219]
OTHER [*]
MULTIPLE [*]

\* Extended Value



## Value Level Metadata - VS [VSORRES]



Variable	Where	Type	Length / Display Format	Controlled Terms or Format	Origin	Derivation/Comment
VSORRES	<a href="#">VSTESTCD</a> EQ DIABP (Diastolic Blood Pressure)	integer	2		CRF Page <a href="#">11</a>	
VSORRES	<a href="#">VSTESTCD</a> EQ FRMSIZE (Body Frame Size)	text	6	["LARGE", "MEDIUM", "SMALL"] < <a href="#">Size</a> >	CRF Page <a href="#">11</a>	
VSORRES	<a href="#">VSTESTCD</a> EQ HEIGHT (Height)	float	5.1		CRF Page <a href="#">11</a>	
VSORRES	<a href="#">VSTESTCD</a> EQ PULSE (Pulse Rate)	integer	2		CRF Page <a href="#">11</a>	
VSORRES	<a href="#">VSTESTCD</a> EQ SYSBP (Systolic Blood Pressure)	integer	3		CRF Page <a href="#">11</a>	
VSORRES	<a href="#">VSTESTCD</a> EQ WEIGHT (Weight)	float	5.1		CRF Page <a href="#">11</a>	

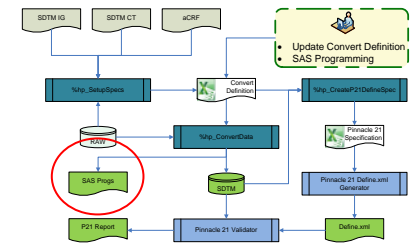
\*) Other output formats (ADaM, ...) are possible!

# Output - SAS Programs

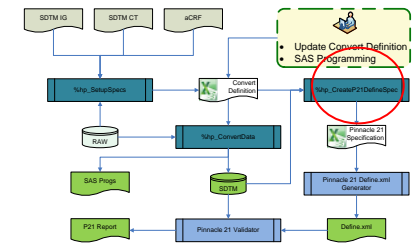
```
*****;
** Module Name:          DM;
** Module Type:         SAS program;
** Topic / Project:     SDTM / xxxx;
** Source File:        ..\DM.sas;
** Software(Dev/Sys):  SAS Version 9.4 / Windows 10;
** -----;
** Short Description:  Create SDTM dataset DM;
** Comments / To Do:  The following libraries must be defined: Raw, SDTM;
** Developer list:    Created by ROSA macro library;
***** HISTORY *****;
** Developer Date    Remarks;
** -----;
** ROSA      20201014  Start of development;
** ROSA      20201121  Implementation finished;
*****;

...

** add data from dataset DEMOG;
data SDTM.dm;
  set SDTM.dm Raw.DEMOG (in=a);
  if a then do;
...
    SEX = upcase(substr(GENDER, 1, 1));
    ETHNIC = ETHORIG;
    RACE = RACEORIG;
...
  end;
run;
```

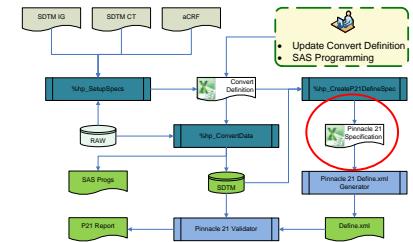


# Create P21 Define Specification



```
%hp_CreateP21DefineSpec (  
  iLibIn      = LIBSDTM  
  ,iSpecsDir  = LIBMETA  
  ,iSpecsIn   = SDTMSpecs.xls  
  ,iTmpDoc    = P21_Template.xls  
  ,iDocDir    = LIBMETA  
  ,iDocOut    = P21DefineSpec.xls  
);
```

# P21 Define Specification

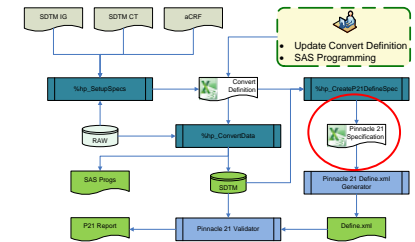


P21DefineSpec.xls

Study	Datasets	Variables	ValueLevel	WhereClauses	Codelists	Dictionaries	Methods	Comments	Documents
-------	----------	-----------	------------	--------------	-----------	--------------	---------	----------	-----------

# P21 Define Specification (Study)

	A	B
1	<b>Attribute</b>	<b>Value</b>
2	StudyName	
3	StudyDescription	
4	ProtocolName	
5	StandardName	
6	StandardVersion	
7	Language	



Study Protocol

SDTM-IG  
ADaM-IG  
SEND-IG

e.g. 3.2

e.g. „en“ or „en-GB“

see Define.XML 2.0, pages 55-58, 73

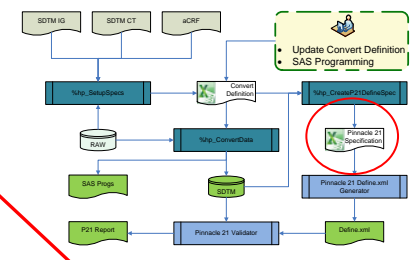
<b>Study Name</b>	CDISC01_1
<b>Study Description</b>	CDISC Test Study Modified to illustrate Define-XML 2.1 features
<b>Protocol Name</b>	CDISC01-1
<b>Metadata Name</b>	Study CDISC01_1, Data Definitions V-1
<b>Metadata Description</b>	Data Definitions for CDISC01-01 SDTM datasets. This metadata version contains only a subset of datasets compared to the prior version. The metadata provided is only intended to illustrate most Define-XML 2.1 new features and not meant to serve as a comprehensive example of all metadata expected to be defined to fully describe an SDTM dataset package. Notes: 1) Supplemental documents released in prior Define-XML versions were not updated to reflect changes to the metadata. 2) This example is prepared for a context other than submission; however, still using the approach of non-standard variables prepared as supplemental qualifiers datasets.

# P21 Define Specification (Datasets)

Dataset	Description	Class	Structure	Purpose	Key Variables	Repeating	Reference Data	Comment
AE	Adverse Events	EVENTS	One record per a Tabulation	STUDYID, USUBJID	Yes	No		

Dataset	Description	Class	Structure	Purpose	Keys*	Location
AE	Adverse Events	Events	One record per adverse event per subject	Tabulation	STUDYID, USUBJID, AEDECOD, AESTDTC	ae.xpt

SDTM IG 3.2



n.a.

Yes  
No

Delimiter: “,”

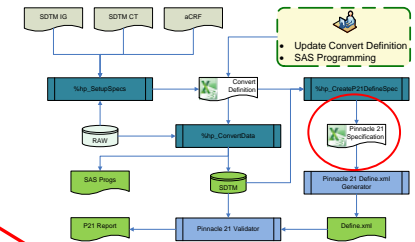
see Define.XML 2.0, pages 64-70

## Tabulation Datasets for Study CDISC01 (SDTM-IG 3.1.2)

Dataset	Description	Class	Structure	Purpose	Keys	Location	Documentation
AE	<a href="#">Adverse Events</a>	EVENTS	One record per adverse event per subject	Tabulation	STUDYID, USUBJID, AEDECOD, AESTDTC	<a href="#">ae.xpt</a>	



# P21 Define Specification (Variables) (1 of 2)



	A	B	C	D	E	F	G	H
1	Order	Dataset	Variable	Label	Data Type	Length	Digits	Format
2	1	AE	STUDYID	Study Identifier	text	11		

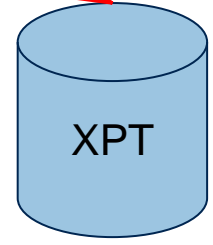
1, 2, ...

CRF

- text
- integer
- float
- datetime
- date
- time
- partialDate
- partialTime
- partialDatetime
- incompleteDatetime
- durationDatetime

Derived

Variable Name	Variable Label	Type	Controlled Terms, Codelist or Format	Role	CDISC Notes	Core
STUDYID	Study Identifier	Char		Identifier	Unique identifier for a study.	Req
DOMAIN	Domain Abbreviation	Char	AE	Identifier	Two-character abbreviation for the domain.	Req



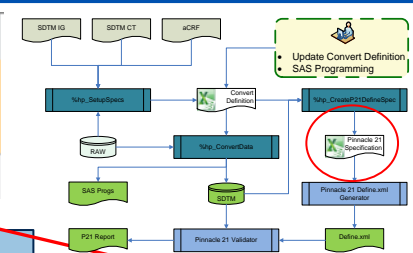
see Define.XML 2.0, pages 20. 60-64, 75-81

## Adverse Events (AE) [Location: [ae.xpt](#)]

Variable	Label	Key	Type	Length	Controlled Terms or Format	Origin	Derivation/Comment
STUDYID	Study Identifier	1	text	7		Protocol	
DOMAIN	Domain Abbreviation		text	2	["AE" = "Adverse Events"] <Domain Abbreviation (AE)>	Assigned	
USUBJID	Unique Subject Identifier	2	text	14		Derived	Concatenation of STUDYID and SUBJID

# P21 Define Specification (Variables) (2 of 2)

	I	J	K	L	M	N	O	P
1	Mandatory	Codelist	Origin	Pages	Method	Predecessor	Role	Comment
2	Yes		Protocol					



CRF  
Derived  
Assigned  
Protocol  
eDT  
Predecessor

CRF

SDTM IG 3.2

Variable Name	Variable Label	Type	Controlled Terms, Codelist or Format	Role	CDISC Notes	Core
STUDYID	Study Identifier	Char		Identifier	Unique identifier for a study.	Req
DOMAIN	Domain Abbreviation	Char	AE	Identifier	Two-character abbreviation for the domain.	Req

n.a.

Convert  
Definition

= <dataset>.<variable>

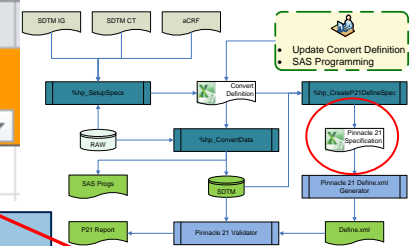
see Define.XML 2.0, pages 20. 60-64, 75-81

## Adverse Events (AE) [Location: [ae.xpt](#)]

Variable	Label	Key	Type	Length	Controlled Terms or Format	Origin	Derivation/Comment
STUDYID	Study Identifier	1	text	7		Protocol	
DOMAIN	Domain Abbreviation		text	2	["AE" = "Adverse Events"] <Domain Abbreviation (AE)>	Assigned	
USUBJID	Unique Subject Identifier	2	text	14		Derived	Concatenation of STUDYID and SUBJID

# P21 Define Specification (ValueLevel) (1 of 2)

	A	B	C	D	E	F	G	H
	<b>Data Significant</b>							
1	Order	Dataset	Variable	Where Clause	Description	Type	Length	Digits
2	2	LB	LBORRES	ALT	Albumin	text	14	



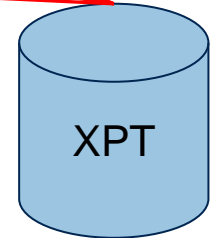
1, 2, ... = Dataset

- text
- integer
- float
- datetime
- date
- time
- partialDate
- partialTime
- partialDatetime
- incompleteDatetime
- durationDatetime

Derived

Variable Name	Variable Label	Type	Controlled Terms, Codelist or Format	Role	CDISC Notes	Core
STUDYID	Study Identifier	Char		Identifier	Unique identifier for a study.	Req
DOMAIN	Domain Abbreviation	Char	AE	Identifier	Two-character abbreviation for the domain.	Req

SDTM IG 3.2



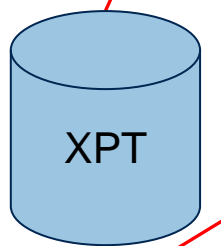
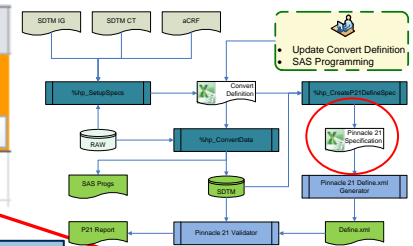
see Define.XML 2.0, pages 63-65, 72-75

## LB (Laboratory Tests Results) - [SDTMIG 3.1.2]

Variable	Where Condition	Label / Description	Type	Length or Display Format	Controlled Terms or ISO Format	Origin / Source / Method / Comment
LBORRES <a href="#">VLM</a>		Result or Finding in Original Units	text	8		Origin specified at Value Level Metadata
	<a href="#">LBTESTCD</a> IN ( "BILI" (Bilirubin), "GLUC" (Glucose) ) and <a href="#">LBSPEC</a> = "BLOOD"	Result or Finding in Orig Units - Set 1	float	3		Collected (Source: Vendor) From Central lab (LB.LBNAM NE "LOCAL LAB")

# P21 Define Specification (ValueLevel) (2 of 2)

	I	J	K	L	M	N	O	P
1	Format	Mandatory	Codelist	Origin	Pages	Method	Predecessor	Comment
2		No		CRF	7			



Variable Name	Variable Label	Type	Controlled Terms, Codelist or Format	Role	CDISC Notes	Core
STUDYID	Study Identifier	Char		Identifier	Unique identifier for a study.	Req
DOMAIN	Domain Abbreviation	Char	AE	Identifier	Two-character abbreviation for the domain.	Req

SDTM IG 3.2

n.a.

Convert Definition

= <dataset>.<variable>

CRF

CRF  
Derived  
Assigned  
Protocol  
eDT  
Predecessor

see Define.XML 2.0, pages 63-65, 72-75

LB (Laboratory Tests Results) - [SDTMIG 3.1.2]

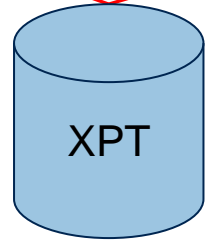
Variable	Where Condition	Label / Description	Type	Length or Display Format	Controlled Terms or ISO Format	Origin / Source / Method / Comment
LBORRES <a href="#">VLM</a>		Result or Finding in Original Units	text	8		Origin specified at Value Level Metadata
	<a href="#">LBTESTCD</a> IN ( "BILI" (Bilirubin), "GLUC" (Glucose) ) and <a href="#">LBSPEC</a> = "BLOOD"	Result or Finding in Orig Units - Set 1	float	3		Collected (Source: Vendor) From Central lab (LB.LBNAM NE "LOCAL LAB")

# P21 Define Specification (WhereClauses)

	A	B	C	D	E
1	ID	Dataset	Variable	Comparator	Value
2	ALT	LB	LBTESTCD	EQ	ALT

= ValueLevel.WhereClause

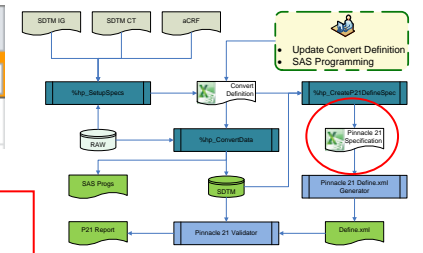
= ValueLevel.Dataset



variable on which the Where clause is based  
(--TESTCD, QNAM, PARAMCD)

- LT
- LE
- GT
- GE
- EQ
- NE
- IN
- NOT IN

see Define.XML 2.0, pages 66-67



## LB (Laboratory Tests Results) - [SDTMIG 3.1.2]

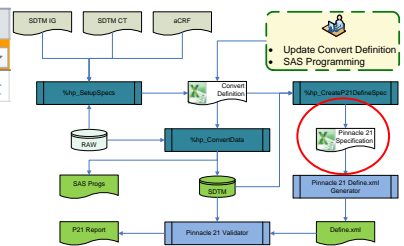
Variable	Where Condition	Label / Description	Type	Length or Display Format	Controlled Terms or ISO Format	Origin / Source / Method / Comment
LBORRES <a href="#">VLM</a>		Result or Finding in Original Units	text	8		Origin specified at Value Level Metadata
	<a href="#">LBTESTCD</a> IN ( "BILI" (Bilirubin), "GLUC" (Glucose) ) and <a href="#">LBSPEC</a> = "BLOOD"	Result or Finding in Orig Units - Set 1	float	3		Collected (Source: Vendor) From Central lab (LB.LBNAM NE "LOCAL LAB")

# P21 Define Specification (Codelists)

ID	Name	NCI Codelist Code	Data Type	Order	Term	NCI Term Code	Decoded Value
LBTRESTCD	Laboratory Test Code	C65047	text	2	ALT	C64433	Alanine Aminotransferase Measurement

= Variables.DataType  
or ValueLevel.DataType

1, 2, ...



Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value	CDISC Synonym(s)	CDISC Definition	NCI Preferred Term
C65047		Yes	Laboratory Test Code	LBTRESTCD	Laboratory Test Code	Terminology used for laboratory test codes of the	CDISC SDTM Laboratory Test Code Terminology
C64433	C65047		Laboratory Test Code	ALT	Alanine Aminotransferase; SGPT	A measurement of the alanine aminotransferase in a biological specimen.	Alanine Aminotransferase Measurement

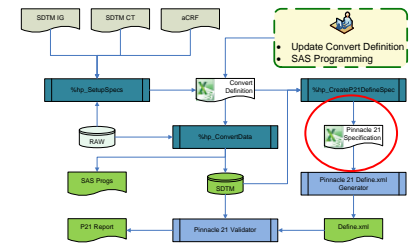
Controlled Terminology

see Define.XML 2.0, pages 74, 82-86

## Laboratory Test Code [C65047] [CDISC/NCI SDTM 2011-12-09]

Permitted Value (Code)	Display Value (Decode)
BILI [C38037]	Bilirubin
BUN [C61019]	Blood Urea Nitrogen

# P21 Define Specification (Dictionaries)



	A	B	C	D	E
1	<b>ID</b>	<b>Name</b>	<b>Data Type</b>	<b>Dictionary</b>	<b>Version</b>
2	MedDRA	Medical Dictionary	Text	MedDRA	22.1

unique ID

free text

text  
integer  
float

MedDRA  
WHODRUG  
...

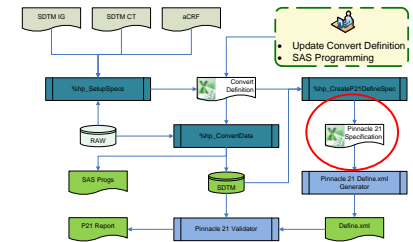
Dictionary

see Define.XML 2.0, pages 82 und 87

## External Dictionaries

Reference Name	External Dictionary	Dictionary Version
Medical Dictionary	MedDRA	22.1

# P21 Define Specification (Methods)



ID	Name	Type	Description	Expression Context	Expression Code	Document	Pages
MC.DM.RFICDTC	MC.DM.RFICDTC	Computation	Derived as date of visit 1 ; input data set : DS ; %CreateISO8601(DSSTDAT, RFICDTC, 1)				

= Variables.Method  
or Variables.Comment  
or ValueLevel.Method  
or ValueLevel.Comment

Computation  
Imputation

n.a.

Convert  
Definition

Document

see Define.XML 2.0, pages 58-60, 72, 88

## Methods

Method	Type	Description
MC.DM.RFICDTC	Computation	Derived as date of visit 1 ; input data set : DS ; %CreateISO8601(DSSTDAT, RFICDTC, 1)



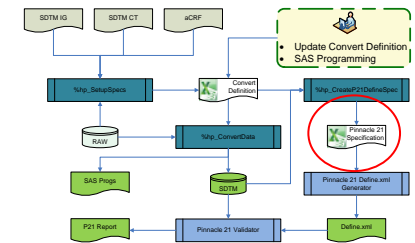
# P21 Define Specification (Documents)

	A	B	C
1	<b>ID</b>	<b>Title</b>	<b>Href</b>
2	ReviewersGuide	Reviewer's Guide	sdrg.pdf

acrf  
ReviewersGuide  
...

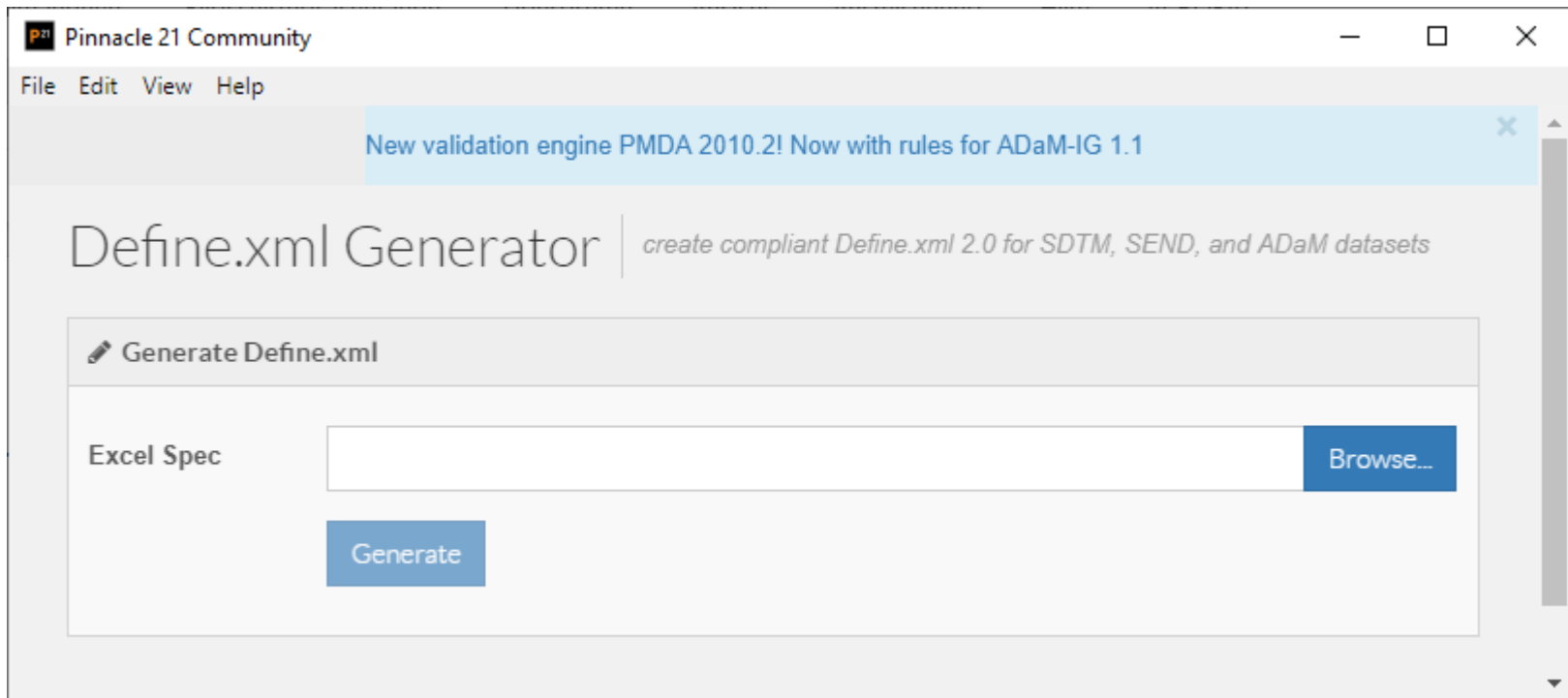
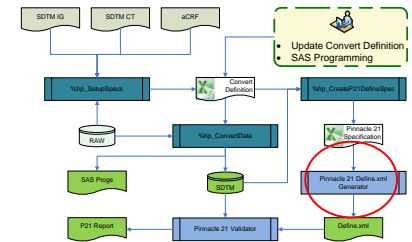
Annotated CRF  
Reviewer's Guide  
...

acrf.pdf  
sdrg.pdf  
adrg.pdf  
...



see Define.XML 2.0, pages 91-92

# Pinnacle 21 Define.xml Generator



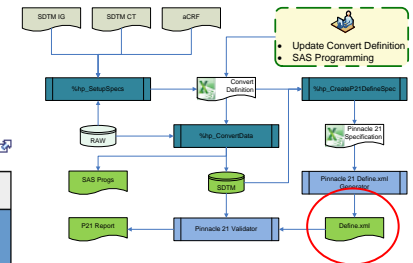
# Output – Define.xml

## DM (Demographics) - [SDTMIG 3.1.2]

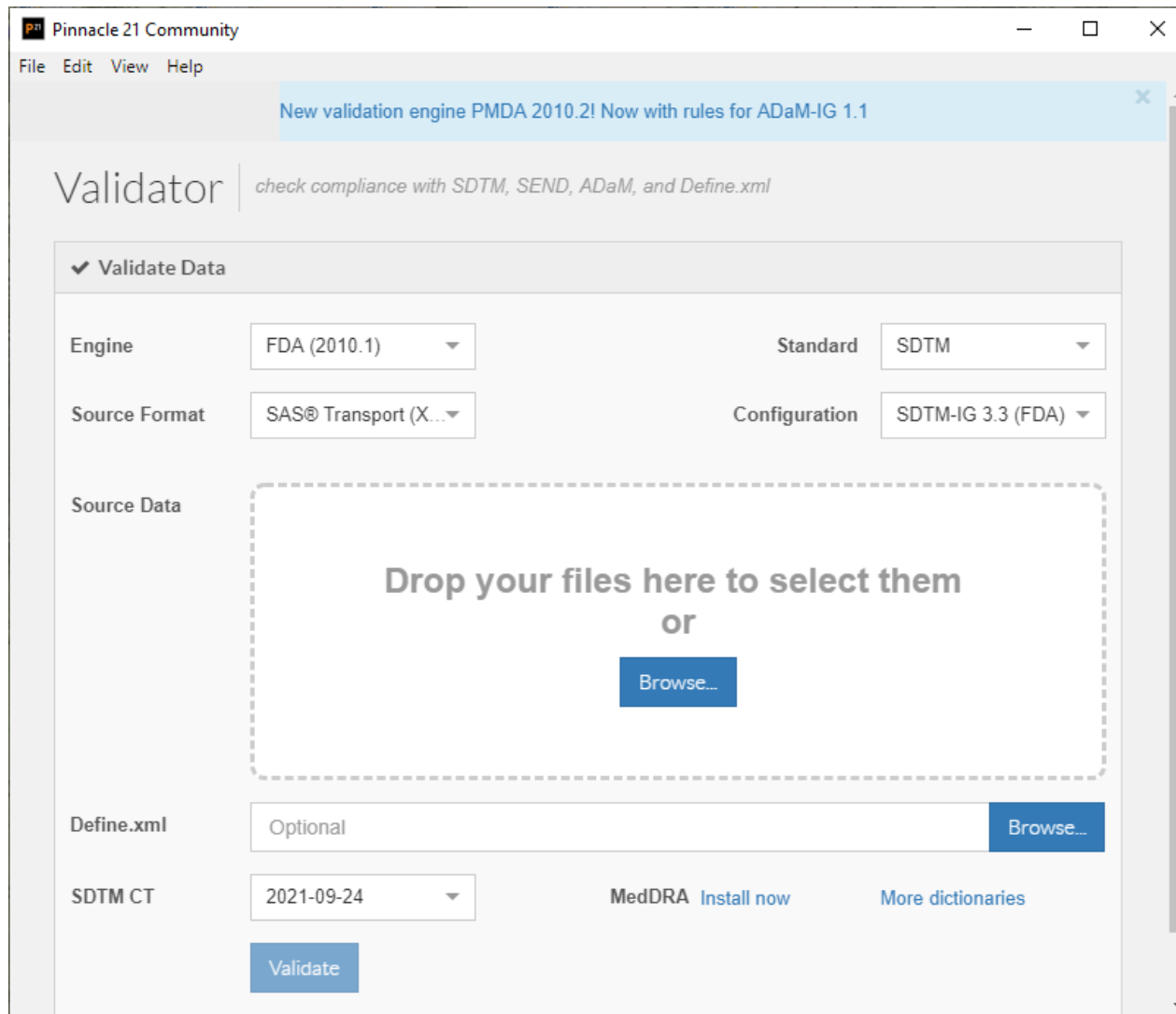
Location: [dm.xpt](#)

Related Supplemental Qualifiers Dataset: [SUPPDM](#) (Supplemental Qualifiers for DM)

Variable	Label / Description	Type	Length or Display Format	Controlled Terms or ISO Format	Origin / Source / Method / Comment
STUDYID	Study Identifier	text	7		Protocol (Source: Sponsor)
DOMAIN	Domain Abbreviation	text	2	<a href="#">Domain Abbreviation (DM)</a> • "DM" = "Demographics"	Assigned (Source: Sponsor)
USUBJID	Unique Subject Identifier	text	14		Derived (Source: Sponsor) Concatenation of STUDYID and SUBJID
SUBJID	Subject Identifier for the Study	text	6		Collected (Source: Investigator) Annotated CRF [3]
RFSTDTC	Subject Reference Start Date/Time	date		ISO 8601	Derived (Source: Sponsor) RFSTDTC = first date/time of study drug, for safety subject. Null for screen failures.
RFENDTC	Subject Reference End Date/Time	date		ISO 8601	Derived (Source: Sponsor) RFENDTC = termination date, for safety subjects. Null for screen failures.
SITEID	Study Site Identifier	text	3		Collected (Source: Investigator) Annotated CRF [3]
BRTHDTC	Date/Time of Birth	date		ISO 8601	Collected (Source: Investigator) Annotated CRF [6]
AGE	Age	integer	2		Derived (Source: Sponsor) Age at Screening Date (Screening Date - Birth date).  For the complete algorithm see the referenced external document. Complex Algorithms [DM]
AGEU	Age Units	text	5		Assigned (Source: Sponsor) Defaulted to YEARS



# Pinnacle 21 Validator



Pinnacle 21 Community

File Edit View Help

New validation engine PMDA 2010.2! Now with rules for ADaM-IG 1.1

## Validator

check compliance with SDTM, SEND, ADaM, and Define.xml

✓ Validate Data

Engine: FDA (2010.1) Standard: SDTM

Source Format: SAS® Transport (X... Configuration: SDTM-IG 3.3 (FDA)

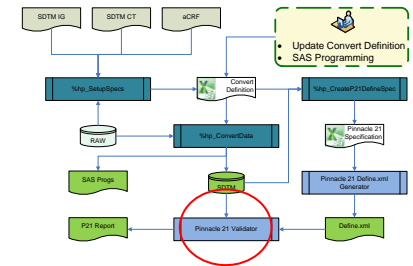
Source Data

Drop your files here to select them  
or  
Browse...

Define.xml: Optional Browse...

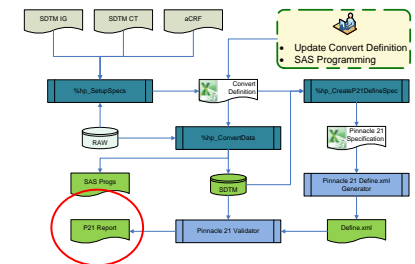
SDTM CT: 2021-09-24 MedDRA [Install now](#) [More dictionaries](#)

Validate



# Output - Pinnacle 21 Report

Pinnacle 21 Validator Report						
Processed Sources						
Domain	Label	Class	Source	Records	Rejects	
GLOBAL	Global Metadata	--	--	0	0	
AE	Adverse Events	EVENTS	ae.xpt	16	0	
CM	Concomitant Medications	INTERVENTIONS	cm.xpt	35	0	
CO	Comments	SPECIAL PURPOSE	co.xpt	4	0	
DA	Drug Accountability	FINDINGS	da.xpt	10	0	
DM	Demographics	SPECIAL PURPOSE	dm.xpt	5	0	
DS	Disposition	EVENTS	ds.xpt	14	0	
EG	ECG	FINDINGS	eg.xpt	56	0	
EX	Exposure	INTERVENTIONS	ex.xpt	17	0	
IE	Inclusion/Exclusion Exceptions	FINDINGS	ie.xpt	1	0	



Pinnacle 21 Validator Report						
Issue Summary						
Source	Pinnacle 21 ID	Message	Severity	Found		
GLOBAL						
	<a href="#">SD1111</a>	Missing SE dataset				1
AE						
	<a href="#">CT2001</a>	AEACN value not found in 'Action Taken with Study Treatment' non-extensible codelist				16
	<a href="#">SD0009</a>	No qualifiers set to 'Y', when AE is Serious				1
	<a href="#">SD0080</a>	AE start date is after the latest Disposition date				1
	<a href="#">SD1077</a>	Regulatory Expected variable EPOCH not found				1
	<a href="#">SD1089</a>	AESTDY variable value is imputed				1
	<a href="#">SD1097</a>	No Treatment Emergent info for Adverse Event				1

# Questions?



**Schnitzer**

Hanspeter Schnitzer Dipl.-Ing.  
SAS Programming

www.hpschnitzer.de      mail@hpschnitzer.de



**estimondo**  
The world of estimands.

**Hanspeter Schnitzer**  
Co-founder & Senior Principal Statistical Programmer

hanspeter.schnitzer@estimondo.com  
+49 177 745 9759

www.estimondo.com  
Estimondo GmbH • Hintergasse 22 • D-61231 Bad Nauheim • Germany