

**USING THE
SAS® CLINICAL STANDARDS TOOLKIT 1.5
FOR DEFINE.XML CREATION**



LEX JANSEN, SAS, CDISC XML TECHNOLOGIES TEAM

**Pharma
SUG**



CHICAGO

May 12-15, 2013

WHAT IS DEFINE-XML ?

HISTORY AND BACKGROUND



THE POWER TO KNOW.

Define-XML HISTORY AND BACKGROUND

- **July 2012** – Study Data Specifications v2.0:
The specification for the data definitions for datasets provided using *CDISC* is included in the Case Report Tabulation Data Definition Specification ...”

(CRT-DDS – define.xml)

Define-XML HISTORY AND BACKGROUND

- **December 2011** CDER Common Data Standards Issues Document (Version 1.1):

“A critical component of data submission is the define file. A properly functioning define.xml file is an important part of the submission of standardized electronic datasets and should not be considered optional.”

“Additionally, sponsors should make certain that every data variable’s code list, origin, and derivation is clearly and easily accessible from the define file. An insufficiently documented define file is a common deficiency that reviewers have noted.”

Define-XML HISTORY AND BACKGROUND



Don't give this message with a define.xml file

Define-XML WHAT IS IT ?

- Case Report Tabulation Data **Specification** (CRT-DDS, or define.xml): Production version: 1.0.0
- Extension of the CDISC Operational Data Model (**ODM**), an XML specification to facilitate the archival and interchange of the metadata and data for clinical research
- Maintained by CDISC's XML Technologies Team
- New **Define-XML** v2.0 has been released with additional metadata support (based on ODM 1.3.2)

Define-XML WHAT IS IT ?

- Provides **metadata** structures to describe data sets:
 - **Study**
 - Name, Description Protocol name
 - **Domains**
 - Name, Description, Structure, Dataset Location, ...
 - **Variables**
 - Name, Label, Data Type, Length, ...
 - **Controlled Terminology**
 - **Derivations** (computations, methods)
 - Supporting **Documents** (aCRF, Supplemental Data Definitions, ..)
 - **Value Level Metadata**
 - Concepts as **SuppQuals**, **Split domains** ...

Define-XML WHAT IS IT ?

- Specification for describing data sets
- Does **not** describe how this metadata should be displayed - that is not part of the standard

Datasets for Study study1					
Dataset	Description	Structure	Purpose	Keys	Location
AE	Adverse Events	Events - One record per adverse event per subject	Tabulation	STUDYID USUBJID AEDECOD AESTDTC	Adverse Events SAS transport file
CE	Clinical Events	Events - One record per event per subject	Tabulation	STUDYID USUBJID CETERM CESTDTC	Clinical Events SAS transport file
CM	Concomitant Medications	Interventions - One record per recorded medication occurrence or constant-dosing interval per subject	Tabulation	STUDYID USUBJID CMTR1 CMSTDTC	Concomitant Medications SAS transport file
CO	Comments	Special Purpose Domains - One record per comment per subject	Tabulation	STUDYID USUBJID COSEQ	Comments SAS transport file
DA	Drug Accountability	Findings - One record per drug accountability finding per subject	Tabulation	STUDYID USUBJID DATESTCD DADTC	Drug Accountability SAS transport file
DM	Demographics	Special Purpose Domains - 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 DSDECOD DSSTDTC	Disposition SAS transport file

- ▢ Annotated Case Report Form
- ▢ Reviewers Guide
- ▢ Datasets
- ▢ Value Level Metadata
- ▢ Computational Algorithms
- ▢ Controlled Terms

Datasets for Study CDISC01					
Dataset	Description	Class	Structure	Purpose	Keys
TA	Trial Arms	Trial Design	One record per planned Element per Arm	Tabulation	STUDYID, ARMCD, TAETORD
TE	Trial Elements	Trial Design	One record per planned Element	Tabulation	STUDYID, ETCD
TI	Trial Inclusion/Exclusion Criteria	Trial Design	One record per I/E criterion	Tabulation	STUDYID, IETESTCD
TS	Trial Summary	Trial Design	One record per trial summary parameter value	Tabulation	STUDYID, TSPARMCD, TSSEQ
TV	Trial Visits	Trial Design	One record per planned Visit per Arm	Tabulation	STUDYID, VISITNUM, ARMCD
DM	Demographics	Special Purpose	One record per subject	Tabulation	STUDYID, USUBJID

- ▢ Annotated Case Report Form
- ▢ SDTM Datasets
- ▢ Value Level Metadata
- ▢ External Dictionary
- ▢ Computational Algorithms

SDTM Datasets for Study CDISCPILOT01

Dataset	Description	Class	Structure	Purpose
TA	Trial Arms	Trial Design	One record per planned Element per Arm	Tabulation
TE	Trial Elements	Trial Design	One record per planned Element	Tabulation
TI	Trial Inclusion/Exclusion Criteria	Trial Design	One record per I/E criterion	Tabulation
TS	Trial Summary	Trial Design	One record per trial summary parameter value	Tabulation
TV	Trial Visits	Trial Design	One record per planned Visit per Arm	Tabulation
DM	Demographics	Special Purpose	One record per subject	Tabulation
SE	Subject Elements	Special Purpose	One record per actual Element per subject	Tabulation
SV	Subject Visits	Special Purpose	One record per actual visit per subject	Tabulation
CM	Concomitant Medications	Interventions	One record per recorded medication occurrence or constant-dosing interval per subject	Tabulation
EX	Exposure	Interventions	One record per constant dosing	Tabulation








































Define-XML WHAT IS IT ?



```
<ItemGroupDef OID="ADSL" Name="ADSL" Repeating="No" IsReferenceData="No"
  Purpose="Analysis" def:Label="Subject-Level Analysis"
  def:Structure="one record per subject" def:DomainKeys="USUBJID"
  def:Class="ADSL" def:ArchiveLocationID="Location.ADSL">
  <ItemRef ItemOID="ADSL.STUDYID" OrderNumber="1" Mandatory="No"/>
  <ItemRef ItemOID="ADSL.USUBJID" OrderNumber="2" Mandatory="No"/>
  <ItemRef ItemOID="ADSL.SUBJID" OrderNumber="3" Mandatory="No"/>
  ...
  <ItemRef ItemOID="ADSL.ARM" OrderNumber="6" Mandatory="No"/>
  ...
  <def:leaf ID="Location.ADSL" xlink:href="adsl.xpt">
    <def:title>adsl.xpt </def:title>
  </def:leaf>
</ItemGroupDef>
...
<ItemDef OID="ADSL.ARM" Name="ARM" DataType="text" Length="20"
  Origin="Derived" Comment="DM.ARM" def:Label="Description of Planned Arm">
  <CodeListRef CodeListOID="ARM"/>
</ItemDef>
```

Define-XML SAS AND DEFINE.XML

- SAS provides data model that represents CRT-DDS Version 1.0 format in 39 SAS data sets
- 20 of these are typically used for the define.xml (*)
- Patterned to match the XML element and attribute structure of the define.xml file
- XML element → SAS dataset
- XML attribute → SAS variable

★  annotatedcrfs	★  clitemdecodetranslatedtext	★  codelistitems	★  codelists
★  computationmethods	★  definedocument	★  externalcodelists	 formdefarchlayouts
 formdefitemgrouprefs	 formdefs	 imputationmethods	 itemaliases
★  itemdefs	 itemgroupaliases	★  itemgroupdefitemrefs	★  itemgroupdefs
★  itemgroupleaf	★  itemgroupleaftitles	 itemmurefs	 itemquestionexternal
 itemquestiontranslatedtext	 itemrangechecks	 itemrangecheckvalues	 itemrole
★  itemvaluelistrefs	★  mdvleaf	★  mdvleaftitles	 measurementunits
★  metadataversion	 mutranslatedtext	 presentation	 protocoleventrefs
 rcerrortranslatedtext	★  study	 studyeventdefs	 studyeventformrefs
★  supplementaldocs	★  valuelistitemrefs	★  valuelists	

SAS® CLINICAL STANDARDS TOOLKIT (CST)



- **Framework** to primarily support **Clinical Research** activities.
- Initially focusing on standards as defined by CDISC, but not limited to CDISC.
- Designed as an integral part of **SAS Clinical Data Integration** (CDI), but is available to all licensed SAS customers as open source SAS Macros and metadata at no additional charge.
- Designed to supersede PROC CDISC
- Framework: designed to **customize** and **extend**

SAS Clinical Standards Toolkit 1.5

INTRODUCTION

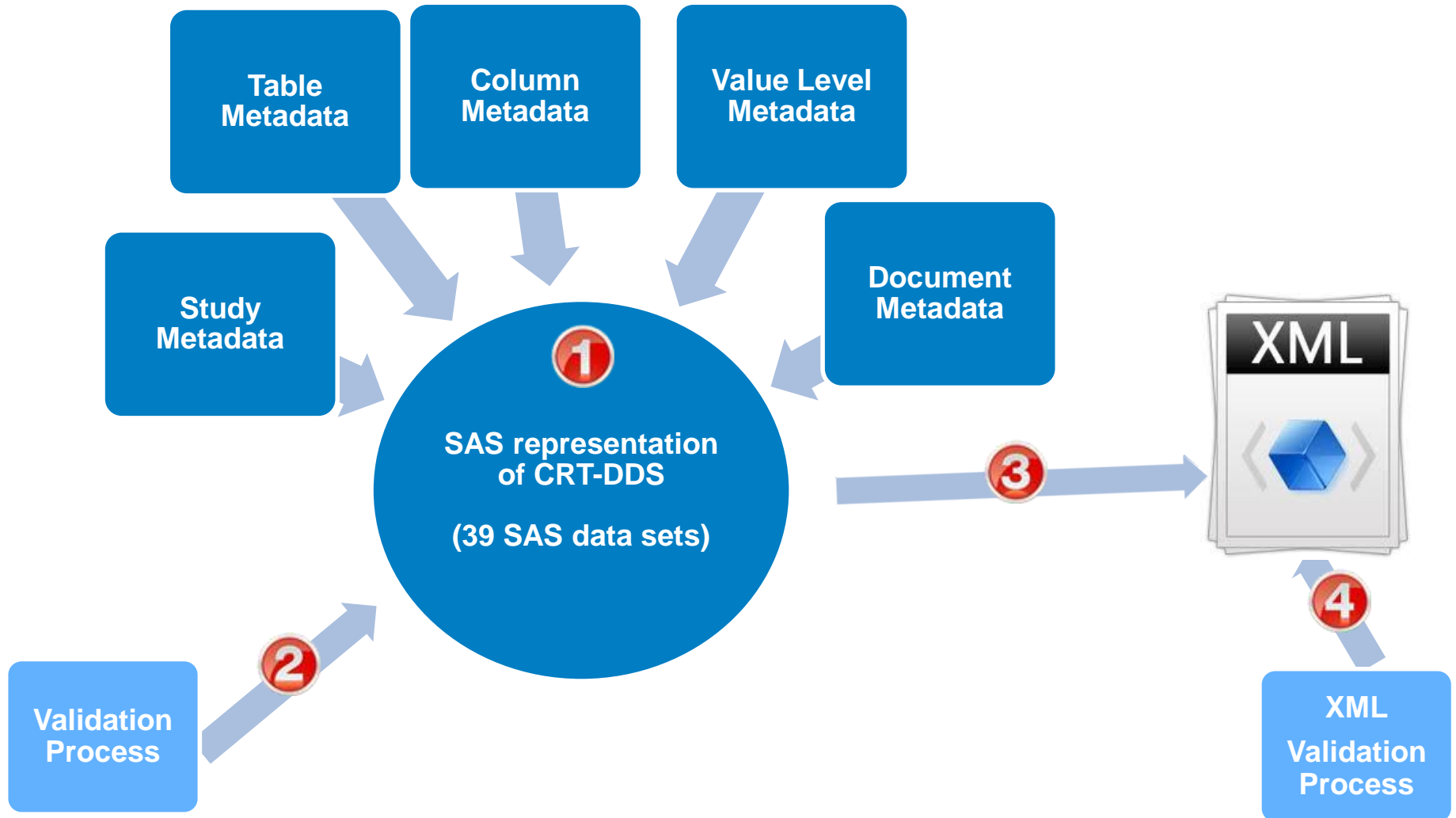
- Supported CDISC standards in Toolkit 1.5 (**just released**):
 - **SDTM** 3.1.1, 3.1.2 and 3.1.3
 - **ADaM** 2.1 (ADSL, Basic Data Structure, ADAE and ADTTE) and Analysis Results Metadata templates; v1.1 of the ADaM validation checks)
 - **SEND** 3.0 (initial implementation)
 - **CRT-DDS** 1.0 (Define-XML - Create / Import / Validate / PDF)
 - **ODM** 1.3.0, 1.3.1 - Read / Write / Validate
 - **NCI CDISC Controlled Terminology** (December 2012)
(import/export of ODM XML through CT 1.0 standard)

SAS Clinical Standards Toolkit 1.5

INTRODUCTION

- Supported with SAS 9.3M2 on the following operating systems:
 - Windows 32
 - Windows for x64
 - Linux for x64
 - Solaris x64 SPARC
- Separately orderable component
- Available at no additional charge to currently licensed SAS customers.
- Contact your SAS Account Representative concerning availability

DEFINE.XML CREATION PROCESS



Define-XML SOURCE METADATA

VIEWTABLE: Sdtmmeta.Source_study (Source Study Metadata)

	definedocumentname	sasref	studyname	protocolname	studydescription
1	define1	SRCDATA	study1	Protocol abc	first study

VIEWTABLE: Sdtmmeta.Source_tables (Source Table Metadata)

	SASref	Table	Label	Class	XmlPath	XmlTitle	Structure	Purpose	Keys
1	SRCDATA	AE	Adverse Events	Events	../transport/ae.xpt	Adverse Events SAS transport file	One record per adverse event per subject	Tabulation	STUDYID USUBJID AEDECOD AESTDTC
2	SRCDATA	CE	Clinical Events	Events	../transport/ce.xpt	Clinical Events SAS transport file	One record per event per subject	Tabulation	STUDYID USUBJID CETERM CESTDTC
3	SRCDATA	CM	Concomitant Medications	Interventions	../transport/cm.xpt	Concomitant Medications SAS transport file	One record per recorded medication occurrence or constant-dosing interval per subject	Tabulation	STUDYID USUBJID CMTRT CMSTDTC

VIEWTABLE: Sdtmmeta.Source_columns (Source Column Metadata)

	SASref	table	column	label	order	type	length	displayformat	xmldatatype	xmlcodelist	core	origin	role
1	SRCDATA	AE	STUDYID	Study Identifier	1	C	40		text		Req		Identifier
2	SRCDATA	AE	DOMAIN	Domain Abbreviation	2	C	8		text		Req		Identifier
3	SRCDATA	AE	USUBJID	Unique Subject Identifier	3	C	40		text		Req		Identifier
4	SRCDATA	AE	AESQ	Sequence Number	4	N	8		integer		Req		Identifier
5	SRCDATA	AE	AEGRPID	Group ID	5	C	40		text		Pem		Identifier
6	SRCDATA	AE	AEREFID	Reference ID	6	C	40		text		Pem		Identifier
7	SRCDATA	AE	AESPID	Sponsor-Defined Identifier	7	C	40		text		Pem		Identifier
8	SRCDATA	AE	AETERM	Reported Term for the Adverse Event	8	C	200		text		Req		Topic
9	SRCDATA	AE	AEMODIFY	Modified Reported Term	9	C	200		text		Pem		SynonymQualifier
10	SRCDATA	AE	AEDECOD	Dictionary-Derived Term	10	C	200		text		Req		SynonymQualifier
11	SRCDATA	AE	AECAT	Category for Adverse Event	11	C	40		text		Pem		GroupingQualifier
12	SRCDATA	AE	AESCAT	Subcategory for Adverse Event	12	C	40		text		Pem		GroupingQualifier
13	SRCDATA	AE	AEPRESP	Pre-Specified Adverse Event	13	C	2		text	NY	Pem		RecordQualifier
14	SRCDATA	AE	AEBODSYS	Body System or Organ Class	14	C	80		text		Exp		RecordQualifier
15	SRCDATA	AE	AELC	Location of Event	15	C	40		text	LOC	Pem		RecordQualifier
16	SRCDATA	AE	AESEV	Severity/Intensity	16	C	20		text	AESEV	Pem		RecordQualifier
17	SRCDATA	AE	AESER	Serious Event	17	C	2		text	NY	Exp		RecordQualifier
18	SRCDATA	AE	AEACN	Action Taken with Study Treatment	18	C	40		text	ACN	Exp		RecordQualifier

Define-XML VALUE LEVEL METADATA

- Value Level Metadata defines **metadata** for a **variable** under a specific **condition**
- Needed in the highly normalized data structure of SDTM, SEND and ADaM (generally one record per subject per test code or parameter per visit or observation)
- Examples:
 - VSORRES, VSSTRESN or VSTRESU based on the value of VSTESTCD
 - QVAL bases on the value of QNAM

USUBJID	VSTESTCD	VSTEST	VSORRES	VSSRESN	VSSRESC	VISITNUM
00001	HEIGHT	Height in cm	183	183		1
00001	WEIGHT	Weight in kg	88.5	88.5		1
00001	FRAME	Body Frame Size	Large		Large	1

Define-XML VALUE LEVEL METADATA

VIEWTABLE: Sdtmmeta.Source_columns (Source Column Metadata)

	SASref	table	column	label	order	type	length	displayfor	xmldatatype	xmlcodelist	core	origin
1	SRCDATA	AE	STUDYID	Study Identifier	1	C	40		text		Req	
2	SRCDATA	AE	DOMAIN	Domain Abbreviation	2	C	8		text		Req	
3	SRCDATA	AE	USUBJID	Unique Subject Identifier	3	C	40		text		Req	
4	SRCDATA	AE	AESEQ	Sequence Number	4	N	8		integer		Req	
5	SRCDATA	AE	AEGRPID	Group ID	5	C	40		text		Perm	
6	SRCDATA	AE	AEREFID	Reference ID	6	C	40		text		Perm	
7	SRCDATA	AE	AESPID	Sponsor-Defined Identifier	7	C	40		text		Perm	
8	SRCDATA	AE	AETERM	Reported Term for the Adverse Event	8	C	200		text		Req	
9	SRCDATA	AE	AEMODIFY	Modified Reported Term	9	C	200		text		Perm	
10	SRCDATA	AE	AEDECOD	Dictionary-Derived Term	10	C	200		text		Req	
11	SRCDATA	AE	AECAT	Category for Adverse Event	11	C	40		text		Perm	
12	SRCDATA	AE	AESCAT	Subcategory for Adverse Event	12	C	40		text		Perm	

VIEWTABLE: Sdtmmeta.Source_values (Source Value Metadata)

	SASref	table	column	value	label	order	type	length	displayfor	xmldatatype	xmlcodelist	core	origin
6	SRCDATA	IE	IETESTCD	INCL02	Acceptable chest X-Ray	1	C	2		text	NY	Perm	CRF Page 5
7	SRCDATA	IE	IETESTCD	INCL10	Systolic BP > 180	2	C	2		text	NY	Perm	CRF Page 5
8	SRCDATA	LB	LBTESTCD	CALCIUM	Calcium	1	N	8 8.2		float		Perm	eDT
9	SRCDATA	LB	LBTESTCD	CHLORIDE	Chloride	2	N	8		integer		Perm	eDT
10	SRCDATA	LB	LBTESTCD	POTASS	Potassium	3	N	8 8.2		float		Perm	eDT
11	SRCDATA	LB	LBTESTCD	SODIUM	Sodium	4	N	8		integer		Perm	eDT
12	SRCDATA	PE	PETESTCD	CARDIO	Cardiovascular	1	N	8		integer		Perm	CRF Page 10
13	SRCDATA	PE	PETESTCD	ENT	Ear/Nose/Throat	2	N	8		integer		Perm	CRF Page 10
14	SRCDATA	PE	PETESTCD	RESP	Respiratory	3	N	8		integer		Perm	CRF Page 10
15	SRCDATA	PE	PETESTCD	SKIN	Skin	4	N	8		integer		Perm	CRF Page 10
16	SRCDATA	SC	SCTESTCD	INITIALS	Initials	1	C	3		text		Perm	CRF Page 6
17	SRCDATA	SC	SCTESTCD	RACEOTH	Race, Other	2	C	10		text		Perm	CRF Page 6
18	SRCDATA	SUPPAE	QNAM	AECONIA	Interaction between add1 and trial meds	1	C	2		text	NY	Perm	CRF Page 17
19	SRCDATA	SUPPAE	QNAM	AETRTEM	Treatment emergent	2	C	2		text	NY	Perm	CRF Page 17
20	SRCDATA	TI	IETESTCD	EXCL01	Systolic BP > 180	4	C	2		text	NY	Perm	CRF Page 14
21	SRCDATA	TI	IETESTCD	EXCL02	Diastolic BP > 120	3	C	2		text	NY	Perm	CRF Page 14
22	SRCDATA	TI	IETESTCD	INCL01	Age between 18 and 70	2	C	2		text	NY	Perm	CRF Page 14
23	SRCDATA	TI	IETESTCD	INCL02	Acceptable chest X-Ray	1	C	2		text	NY	Perm	CRF Page 14
24	SRCDATA	VS	VSTESTCD	DIABP	Diastolic Blood Pressure	1	N	8		integer		Perm	CRF Page 11
25	SRCDATA	VS	VSTESTCD	FRMSIZE	Frame Size	2	C	6		text		Perm	CRF Page 11
26	SRCDATA	VS	VSTESTCD	HRATE	Heart Rate	3	N	8		integer		Perm	CRF Page 11

- Define global macro variables ("properties")
 - **%cst_setStandardProperties**
(_cstStandard=CST-FRAMEWORK,_cstSubType=initialize);
- Define inputs / outputs (libname refs, filename refs, SAS autocall macros, ...)
 1. Create SASReferences dataset
 2. **%cstutil_processsetup()**; (default: use WORK.SASReferences)
- Run process specific macro:
 - **%crtdds_sdtmtodefine**
 - **%crtdds_validate**
 - **%crtdds_write**
 - **%crtdds_writepdf**
 - **%crtdds_xmlvalidate**
 - **%crtdds_read**

DEMO



Thank You !

Questions ?



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

**THE
POWER
TO KNOW.**

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

E-mail: lex.jansen@sas.com