

Traceability between SDTM and ADaM converted analysis datasets



1

Introduction

2

ADaM Conversion

3

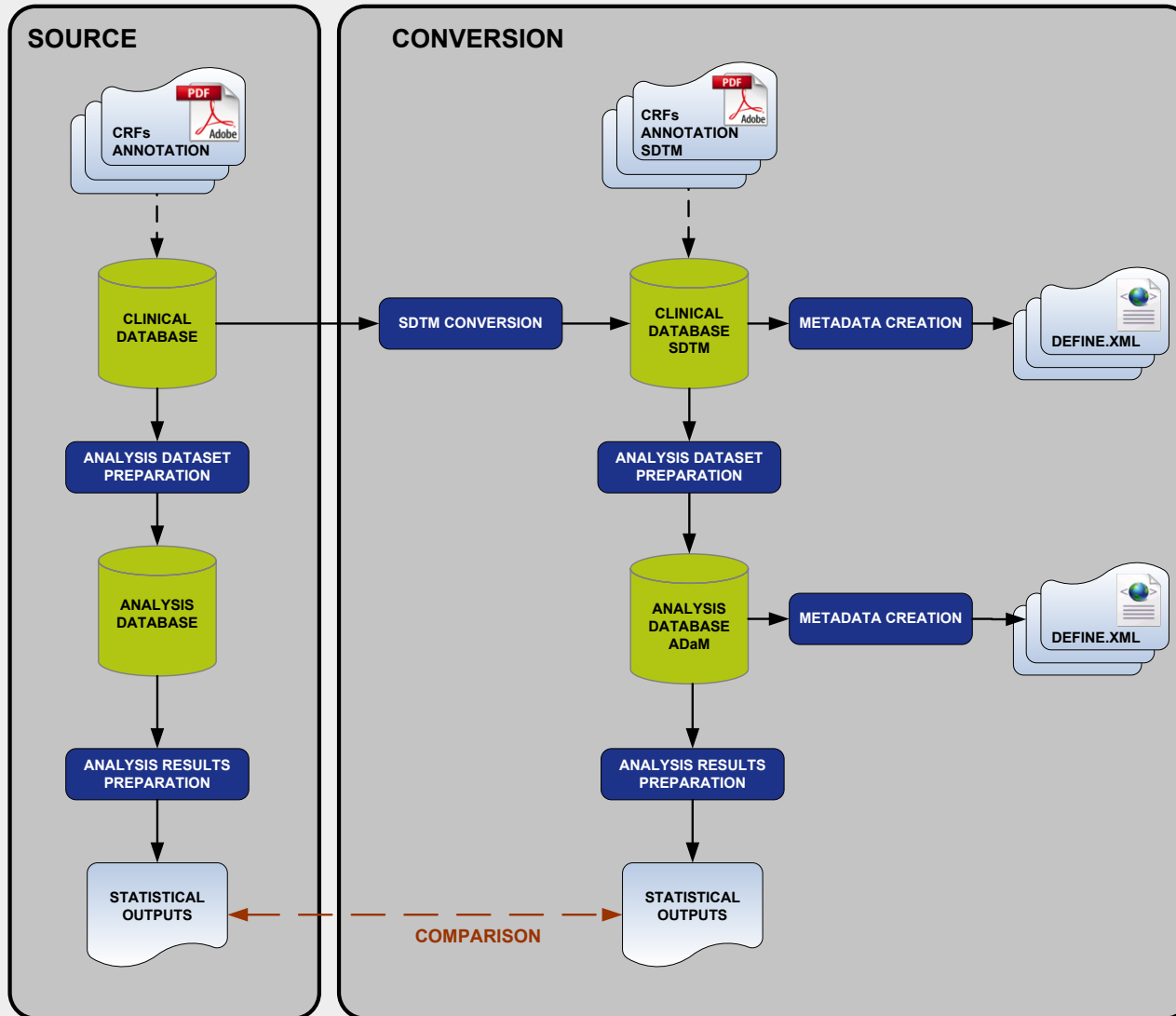
Quality Control

4

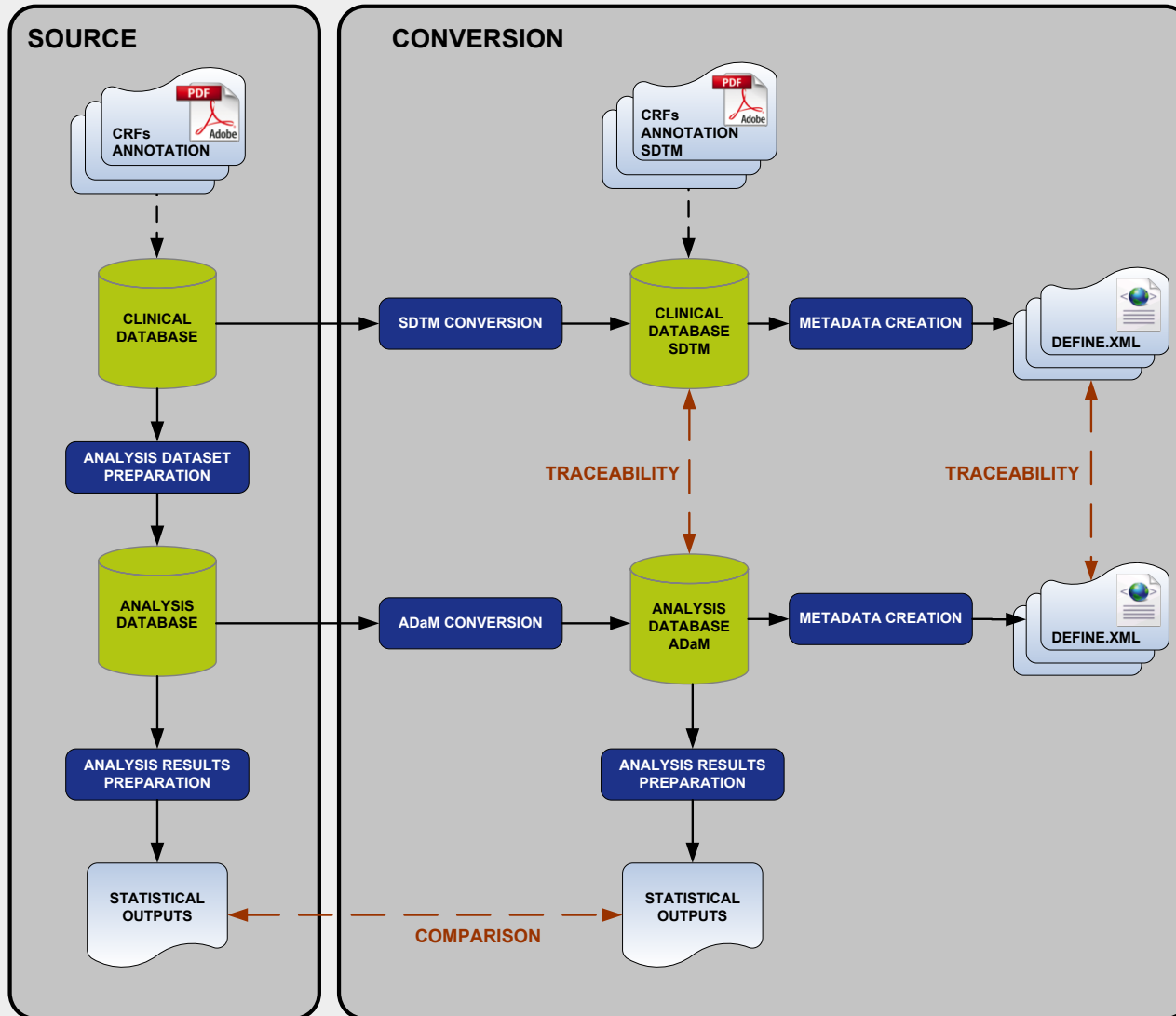
Challenges & Conclusion

- SDTM is expected to be « required for FDA submission » within 2 years
 - CDER is accepting SDTM submissions
 - CBER is accepting SDTM submissions since May 2010
 - CDRH interest is rising, CDISC SDTM team has formed a medical devices subteam
- FDA CDER:
 - Requesting sponsors to submit in SDTM format
 - Encouraging sponsors to submit in ADaM format
- Continuous FDA pilot projects, both CDER and CBER

Implementation approaches: strategy 1



Implementation approaches: strategy 2



- Understanding relationship between the analysis results, the analysis datasets and the SDTM domains
- Establishing the path between an element and its immediate predecessor
- Two levels:
 - Metadata traceability
 - Relationship between an analysis result and analysis dataset(s)
 - Relationship of the analysis variable to its source dataset(s) and variable(s)
 - Data point traceability
 - Predecessor record(s)

SDTM aCRF

• Patient Demographics – Part I

Patient X5 Page 4 (Demo_V1a for Visit 1a) Page 1 of 1.

Visit Date **11-Dec-2007** Blank ☐ Comment

PATIENT DEMOGRAPHICS - Part I **DSCAT = "PROTOCOL MILESTONE"**

Informed consent was obtained on **DSSTDT**

Gender **SEX** ☐ 1 = male, 2 = female

Date of birth **BRTHDTC** Age **AGE** years **AGEU**

(Age is automatically calculated when screen is saved and closed)

Height cm

Weight kg

Waist circumference cm

VSORRES / VSORRESU where VSTESTCD = "HEIGHT", "WEIGHT", "WAIST"

SDTM define.xml

Vital Signs Dataset (VS)						..Tabulations\vs.xpt
Variable	Label	Type	Controlled Terminology	Origin	Role	Comment
STUDYID	Study Identifier	text		Protocol, CRF Page 1	IDENTIFIER	The STUDYID variable has a fixed format: 'XXXX-YYYY', where 'XXXX' indicates the 4-digit compound code and the 'YYYY' the 4-digit study code
DOMAIN	Domain Abbreviation	text	DOMAIN	Assigned	IDENTIFIER	
USUBJID	Unique Subject Identifier	text		Protocol	IDENTIFIER	The USUBJID variable has a fixed format: 'XXXX-YYYY-ZZZZZ', where 'XXXX' indicates the 4-digit compound code, 'YYYY' the 4-digit study code and 'ZZZZZ' the 6-digit patient code
VSSEQ	Sequence Number	integer	Derived		IDENTIFIER	Sequence number (automatically generated) to ensure uniqueness within a dataset for a subject
VSTESTCD	Vital Signs Test Short Name	text	Assigned		TOPIC	
VSTEST	Vital Signs Test Name	text	Derived		SYNONYM QUALIFIER	
VSTRESN	Numeric Result Finding in Standard Units	float	3.1	Derived	RESULT QUALIFIER	

Analysis Results

Table 1 Demographic Data - Per-Protocol

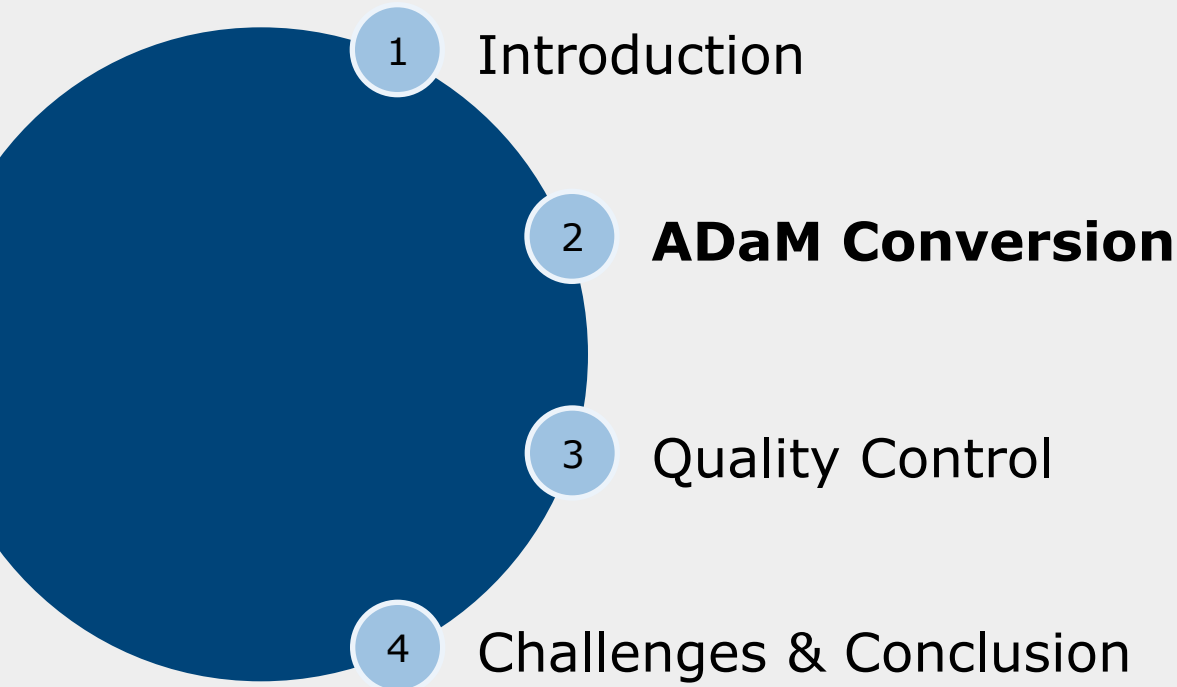
	Treatment 1	Treatment 2
Baseline body mass index (BMI) [kg/m**2]		
N	167	167
Mean	29.08	29.04
SD	4.84	4.80
Min	20.3	16.0
Median	28.69	28.47
Max	40.1	41.2
Baseline BMI (categorical) [N (%)]		
<25 kg/m**2	41 (24.6%)	71 (21.1%)
25-<30 kg/m**2	60 (35.9%)	130 (38.7%)
>=30 kg/m**2	66 (39.5%)	135 (40.2%)

Analysis Dataset

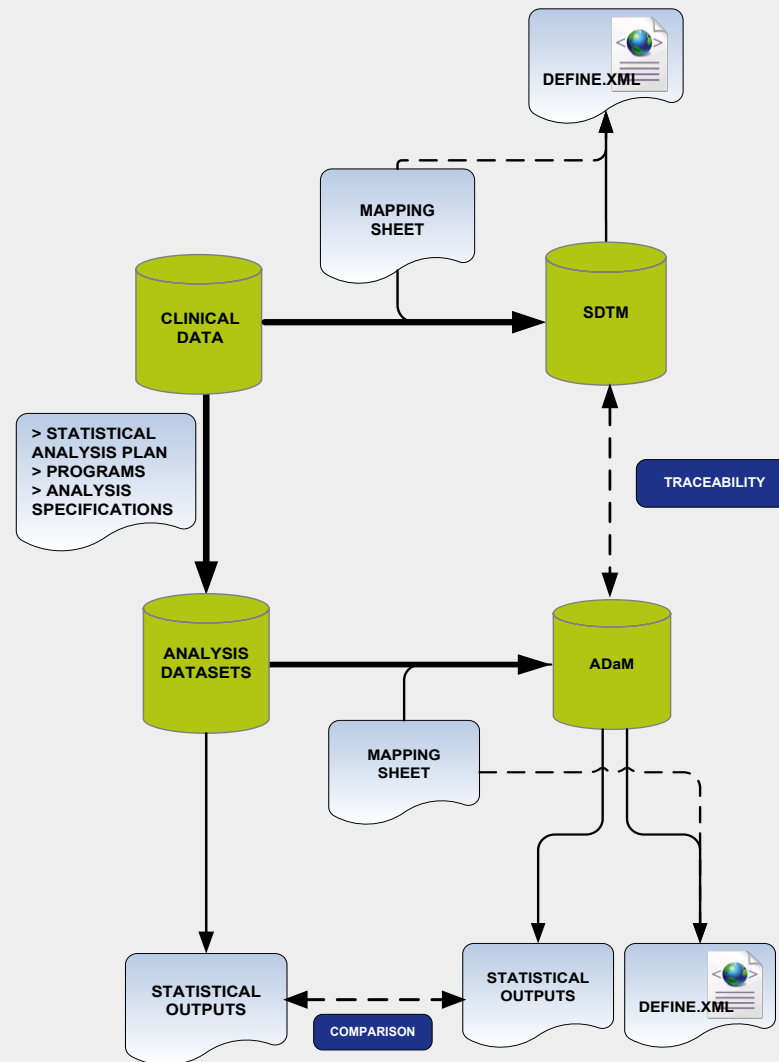
	STUDYID	USUBJID	SUBJID	BMI	BMIGR1	BMIGR1N	BMIGR2	BMIGR2N
2	9999-0001	9999-0001-000001	000001	27.77777778	<30 kg/m**2		1 25-<30 kg/m**2	2
3	9999-0001	9999-0001-000002	000002	25.503615702	<30 kg/m**2		1 25-<30 kg/m**2	2
4	9999-0001	9999-0001-000003	000003	26.175194521	<30 kg/m**2		1 25-<30 kg/m**2	2
5	9999-0001	9999-0001-000004	000004	35.15625	>=30 kg/m**2		2 >=30 kg/m**2	3
6	9999-0001	9999-0001-000005	000005	30.968858131	>=30 kg/m**2		2 >=30 kg/m**2	3
7	9999-0001	9999-0001-000006	000006	39.697163916	>=30 kg/m**2		2 >=30 kg/m**2	3
8	9999-0001	9999-0001-000007	000007	25.826446281	<30 kg/m**2		1 25-<30 kg/m**2	2
9	9999-0001	9999-0001-000008	000008	30.103806228	>=30 kg/m**2		2 >=30 kg/m**2	3
10	9999-0001	9999-0001-000009	000009	32.280962683	>=30 kg/m**2		2 >=30 kg/m**2	3
11	9999-0001	9999-0001-000010	000010	28.876133787	<30 kg/m**2		1 25-<30 kg/m**2	2
12	9999-0001	9999-0001-000011	000011	29.372397383	<30 kg/m**2		1 25-<30 kg/m**2	2
13	9999-0001	9999-0001-000012	000012	26.714852608	<30 kg/m**2		1 25-<30 kg/m**2	2
14	9999-0001	9999-0001-000013	000013	32.718619869	>=30 kg/m**2		2 >=30 kg/m**2	3
15	9999-0001	9999-0001-000014	000014	28.719723183	<30 kg/m**2		1 25-<30 kg/m**2	2
16	9999-0001	9999-0001-000015	000015	32.270420377	>=30 kg/m**2		2 >=30 kg/m**2	3

ADaM define.xml

Subject-Level Analysis Dataset Dataset (ADSL)						..Analysis\adsl.xpt	
Variable	Label	Type	Controlled Terms or Format	Computational Algorithm or Method	Origis	Role	Comment
STUDYID	Study Identifier	text			DM	IDENTIFIER	The STUDYID variable has a fixed format: 'XXXX-YYYY', where 'XXXX' indicates the 4-digit compound code and the 'YYYY' the 4-digit study code
USUBJID	Unique Subject Identifier	text			DM	IDENTIFIER	The USUBJID variable has a fixed format: 'XXXX-YYYY-ZZZZZ', where 'XXXX' indicates the 4-digit compound code, 'YYYY' the 4-digit study code and 'ZZZZZ' the 5-digit patient code
SUBJID	Subject Identifier for the Study	text			DM	IDENTIFIER	
HEIGHT	Baseline Height (cm)	integer		ADSL HEIGHT	Derived	ANALYSIS	
WEIGHT	Baseline Weight (kg)	integer		ADSL WEIGHT	Derived	ANALYSIS	
BMI	Baseline BMI (kg/m**2)	integer		ADSL BMI	Derived	ANALYSIS	
BMIGR1	Category 1 of Baseline BMI	text	BMIGR1L	ADSL BMIGR1	Derived	ANALYSIS	
BMIGR1N	Category 1 of Baseline BMI (N)	integer	BMIGR1N	ADSL BMIGR1N	Derived	ANALYSIS	
BMIGR2	Category 2 of Baseline BMI	text	BMIGR2L	ADSL BMIGR2	Derived	ANALYSIS	
BMIGR2N	Category 2 of Baseline BMI (N)	integer	BMIGR2N	ADSL BMIGR2N	Derived	ANALYSIS	



 Business & Decision
Life Sciences **Boehringer**
Ingelheim



- Submission included 11 trials
- For each trial:
 - ADSL (Subject Level Analysis Dataset)
 - AD with baseline conditions
 - AD with treatment administration
 - AD with efficacy endpoints
- For some trials:
 - 2 Pharmacokinetic datasets

- **CRO Manager**
 - CDISC expert support
- **Project Manager**
Project Manager back-up
 - Assigned for the duration of the project
 - Single point of contact
- **Mappers (4)**
 - ADaM experts
 - Define mapping
 - Investigate traceability
- **Programmers (2.5)**
 - Create the conversions programs
 - Perform peer review
- **Data Steward (0.5)**
 - Maintains the consistency across the project
- **Quality Checker (4)**
 - Perform ADaM datasets review
 - Perform define.xml review



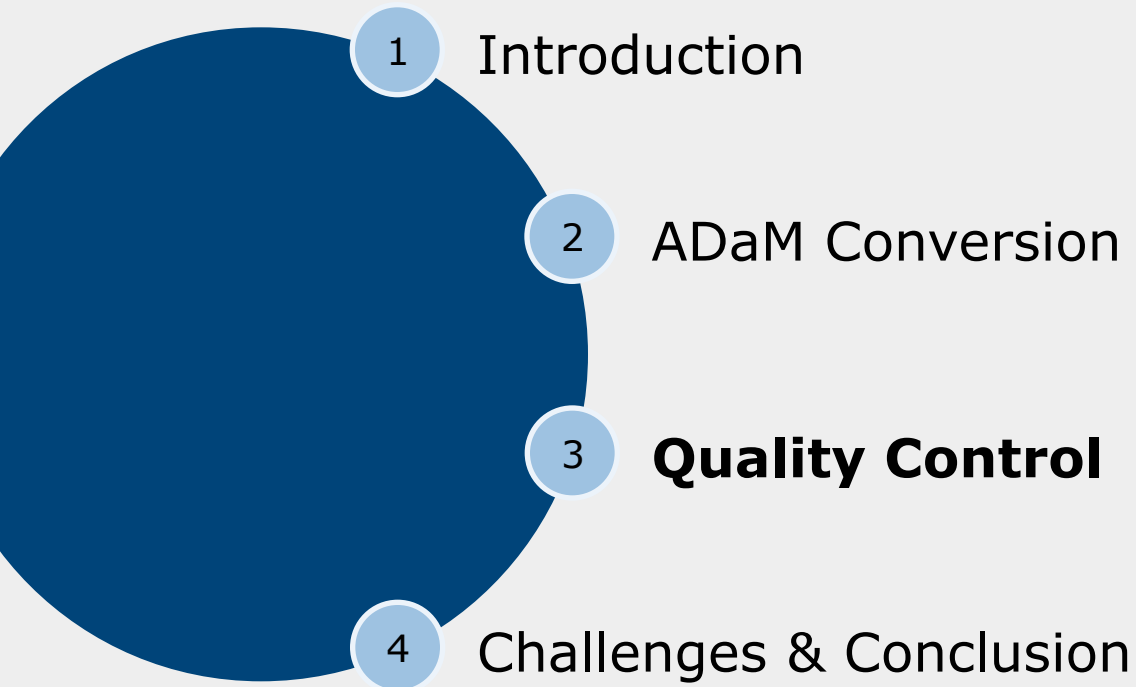
- **Creation of SDTM variables**
 - Variables like USUBJID which were created during the SDTM conversion
- **Minor conversion**
 - Contents unchanged, metadata changes
 - Change variable name and label of the age group variable
- **Format values**
 - Content and metadata changes
 - The content of the SEX variable had to be changed in order to reflect the SDTM values
- **Transpose**
 - Observations become variables
 - Populations in the ADSL dataset

- **Variables originating from SDTM**

- SDTM variables are retained in ADaM ADs for traceability
- SDTM variables are unchanged
 - same name, same type, same label (metadata)
 - and same content (data)

- **Derived variables**

- Original computational algorithm for derived AD variable(s) based on original clinical database
- New computational algorithm needs to be based on SDTM database
- New computational algorithm is included into ADaM define.xml



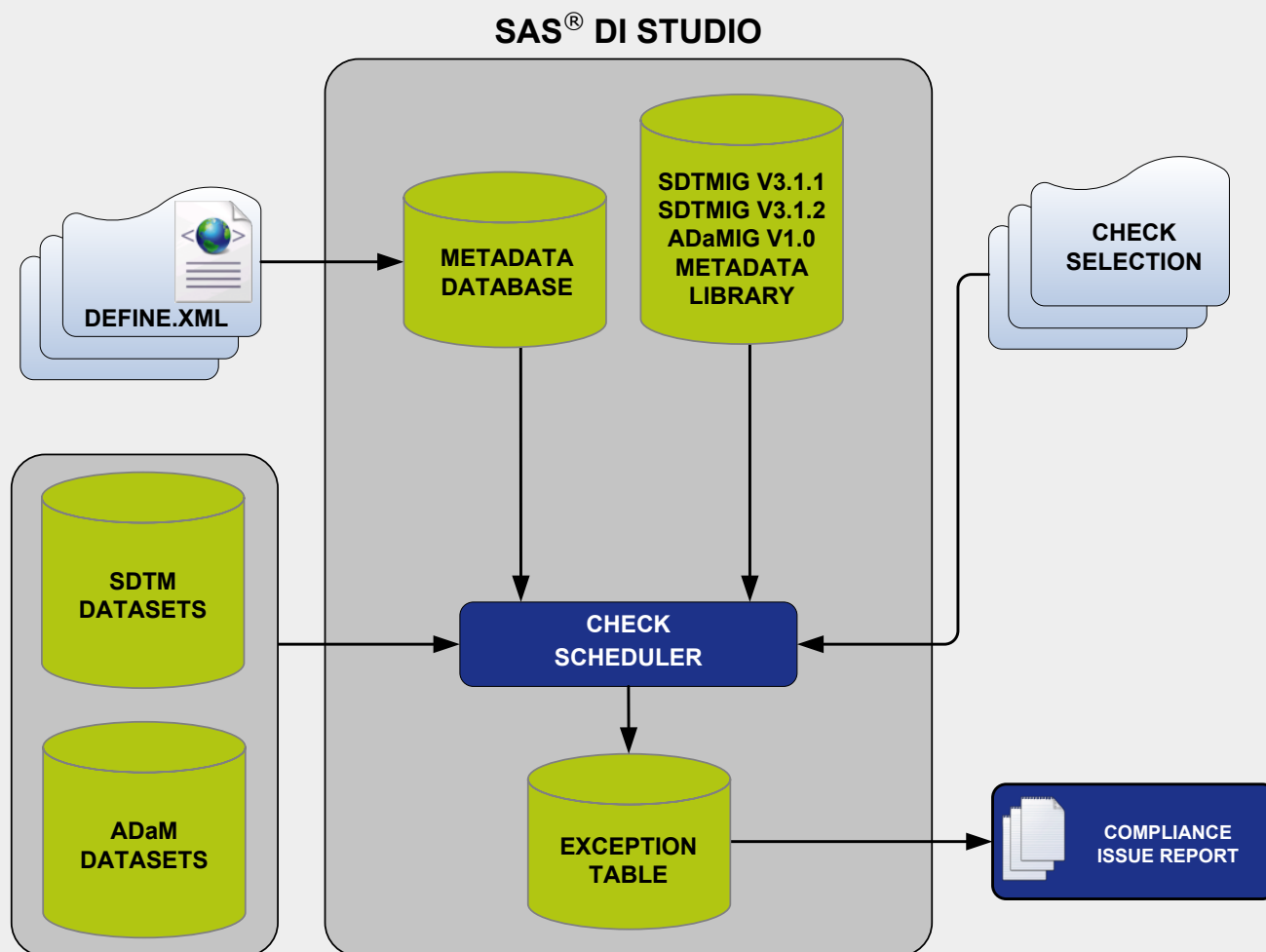
- **QC is partially automated**
 - Electronic QC (CDISC Compliance Checks – SDTM&ADaM)
 - Manual QC
 - QC on Consistency (Data Steward)
- **QC on:**
 - Mapping
 - ADaM Datasets
 - Define.xml
 - Statistical Results
- **QC is supported by documentation**

We have created an expanded & enhanced list of checks

- 154 WebSDM TM checks
- Total check package:

	SDTMIG V3.1.1	SDTMIG V3.1.2	ADaMIG V1.0
Data checks	141	219	45
Metadata checks	68	117	51
Mapping checks	56	57	12
Project consistency checks	20	20	20

CDISC compliance checks list is growing continuously



- 100% manual QC on a random sample
- Supported by checklists
- Supported by a QC content tool on source and target

DATA.ADSL

Created: 08APR10:08:52:02
obs : 503
Keys :

#	Variable	Label	Type	Length	Contents
1	STUDYID	Study Identifier	CHAR	200	+ 1 distinct values, 0 missing values, all shown
2	USUBJID	Unique Subject Identifier	CHAR	200	+ 503 distinct values, 0 missing values, all shown
3	SUBJID	Subject Identifier for the Study	CHAR	200	+ 503 distinct values, 0 missing values, all shown
4	SITEID	Study Site Identifier	CHAR	200	+ 63 distinct values, 0 missing values, all shown
5	PTNOTYP	Type of trial participant	CHAR	200	+ 1 distinct values, 0 missing values, all shown
6	INVTNAM	Investigator Name	CHAR	200	+ 62 distinct values, 0 missing values, all shown
7	COUNTRY	Country	CHAR	200	+ 11 distinct values, 0 missing values, all shown
8	COUNTRDC	Country, Decode	CHAR	40	+ 11 distinct values, 0 missing values, all shown
9	REGION	Region	CHAR	40	+ 2 distinct values, 0 missing values, all shown
10	AGE	Age	NUM	8	+ 52 distinct values, 0 missing values, all shown Mean: 65.708 Minimum: 24 Maximum: 74 Missing values: 0
11	AGEU	Age Units	CHAR	200	+ 1 distinct values, 0 missing values, all shown
12	AGEGR1	Age Group 1	CHAR	20	+ 3 distinct values, 0 missing values, all shown 65-74 years (97x) <65 years (398x) >=75 years (8x)
13	AGEGR1N	Age Group 1, (N)	NUM	8	+ 3 distinct values, 0 missing values, all shown Mean: 1.225 Minimum: 1 Maximum: 3 Missing values: 0
14	AGEGR2	Age Group 2	CHAR	20	+ 5 distinct values, 0 missing values, all shown
15	AGEGR2N	Age Group 2, (N)	NUM	8	+ 5 distinct values, 0 missing values, all shown Mean: 2.646 Minimum: 1 Maximum: 5 Missing values: 0
16	AGEGR3	Age Group 3	CHAR	20	+ 2 distinct values, 0 missing values, all shown
17	AGEGR3N	Age Group 3, (N)	NUM	8	+ 2 distinct values, 0 missing values, all shown

11 distinct values, 0 missing values, all shown

Croatia	(17x)
India	(135x)
Israel	(21x)
Italy	(15x)
Malaysia	(72x)
Netherlands	(33x)
Poland	(11x)
Romania	(24x)
Slovakia	(62x)
Thailand	(24x)
Ukraine	(89x)

3 distinct values, 0 missing values, all shown

65-74 years	(97x)
<65 years	(398x)
>=75 years	(8x)

- **Maintains consistency of metadata across project**
- **Uses the metadata repository**
- **Electronic consistency checks**

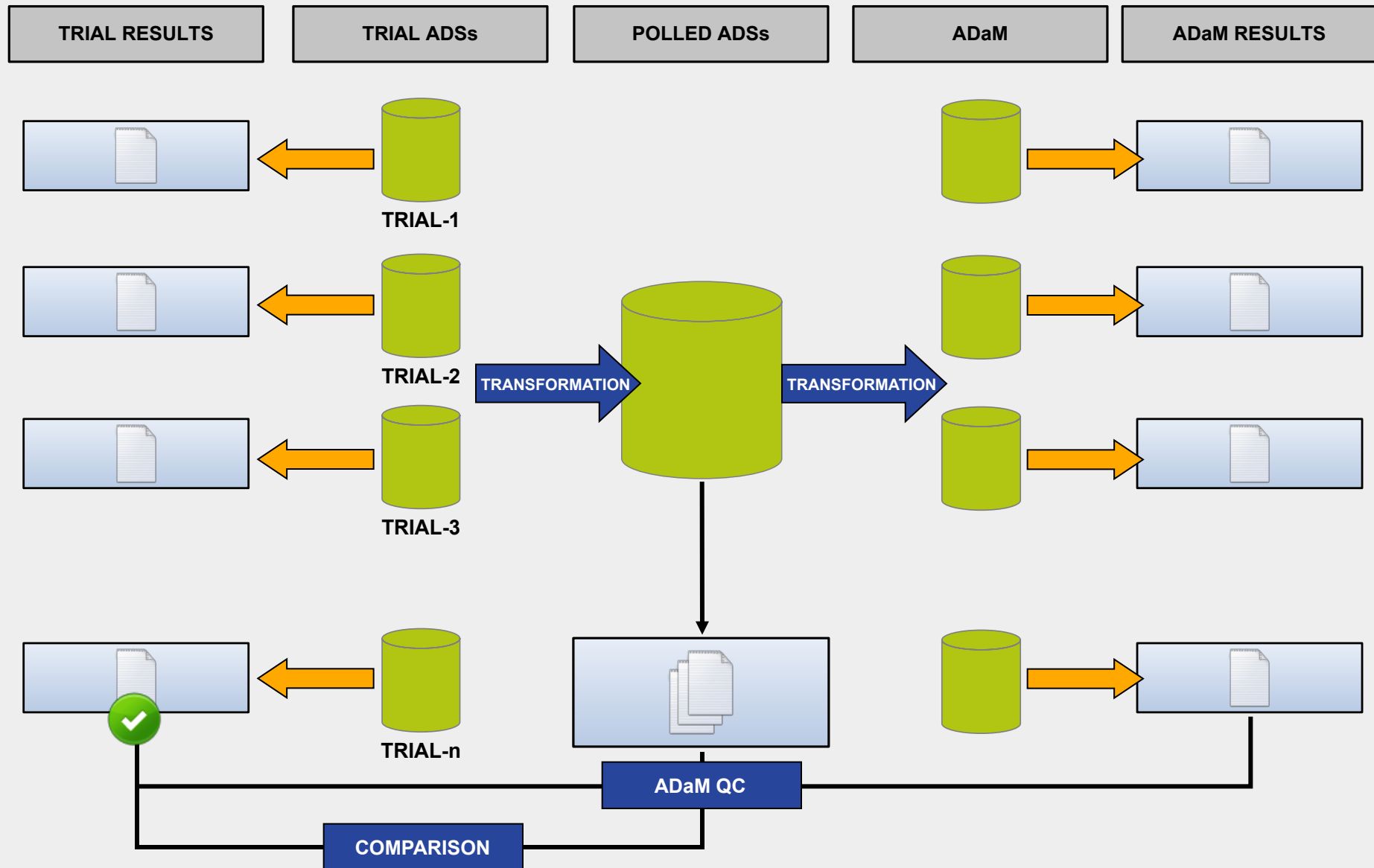
Table of Contents

1. DATASET
2. DATASET DESCRIPTION
3. DATASET CLASS
4. DATASET STRUCTURE
5. DATASET PURPOSE
6. DATASET SORTING KEYS

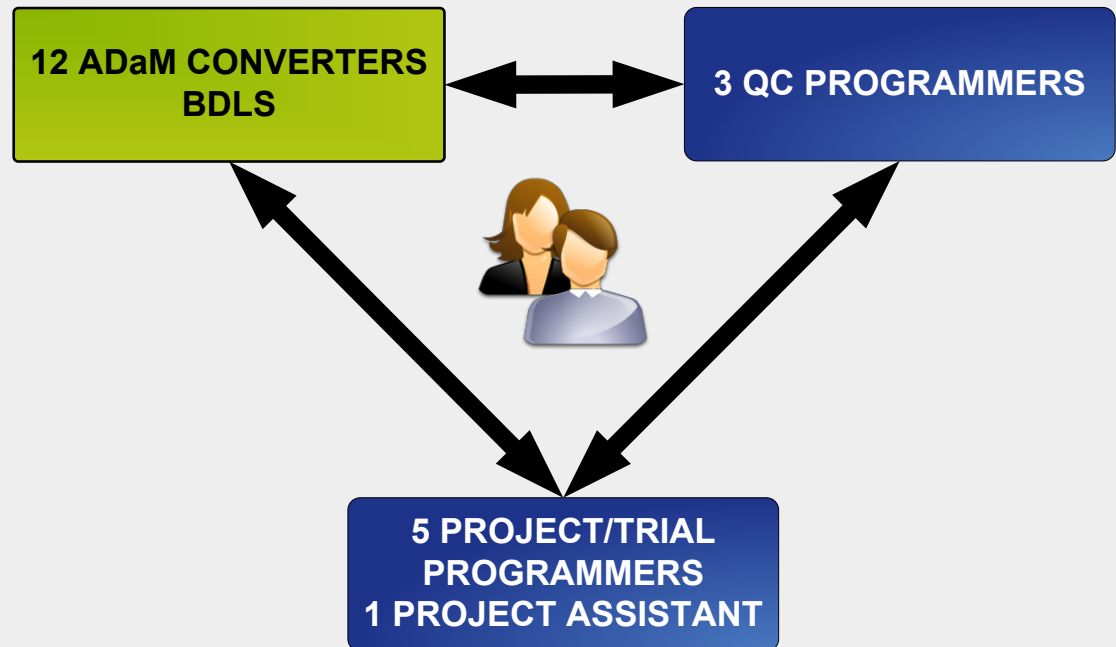
VARIABLE LABEL

DOMAIN	VARIABLE	VARIABLE LABEL	S1218_0005	S1218_0006	S1218_0017	GLBSTND
ADSL	FASCOFN	FAS Completers Population Flag, (N)	X	X		.
ADSL	FASCOFN	FAS Completers Population Flag, Num			X	.
ADSL	FASFN	Full Analysis Set Population Flag, (N)	X	X		.
ADSL	FASFN	Full Analysis Set Population Flag, Num			X	.
ADSL	FASMTTFN	FAS MTT Population Flag, (N)	X	X		.
ADSL	FASMTTFN	FAS MTT Population Flag, Num			X	.
ADSL	PPROTFN	Per-Protocol Population Flag, (N)	X	X		.
ADSL	PPROTFN	Per-Protocol Population Flag, Num			X	.
ADSL	RACEGR1	Pooled Race Group 1			X	.
ADSL	RACEGR1N	Pooled Race Group 1 (N)			X	.
ADSL	RANDFN	Randomized Population Flag, (N)	X	X		.

QC Tier 4: Statistical Results



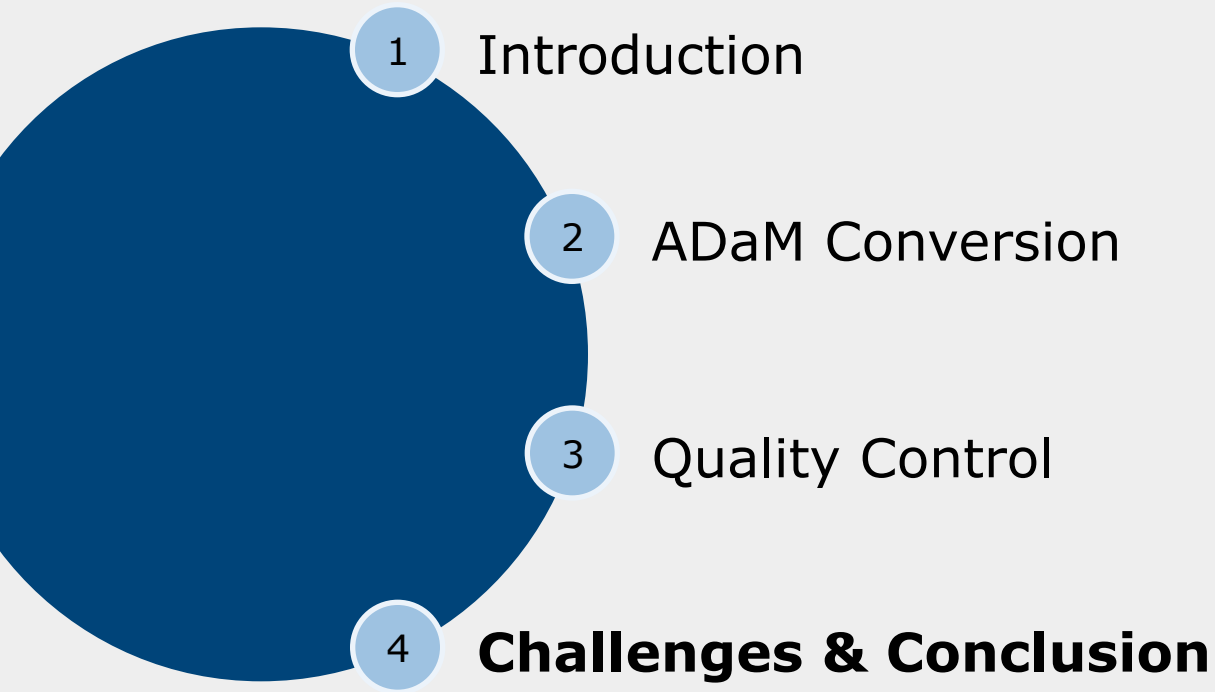
- **Project-/Trial Programmer (3)**
 - Coordination
 - Single point of contact
- **Project Statistician (1)**
 - Specifications of results subject to QC
- **QC Programmers (3)**
 - Re-production of statistical results



- Compilation of selected result-tables
 - ~ 55 table types
 - ~ 220 tables
 - mainly descriptive statistics
 - few inferential statistics (ANCOVA)
- Set-up of work environment
 - e.g. directories, access rights
- Learning the project, trials
- QC Programming
 - Recreate results from CTR / ISE
 - Based on Pooled BI Analysis Datasets (initially)
 - Based on ADaM (once available)
- Documenting QC progress
- Comparison of results

- Report Source Data Issues
 - Empty variables
 - Exclusion of screen failures
 - Unclear computational algorithms
 - Traceability issues with SDTM
- Sponsor Feedback
 - Clarifications computational algorithms
 - QC comments

- Addressing and solving issues and deciding further proceedings in
 - weekly T*C with representatives from each of the 3 subteams
 - daily brief QC Programmers meeting
- Communication was:
 - Timely and immediate
 - Focused
 - For some last minute changes to ADaM, communication was not effective
 - e.g. renaming of variables
 - data changes due to B&D Life Sciences QC, e.g. indicator variables



- Learning the project / trials
- Understanding original analysis datasets and computational algorithms
- Finding all QC relevant result tables
 - Initially some wrong tables selected
 - Transformation from trial to pooled ADs not clearly documented
- This type of project is always on critical path for a submission
 - Short timelines
 - Large team

- We now understand better how FDA feels
- SDTM is the basis for analysis and therefore needs to be complete
- Results in the clinical study report must be reproducible by FDA reviewers from the newly created ADaM analysis datasets
- Traceability most difficult part in ADaM conversion
- Familiarization with usage of ADaM for programming was minimal
 - Due to similarity of ADaM with BI-ADs structure
- Relatively straightforward to program from ADaM
- In an ideal world, analysis datasets are created from SDTM datasets, thereby ensuring 100% traceability