

Introduction to SAS[®] Clinical Standards Toolkit

Tutorial

PhUSE 2010 – October 20th – Paper TU06

Agenda

- Introduction
- Background of the SAS Clinical Standards Toolkit
- Software architecture, system requirements, installation
- Validation of study data against the SDTM Standard
 - Simple example
 - More complete example
- Generation of define.xml
 - Simple example
 - Extended example
- Further Steps

Company

- HMS Analytical Software is a specialist for Information Technology in the field of Data Analysis and Business Intelligence Systems
- Profile
 - 40 employees in Heidelberg, Germany
 - SAS Institute Silver Consulting Partner for 14 years
 - Doing data analysis software projects for more than 20 years
- Technologies
 - Analytics and Data Management:
SAS, JMP, R, Microsoft SQL Server
 - Application Development: Microsoft .NET, Java

Services – see our booth in exhibition area

- Consulting for the application of software tools
- Validation, Auditing, SOP definition and training for analytic software application environments
- Custom software development
- Migration of software systems to new versions
- Outsourcing of data management, data analysis and CDISC-conversion
- Contracting
- Training
 - Own curriculum: Validation, Clinical Standards Toolkit
 - SAS curriculum
- Support

Background of the SAS Clinical Standards Toolkit

- Clinical data standards are increasingly used for
 - submission of results of clinical research to the FDA
 - Data interchange between companies
 - Consistent storage of data within companies
- Deep knowledge is necessary
 - about clinical data management
 - about standards and their implementation
 - and cannot be superseded by tools
- But tools are useful for
 - Management of data and metadata
 - Mapping of data to elements of standard models
 - Validation of standard compliance
 - Generation of documentation

SAS Clinical DI

SAS Clinical Standards Toolkit

Software architecture, system requirements, installation

- System Requirements
- Versions and their Support for Standards
- The Global Standards Library
- Directory Structures
- Installation

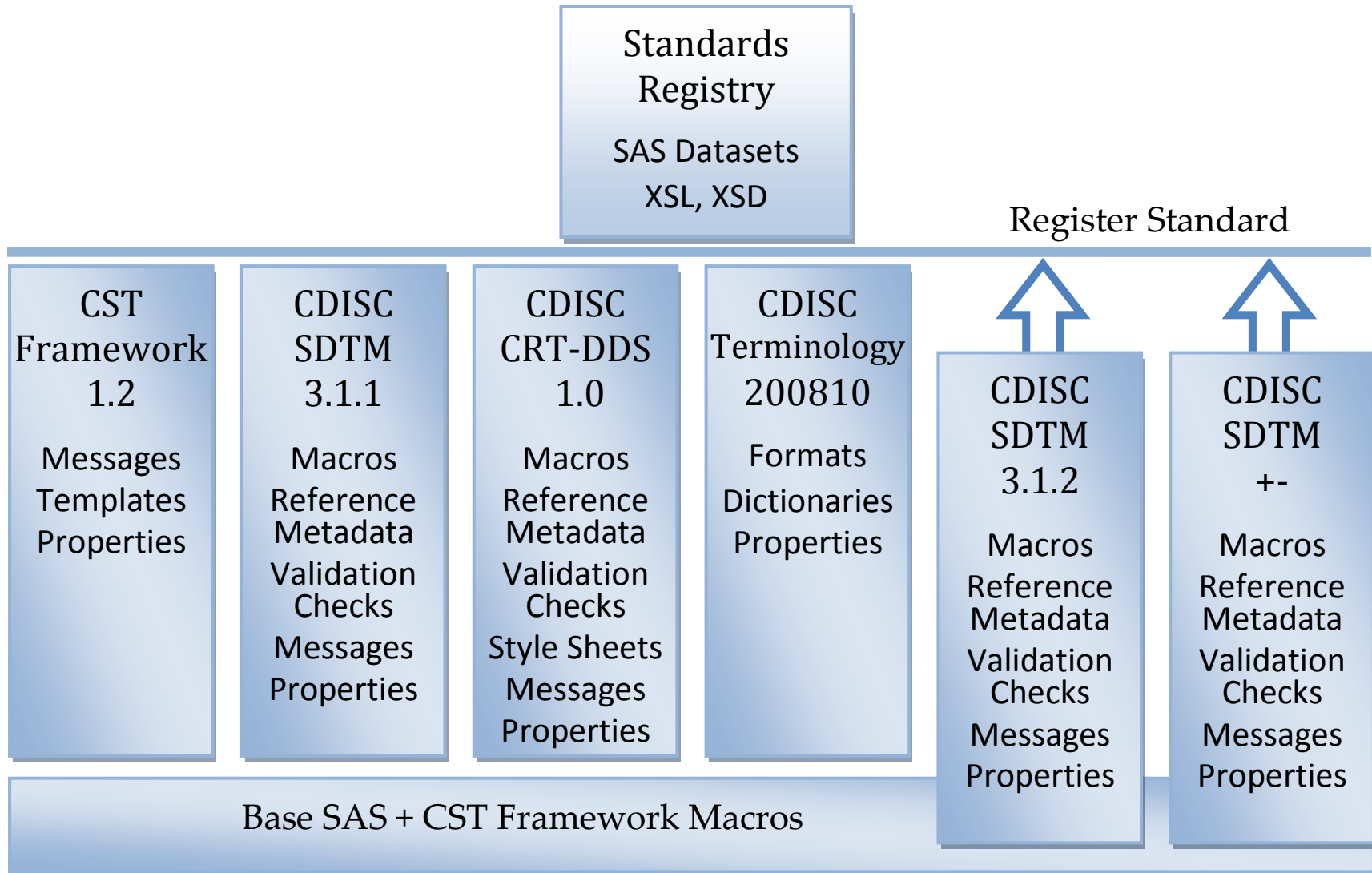
System Requirements

- Available for
 - SAS 9.1.3 on Microsoft Windows
 - SAS 9.2 on Microsoft Windows (not 64 bit) and UNIX
- Requirements
 - SAS: only BASE
 - Java virtual machine for creation and validation of define.xml
- Installation media
 - SAS 9.2: delivered free of charge from SAS Institute
 - SAS 9.1.3: download from the SAS website

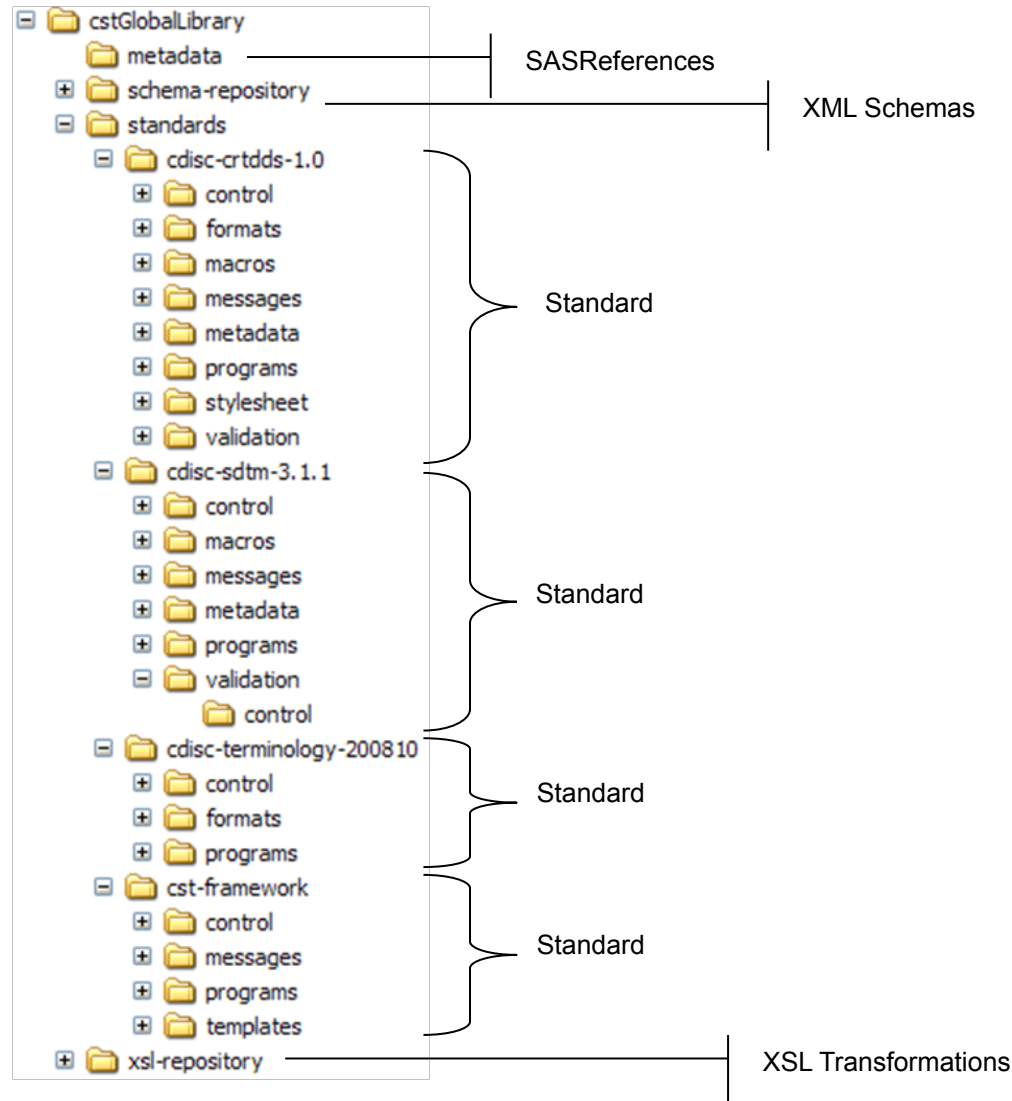
Versions and their Support for Standards

- Current version is 1.2, supporting
 - SDTM 3.1.1
 - CDISC terminology 2008-10
- Preproduction update can be downloaded*
 - SDTM 3.1.2 (other than updated validation checks)
 - CDISC terminology 2010-03
 - Reporting framework
- Version 1.3 is announced for end of the year
 - Full SDTM 3.1.2 and terminology 2010-03 support
 - Reporting framework
- Support for further standard (e.g. ADaM)
 - Has been announced without timeline

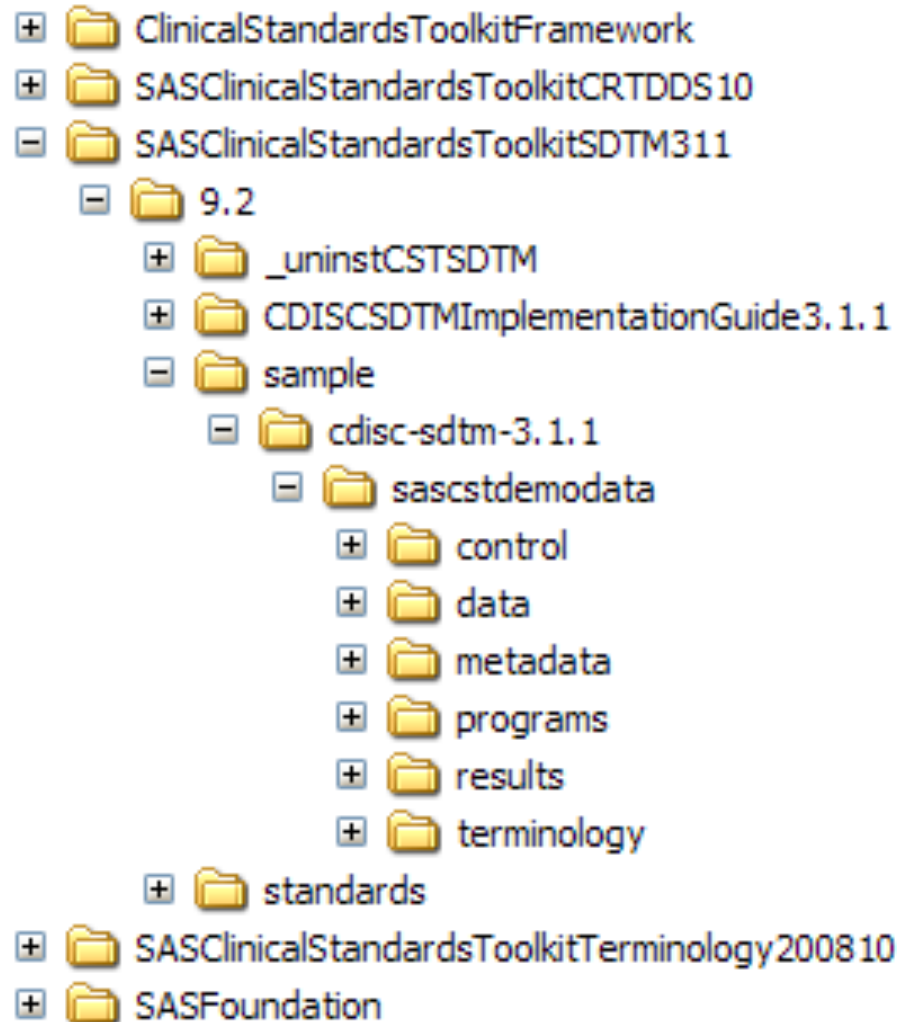
The Global Standards Library



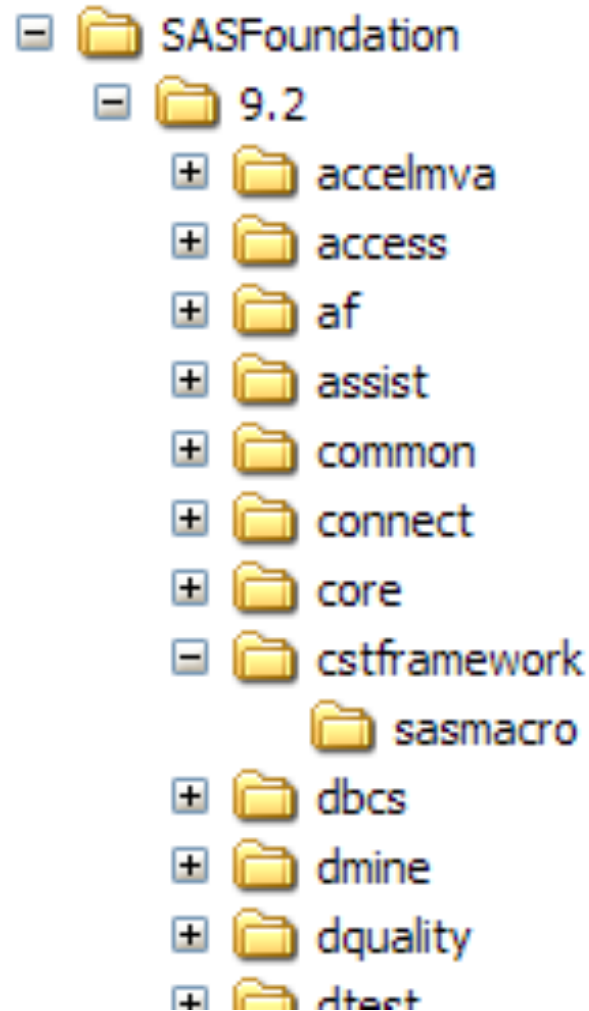
Directory Structure – global standards library



Directory Structure – samples per standard



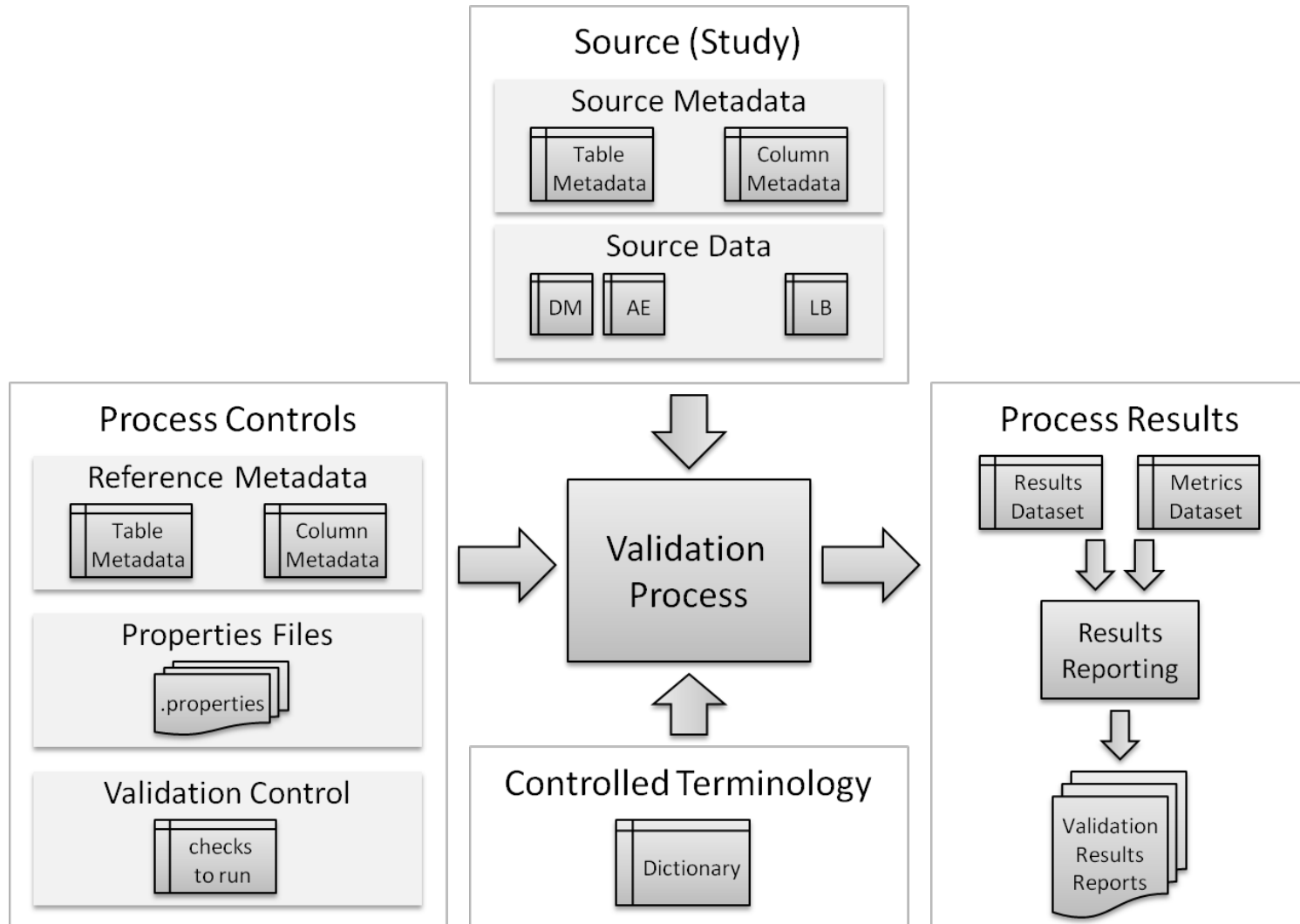
Directory Structure – framework macros



Installation (SAS 9.2)

- Use deployment wizard like for any other SAS product
- SAS Foundation has to be installed together with the toolkit even if it was installed before
- A path to the global standards library has to be provided in the course of the installation process
 - This might be local or shared. In a productive environment, it must be shared and read only.
- After installing the product, an installation qualification procedure should be followed*

Validation of Study Data against the SDTM Standard



Validation of study data against the SDTM standard – simple example

```
/*-- root location of the process input and output --*/  
%let studyRootPath=C:\projects\PhUSE\demo1;  
  
/*-- load basic configuration to macro variables --*/  
%cst_setStandardProperties(  
    _cstStandard=CST-FRAMEWORK  
    ,_cstStandardVersion=1.2  
    ,_cstSubType=initialize);  
%cst_setStandardProperties(  
    _cstStandard=CDISC-SDTM  
    ,_cstStandardVersion=3.1.1  
    ,_cstSubType=initialize);  
  
/*-- make known the existing sasreferences dataset --*/  
%let _cstSASRefsLoc=&studyRootPath\control;  
%let _cstSASRefsName=sasreferences;  
  
/*-- process sasreferences: allocate librefs etc. --*/  
%cstutil_allocatesasreferences;  
  
/*-- run validation, write results and metrics --*/  
%sdtm_validate;
```

Validation of study data against the SDTM standard – results dataset

Result identifier	Validation check id	Seq. no.	Source data	Resolved message text from message file	Result severity
CST0108		1	CST_SETPROPERTIES	The properties were processed from the PATH C: \Programme\SAS\cstGlobalLibrary/standards/cst-framework/programs/initialize.properties	Info
CST0108		1	CST_SETPROPERTIES	The properties were processed from the PATH C: \Programme\SAS\cstGlobalLibrary/standards/cdisc-sdtm-3.1.1/programs/initialize.properties	Info
CST0200		1	SDTM_VALIDATE	PROCESS STANDARD: CDISC-SDTM	Info
CST0200		2	SDTM_VALIDATE	PROCESS STANDARDVERSION: 3.1.1	Info
CST0200		3	SDTM_VALIDATE	PROCESS DRIVER: SDTM_VALIDATE	Info
CST0200		4	SDTM_VALIDATE	PROCESS DATE: 2010-10-12T13:38:05	Info
CST0200		5	SDTM_VALIDATE	PROCESS TYPE: VALIDATION	Info
CST0200		6	SDTM_VALIDATE	PROCESS SASREFERENCES: C:\projects\PhUSE\demo1\control\sasreferences.sas7bdat	Info
CST0100	SDTM0011	1	WORK._CSTSRCCOLUMN METADATA	No errors detected in source data	Info
...	...	...	...	...	...
SDTM0015	SDTM0015	1	SUPPAE	Variable IDVAR appears in dataset but is not in SDTM standard	Warning
SDTM0015	SDTM0015	2	SUPPAE	Variable IDVARVAL appears in dataset but is not in SDTM standard	Warning
...	...	...	...	...	...
CST0100	SDTM0019	1	WORK._CSTSRCCOLUMN METADATA	No errors detected in source data	Info
...	...	...	...	...	...
SDTM0452	SDTM0452	1	SRCDATA.AE	AE is Serious but no qualifiers set to 'Y'	Note
CST0029	SDTM0453	1	CSTCHECK_NOTINCodel IST	Format catalog WORK.FORMATS in fmtsearch could not be found	Info
CST0033	SDTM0453	2	CSTCHECK_NOTINCodel IST	Format search path has been set to WORK.FORMATS SRCFMT.FORMATS CSTFMT.CTERMS	Info
CST0100	SDTM0453	3	SRCDATA.AE.AESER	No errors detected in source data	Info

Validation of study data against the SDTM standard – validation checks

Validation check identifier	Source of check	Severity of check	Category of check	SAS macro module name	Domains to which check applies	Columns to which check applies	SAS format name
SDTM0011	Janus	Note	Metadata	cstcheck_metamismatch	_ALL_		
SDTM0012	JanusFR	Error	Metadata	cstcheck_metamismatch	_ALL_		
SDTM0013	Janus	Note	Metadata	cstcheck_metamismatch	_ALL_		
SDTM0014	SAS	Note	Metadata	cstcheck_metamismatch	_ALL_		
SDTM0015	Janus	Warning	Metadata	cstcheck_metamismatch	_ALL_		
SDTM0019	JanusFR	Warning	Metadata	cstcheck_metamismatch	_ALL_		
SDTM0020	SAS	Warning	Metadata	cstcheck_metamismatch	_ALL_		
SDTM0022	SAS	Note	Metadata	cstcheck_metamismatch	_ALL_		
SDTM0023	SAS	Error	Metadata	cstcheck_metamismatch	_ALL_		
SDTM0030	SAS	Note	Metadata	cstcheck_metamismatch	_ALL_		
SDTM0031	SAS	Error	Metadata	cstcheck_metamismatch	_ALL_		
SDTM0032	SAS	Note	Metadata	cstcheck_metamismatch	_ALL_		
SDTM0452	Janus	Note	ColumnValue	cstcheck_column	AE	AESER	
SDTM0453	JanusFR	Error	Cntlterm	cstcheck_notincodelist	AE	AESER	\$NY

Validation of study data against the SDTM standard – messages for checks

Result identifier	Rule description from checksource	Message text
SDTM0011	Identifies a column that was described in the domain description but not included in the SAS dataset for that domain	Variable &_cstparm1 in description file not in dataset
SDTM0012	Identifies a column listed in the domain description as Required ('Req') but not included in the SAS dataset for that domain	SDTM required variable &_cstparm1 not found
SDTM0013	Identifies a column listed in the domain description as Expected ('Exp') but not included in the SAS dataset for that domain	SDTM expected variable &_cstparm1 not found
SDTM0015	Identifies a column that appears in the SAS dataset but is not listed in the domain description	Variable &_cstparm1 appears in dataset but is not in SDTM standard
SDTM0019	Identifies a variable where datatype in (study specific) description is not consistent with datatype implicit in SAS dataset	Description file/dataset variable type mismatch for &_cstparm1
SDTM0020	Column order does not match standard	Column order does not match standard for &_cstparm1
SDTM0022	Column length < length defined in standard	Column length < length defined in standard for &_cstparm1
SDTM0023	Column length > length defined in standard	Column length > length defined in standard for &_cstparm1
SDTM0030	Column label inconsistent with label defined in standard	Column label inconsistent with label defined in standard for &_cstparm1
SDTM0031	Column format found but column not subject to controlled terminology	Column not subject to controlled terminology for &_cstColumn
SDTM0032	Column format found but format name mismatch with standard controlled terminology name	Column format name mismatch with standard for &_cstparm1
SDTM0452	Identifies records where Serious Event (AESER)='Y' but none of Involves Cancer (AESCAN), Congenital Anomaly or Birth Defect (AESCONG), Persist or Signif Disability/Incapacity (AESDISAB), Results in Death (AESDTH), Requires or Prolongs Hospitalization (AESHOSP), Is Life Threatening (AESLIFE), Other Medically Important Serious Event (AESMIE), or Occurred with Overdose (AESOD) equals 'Y'	AE is Serious but no qualifiers set to 'Y'
SDTM0453	Identifies records where value for [Serious Event (AESER)] is not found in Codelist [YESNO]	Invalid YESNO code

Validation of study data against the SDTM standard – more complete example

- Generate the SASReferences dataset
 - See next slide
- Select validation checks

```
data work.checks;  
  set refcntl.validation_master;  
  where checkid='SDTM0452' and checksource='Janus'  
    or checkid='SDTM0453' and checksource='JanusFR'  
    or checkid='SDTM0011' and checksource='Janus';  
run;
```

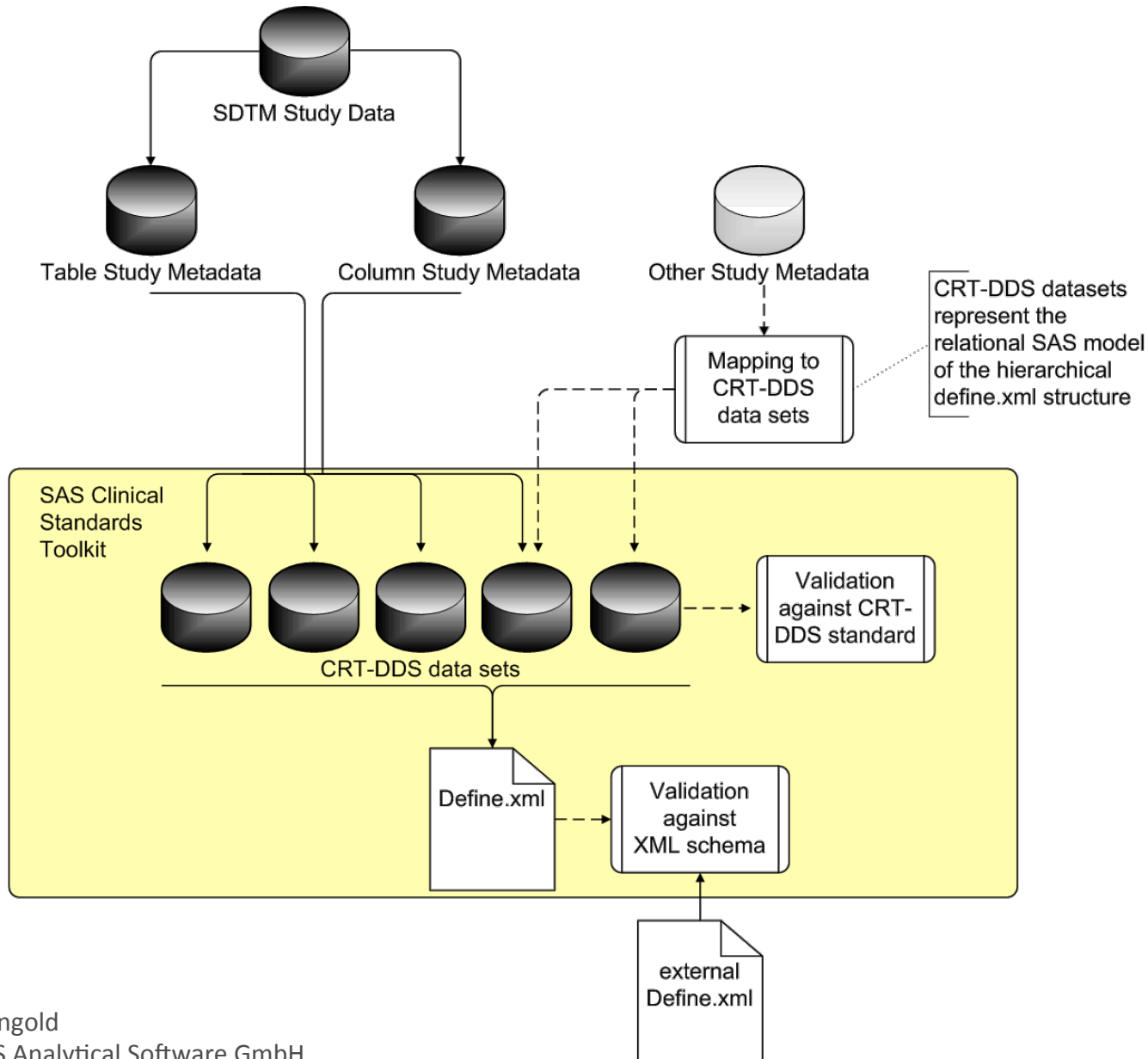
- Save and restore options

```
%cstutil_cleanupcstsession(_cstClearCompiledMacros=1  
  ,_cstClearLibRefs=1      ,_cstResetSASAutos=1  
  ,_cstResetFmtSearch=1    ,_cstResetSASOptions=1  
  ,_cstDeleteFiles=1       ,_cstDeleteGlobalMacroVars=1);  
options mrecall;
```

Validation of study data against the SDTM standard – SASReferences control dataset

Standard	Version	Data or metadata type	subtype	SAS libref or fileref	Reference type	Relative path	Filename (null for libraries)
CDISC-SDTM	3.1.1	sourcedata		srcdata	libref	&studyRootPath\data	
CDISC-SDTM	3.1.1	sourcemetadata	table	srcmeta	libref	&studyRootPath\metadata	source_tables.sas7bdat
CDISC-SDTM	3.1.1	sourcemetadata	column	srcmeta	libref	&studyRootPath\metadata	source_columns.sas7bdat
CDISC-SDTM	3.1.1	autocall		sdtmcode	fileref	&_cstGRoot\standards\cdisc-sdtm-3.1.1\macros	
CDISC-SDTM	3.1.1	fmtsearch		srcfmt	libref	&studyRootPath\terminology\formats	formats.sas7bcat
CDISC-TERMINOLOGY	200810	fmtsearch		cstfmt	libref	&_cstGRoot\standards\cdisc-terminology-200810\formats	cterms.sas7bcat
CDISC-SDTM	3.1.1	control	validation	control	libref	&studyRootPath\control	validation_control.sas7bdat
CDISC-SDTM	3.1.1	control	reference	control	libref	&studyRootPath\control	sasreferences.sas7bdat
CDISC-SDTM	3.1.1	messages		sdtmmsg	libref	&_cstGRoot\standards\cdisc-sdtm-3.1.1\messages	messages.sas7bdat
CST-FRAMEWORK	1.2	messages		cstmsg	libref	&_cstGRoot\standards\cst-framework\messages	messages.sas7bdat
CDISC-SDTM	3.1.1	properties	validation	valprop	fileref	&studyRootPath\programs	validation.properties
CDISC-SDTM	3.1.1	results	validationresults	results	libref	&studyRootPath\results	validation_results.sas7bdat
CDISC-SDTM	3.1.1	results	validationmetrics	results	libref	&studyRootPath\results	validation_metrics.sas7bdat

Generation of define.xml



Generation of define.xml – simple example

```
/*-- root location of the process input and output --*/
%let studyRootPath=C:\projects\PhUSE\demo3;

/*-- load basic configuration to macro variables --*/
%cst_setStandardProperties(_cstStandard=CST-FRAMEWORK, _cstSubType=initialize);
%cst_setStandardProperties(_cstStandard=CDISC-CRTDDS , _cstSubType=initialize);
%cst_setStandardProperties(
    _cstStandard=CDISC-TERMINOLOGY, _cstSubType=initialize);

/*-- process sasreferences: allocate librefs etc. --*/
%let _cstSASRefsLoc=&studyRootPath\control;%let _cstSASRefsName=sasrefs;
%cstutil_allocatesasreferences;

/*-- create intermediate CRTDDS format --*/
libname meta "&studyRootPath/metadata";
%crtdds_sdtm311todefinel0(
    _cstOutLib=srcdata /* allocated by sasrefs */
    ,_cstSourceTables=meta.source_tables
    ,_cstSourceColumns=meta.source_columns
    ,_cstSourceStudy=meta.source_study
);

/*-- generate define.xml --*/
%crtdds_write(
    _cstCreateDisplayStyleSheet=1
    ,_cstResultsOverrideDS=&_cstResultsDS
);
```

Generation of define.xml – simple example – output

Datasets for Study study1					
Dataset	Description	Structure	Purpose	Keys	Location
AE	Adverse Events	Events - One record per event per subject	Tabulation	STUDYID USUBJID AETERM AESTDTC	
CM	Concomitant Medications	Interventions - One record per medication intervention episode per subject	Tabulation	STUDYID USUBJID CMTRT CMSTDTC	
CO	Comments	Special Purpose - One record per comment per subject	Tabulation	STUDYID USUBJID COSEQ	
DM	Demographics	Special Purpose - One record per subject	Tabulation	STUDYID USUBJID	
DS	Disposition	Events - One record per disposition status or protocol milestone per subject	Tabulation	STUDYID USUBJID DSSTDTC	
DV	Protocol Deviations	Events - One record per protocol deviation per subject	Tabulation	STUDYID USUBJID DVSEQ	
EG	ECG Test Results	Findings - One record per ECG observation per time point per visit per subject	Tabulation	STUDYID USUBJID EGTESTCD VISITNUM EGTPTNUM EGSEQ	
EX	Exposure	Interventions - One record per constant dosing interval per subject	Tabulation	STUDYID USUBJID EXTRT EXSTDTC	
IE	Inclusion/Exclusion Exceptions	Findings - One record per Inclusion/Exclusion criteria exception per subject	Tabulation	STUDYID USUBJID IETESTCD	
LB	Laboratory Tests	Findings - One record per lab test per time point per visit per subject	Tabulation	STUDYID USUBJID LBTESTCD VISITNUM LBTPPTNUM	
MH	Medical History	Events - One record per medical history event per subject	Tabulation	STUDYID USUBJID MHTERM	
PE	Physical Examination	Findings - One record per body system per visit per subject	Tabulation	STUDYID USUBJID VISITNUM PETESTCD	

Generation of define.xml – extended example

```
/*-- initialize --*/
* ...;
/*-- create all 39 CRT-DDS data sets --*/
%cst_createTablesForDataStandard(_cstStandard=CDISC-CRTDDS
    ,_cstOutputLibrary=srcdata);

/*-- fill 9 of the 39 tables --*/
libname meta "&studyRootPath/metadata";
%crtdds_sdtm311todefined10(
    _cstOutLib=srcdata,_cstSourceTables=meta.source_tables
    ,_cstSourceColumns=meta.source_columns,_cstSourceStudy=meta.source_study);

/*-- Add information about archive locations --*/
proc sql;
    update srcdata.itemgroupdefs set archivelocationid = 'ALID'!!oid;
    insert into srcdata.itemgroupleaf (id, href, fk_itemgroupdefs)
    select 'ALID'!!i.oid, s.xmlpath, i.oid
        from meta.source_tables s join srcdata.itemgroupdefs i on s.table=i.name;
    delete from srcdata.itemgroupleaf where id=' ';
    insert into srcdata.itemgroupleaftitles (fk_itemgroupleaf, title)
    select 'ALID'!!i.oid, s.xmltitle
        from meta.source_tables s join srcdata.itemgroupdefs i on s.table=i.name;
    delete from srcdata.itemgroupleaftitles where fk_itemgroupleaf=' ';
quit;

/*-- create define.xml --*/
*%crtdds_write(...);
```

Generation of define.xml – extended example – output

Datasets for Study study1					
Dataset	Description	Structure	Purpose	Keys	Location
AE	Adverse Events	Events - One record per event per subject	Tabulation	STUDYID USUBJID AETERM AESTDTC	Adverse Events SAS transport file
CM	Concomitant Medications	Interventions - One record per medication intervention episode per subject	Tabulation	STUDYID USUBJID CMTRT CMSTDTC	Concomitant Medications SAS transport file
CO	Comments	Special Purpose - One record per comment per subject	Tabulation	STUDYID USUBJID COSEQ	Comments SAS transport file
DM	Demographics	Special Purpose - One record per subject	Tabulation	STUDYID USUBJID	Demographics SAS transport file
DS	Disposition	Events - One record per disposition status or protocol milestone per subject	Tabulation	STUDYID USUBJID DSSTDTC	Disposition SAS transport file
DV	Protocol Deviations	Events - One record per protocol deviation per subject	Tabulation	STUDYID USUBJID DVSEQ	Protocol Deviations SAS transport file
EG	ECG Test Results	Findings - One record per ECG observation per time point per visit per subject	Tabulation	STUDYID USUBJID EGTESTCD VISITNUM EGTPPTNUM EGSEQ	ECG Test Results SAS transport file
EX	Exposure	Interventions - One record per constant dosing interval per subject	Tabulation	STUDYID USUBJID EXTRT EXSTDTC	Exposure SAS transport file
IE	Inclusion/Exclusion Exceptions	Findings - One record per Inclusion/Exclusion criteria exception per subject	Tabulation	STUDYID USUBJID IETESTCD	Inclusion/Exclusion Exceptions SAS transport file
LB	Laboratory Tests	Findings - One record per lab test per time point per visit per subject	Tabulation	STUDYID USUBJID LBTESTCD VISITNUM LBTPPTNUM	Laboratory Tests SAS transport file
MH	Medical History	Events - One record per medical history event per subject	Tabulation	STUDYID USUBJID MHTERM	Medical History SAS transport file
PE	Physical Examination	Findings - One record per body system per visit per subject	Tabulation	STUDYID USUBJID VISITNUM PETESTCD	Physical Examination SAS transport file

Add information to define.xml – process

- Look at the CDISC "Case Report Tabulation Data Definition Specification"*
- Determine which (sub-)elements and attributes have to be supplied to address the metadata in question
- Follow the section about the CRT-DDS data model in the toolkit user's guide*
- identify the data sets and columns of interest and sort out how tables have to be linked together by foreign keys
- Write a program which fills the data sets accordingly

Further Steps – beyond programming

- Administration of standards
 - installing new versions of standards (e.g. SDTM 3.1.2)
 - modification of existing standards
 - bringing in of new domains
 - development of company specific (variants of) standards
- Different kinds of toolkit users
 - administer metadata and standards
 - use metadata and standards
 - which users needs which access rights?
- Training
 - Knowledge of the data standards
 - CDISC implementation
 - clinical data management practices
 - technical aspects

Thank you for your attention

If you want to try out the examples by yourself, send an e-mail to the authors and request the sample data and programs.

Andreas Mangold
Nicole Wächter

HMS Analytical Software GmbH
Rohrbacher Str. 26 • 69115 Heidelberg
Telefon +49 6221 6051-0
andreas.mangold@analytical-software.de
nicole.waechter@analytical-software.de
www.analytical-software.de