

Traceability between SDTM and ADaM converted analysis datasets

Florence Somers
Manager Statistical Programming
Business & Decision Life Sciences

Tel +32 2 774 11 00
Fax +32 2 774 11 99
Mobile +32 491 15 37 47
florence.somers@businessdecision.com

Sint-Lambertusstraat 141
1200 Brussels

www.businessdecision-lifesciences.com



DRIVE YOUR PERFORMANCE[®]

Topics

- 
- 1 Introduction
 - 2 ADaM Conversion
 - 3 Quality Control
 - 4 Challenges & Conclusion

Topics

1

Introduction

2

ADaM Conversion

3

Quality Control

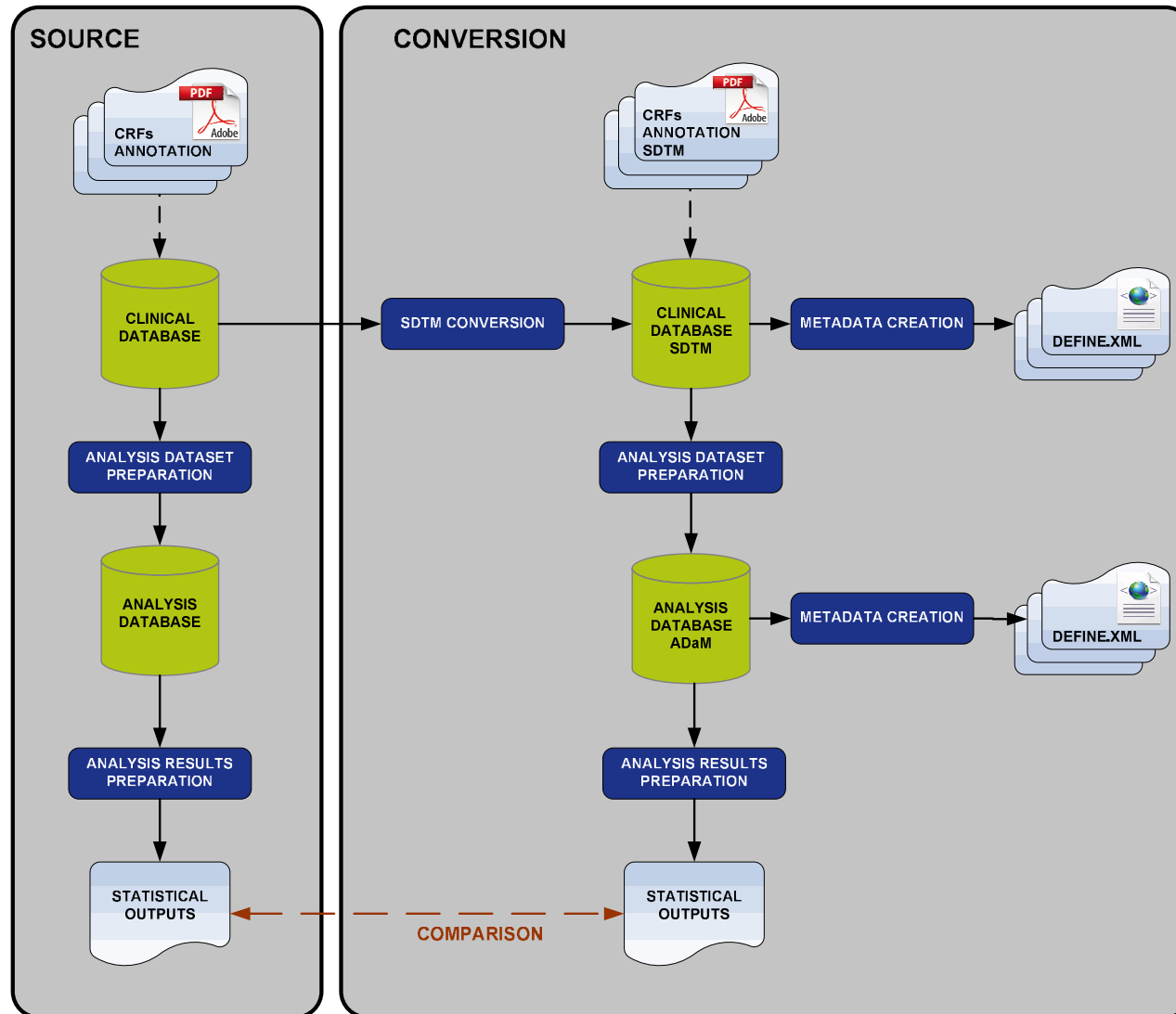
4

Challenges & Conclusion

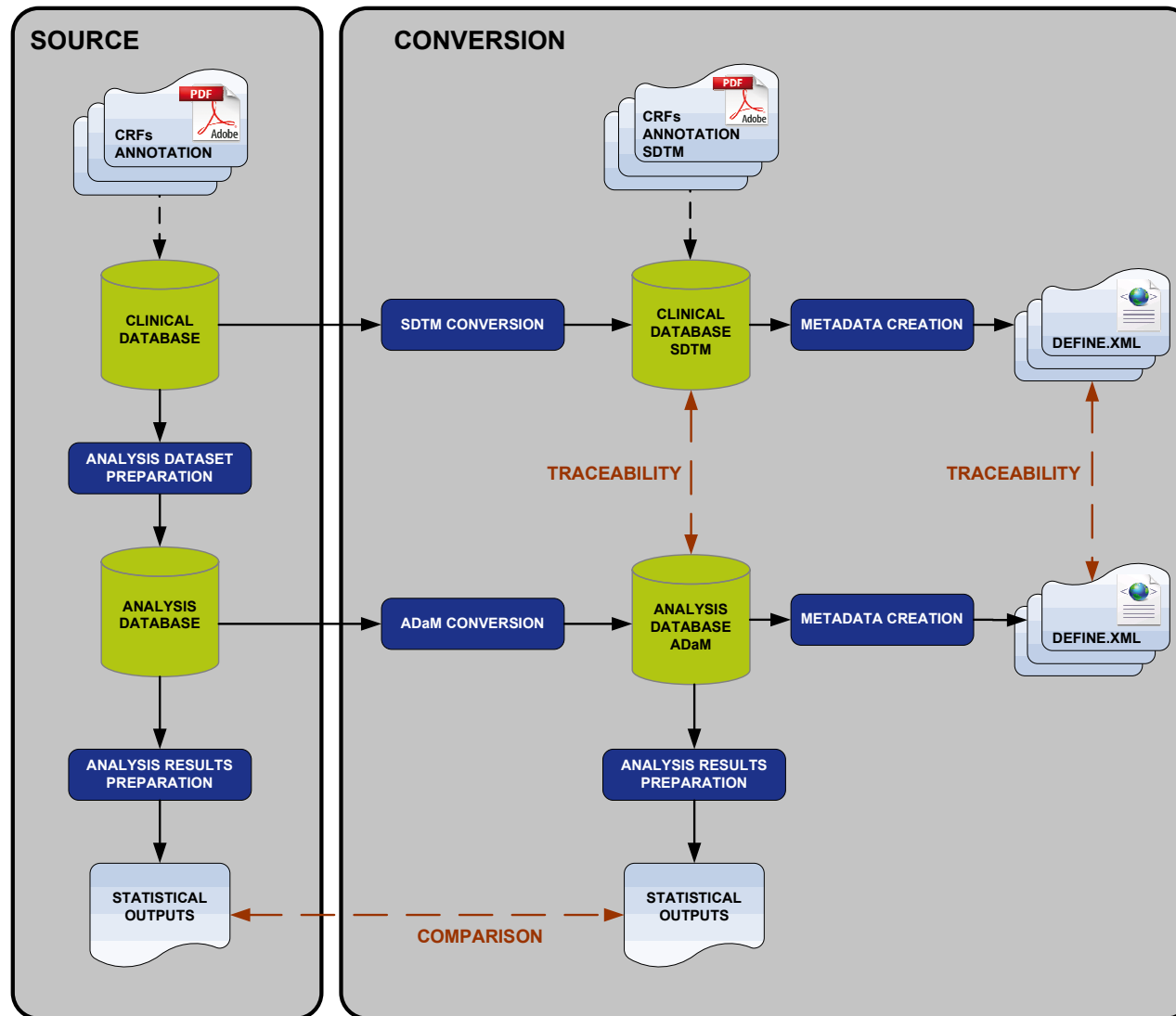
SDTM/ADaM adoption by FDA

- 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



Traceability SDTM and ADaM

- 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)

Traceability SDTM and ADaM

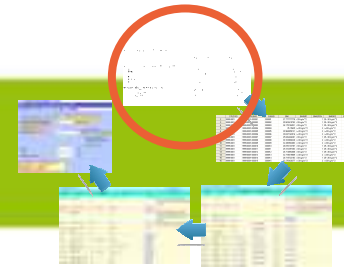
Table 1 Demographic Data - Per-Protocol

	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%)

	STUDYID	USUBID	SUBID	BMI	BMGR1	BMGRN	BMGR2	BMGRZ
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	25.50361502	<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	30.95859191	>=30 kg/m**2		2 >=30 kg/m**2	3
7	9999-0001	9999-0001-000006	000006	39.627632916	>=30 kg/m**2		2 >=30 kg/m**2	3
8	9999-0001	9999-0001-000007	000007	25.825446281	<30 kg/m**2		1 25-<30 kg/m**2	2
9	9999-0001	9999-0001-000008	000008	30.103866228	>=30 kg/m**2		2 >=30 kg/m**2	3
10	9999-0001	9999-0001-000009	000009	32.283862683	>=30 kg/m**2		2 >=30 kg/m**2	3
11	9999-0001	9999-0001-000010	000010	28.976133787	<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	26.714852628	<30 kg/m**2		1 25-<30 kg/m**2	2
14	9999-0001	9999-0001-000013	000013	32.719619869	>=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

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	
USUBID	Unique Subject Identifier	text		Protocol	IDENTIFIER	The USUBID 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.
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	
VSPPOS	Vital Signs Position of Subject	text	POSITION	CRF Page 11	RECORD QUALIFIER	
VSORRES	Result or Finding in Original Units	text		Derived, CRF Page 2, 11	RESULT QUALIFIER	
VSORRESU	Original Units	text	VSRSU	CRF Page 9, 11	VARIABLE QUALIFIER	
VSSTRSC	Character Result Finding in Std Format	text		Derived	RESULT QUALIFIER	
VSSTRSCN	Numeric Result Finding in Standard Units	float	3.1	Derived	RESULT QUALIFIER	

Traceability SDTM and ADaM

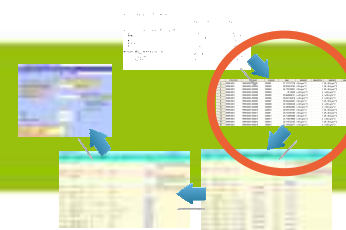


- Analysis Results

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%)

Traceability SDTM and ADaM



- 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

Traceability SDTM and ADaM

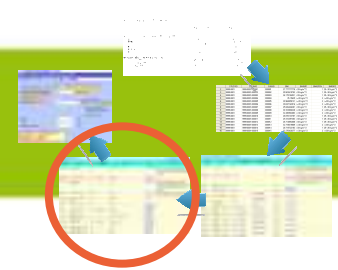


- ADaM define.xml

Computational Algorithms (ADSL.BMI)	
Reference Name	Computation Method
ADSL.BMI	Continuous variable, calculated using $ADSL.WEIGHT / (ADSL.HEIGHT * 0.01) ** 2$ value at visit 3 if visit 3 data not available, the last data collected before randomisation
Computational Algorithms (ADSL.HEIGHT)	
Reference Name	Computation Method
ADSL.HEIGHT	equal to VS.VSSTRESN when VS.VSTESTCD="HEIGHT"
Computational Algorithms (ADSL.WEIGHT)	
Reference Name	Computation Method
ADSL.WEIGHT	equal to VS.VSSTRESN when VS.VSTESTCD="WEIGHT" and VS.VISITNUM=30 if visit 3 data not available, the last data collected before randomisation

WEIGHT	Baseline Weight (kg)	integer		ADSL.HEIGHT	Derived	ANALYSIS
BMI	Baseline BMI (kg/m**2)	integer		ADSL.WEIGHT	Derived	ANALYSIS
BMIGR1	Category 1 of Baseline BMI	text	BMIGR1L	ADSL.BMI	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
BMIGR3	Category 3 of Baseline BMI	text	BMIGR3L	ADSL.BMIGR3	Derived	ANALYSIS
BMIGR3N	Category 3 of Baseline BMI, (N)	integer	BMIGR3N	ADSL.BMIGR3N	Derived	ANALYSIS

Traceability SDTM and ADaM



- SDTM define.xml and aCRF

Value Level Metadata (ValueList.VS.VSTESTCD)							
Source Variable	Value	Label	Type	Controlled Terminology	Origin	Role	Comment
VSTESTCD	DIABP	DIASTOLIC BLOOD PRESSURE	text		CRF Page 13		
VSTESTCD	HEIGHT	HEIGHT	text		CRF Page 9		
VSTESTCD	PULSE	PULSE RATE	text		CRF Page 13		
VSTESTCD	SYSBP	SYSTOLIC BLOOD PRESSURE	text		CRF Page 13		
VSTESTCD	WAIST	WAIST CIRCUMFERENCE	text		CRF Page 9		
VSTESTCD	WEIGHT	WEIGHT	text		CRF Page 9		

VSSEQ	Sequence Number					patient code
VSTESTCD	Vital Signs Test Short Name					Sequence number (automatically generated) to ensure uniqueness within a dataset for a subject
VSTEST	Vital Signs Test Name					
VSPOR	Vital Signs Position of Subject					
VSORRES	Result or Finding in Original Units					
VSORRESU	Original Units					
VSSSTRESC	Character Result Finding in Std Format					
VSSSTRESN	Numeric Result Finding in Standard Units					

• 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"

Topics

1

Introduction

2

ADaM Conversion

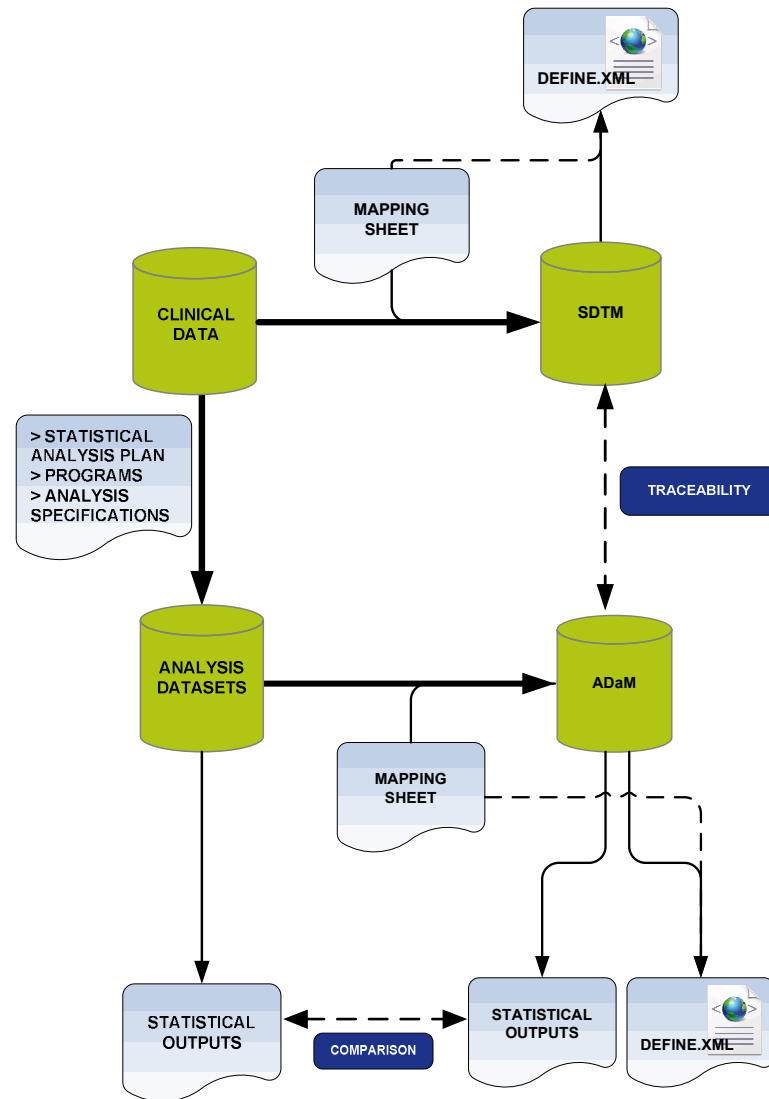
3

Quality Control

4

Challenges & Conclusion

ADaM Conversion



Team Profile and Roles

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



Conversion Types

- Creation of SDTM variables

ORIGINAL AD

	Trial number	Patient number
1	9999-0001	1
2	9999-0001	1
3	9999-0001	1
4	9999-0001	2
5	9999-0001	2
6	9999-0001	2
7	9999-0001	3
8	9999-0001	3
9	9999-0001	3

ADaM AD

	Study Identifier	Unique Subject Identifier	Subject Identifier for the Study
1	9999-0001	9999-0001-000001	000001
2	9999-0001	9999-0001-000002	000002
3	9999-0001	9999-0001-000003	000003

Conversion Types

- Minor conversion
 - Contents unchanged, metadata changes

ORIGINAL AD

	STUDY	PTNO	BAGE	BAGEC1	BAGECDC1	BAGEC2	BAGECDC2
1	9999-0001	1	74	2	65-74 years	4	65-<75 years
2	9999-0001	2	46	1	<65 years	2	45-<55 years
3	9999-0001	3	60	1	<65 years	3	55-<65 years

ADaM AD

	STUDYID	USUBJID	SUBJID	AGE	AGEU	AGEGR1	AGEGR1N	AGEGR2	AGEGR2N
1	9999-0001	9999-0001-000001	000001	74	YEARS	65-74 years	2	65-<75 years	4
2	9999-0001	9999-0001-000002	000002	46	YEARS	<65 years	1	45-<55 years	2
3	9999-0001	9999-0001-000003	000003	60	YEARS	<65 years	1	55-<65 years	3

Conversion Types

- Format values
 - Content and metadata changes

ORIGINAL AD

	STUDY	PTNO	SEX	SEXDC
1	9999-0001	1	1	Male
2	9999-0001	2	2	Female
3	9999-0001	3	1	Male

ADaM AD

	STUDYID	USUBJID	SUBJID	SEX	SEXN	SEXDC
1	9999-0001	9999-0001-000001	000001	M	1	Male
2	9999-0001	9999-0001-000002	000002	F	2	Female
3	9999-0001	9999-0001-000003	000003	M	1	Male

Conversion Types

- Transpose
 - Observations become variables
 - Variables become observations

ORIGINAL AD

	Trial number	Patient number	Analysis set	Analysis set decode	Included in analysis set	Included in analysis set decode
1	9999-0001	1	FAS	Full analysis set	0	No
2	9999-0001	1	PPS	Per protocol set	0	No
3	9999-0001	1	RS	Randomised set	0	No
4	9999-0001	2	FAS	Full analysis set	1	Yes
5	9999-0001	2	PPS	Per protocol set	0	No
6	9999-0001	2	RS	Randomised set	1	Yes
7	9999-0001	3	FAS	Full analysis set	1	Yes
8	9999-0001	3	PPS	Per protocol set	1	Yes
9	9999-0001	3	RS	Randomised set	1	Yes

ADaM AD

	Study Identifier	Unique Subject Identifier	Subject Identifier for the Study	Full Analysis Set Population Flag	Full Analysis Set Population Flag, (N)	Per-Protocol Population Flag	Per-Protocol Population Flag, (N)	Randomized Population Flag	Randomized Population Flag, (N)
1	9999-0001	9999-0001-000001	000001	N	0 N	0	0 N	0	0
2	9999-0001	9999-0001-000002	000002	Y	1 N	0	0 Y	0	1
3	9999-0001	9999-0001-000003	000003	Y	1 Y	1	1 Y	1	1

Traceability

- 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

Topics

- 
- 1 Introduction
 - 2 ADaM Conversion
 - 3 Quality Control**
 - 4 Challenges & Conclusion

Quality Control

- **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
- **QC is supported by documentation**
- **For each study a Data Handling Report is generated**

QC Tier 1: CDISC Compliance Checks

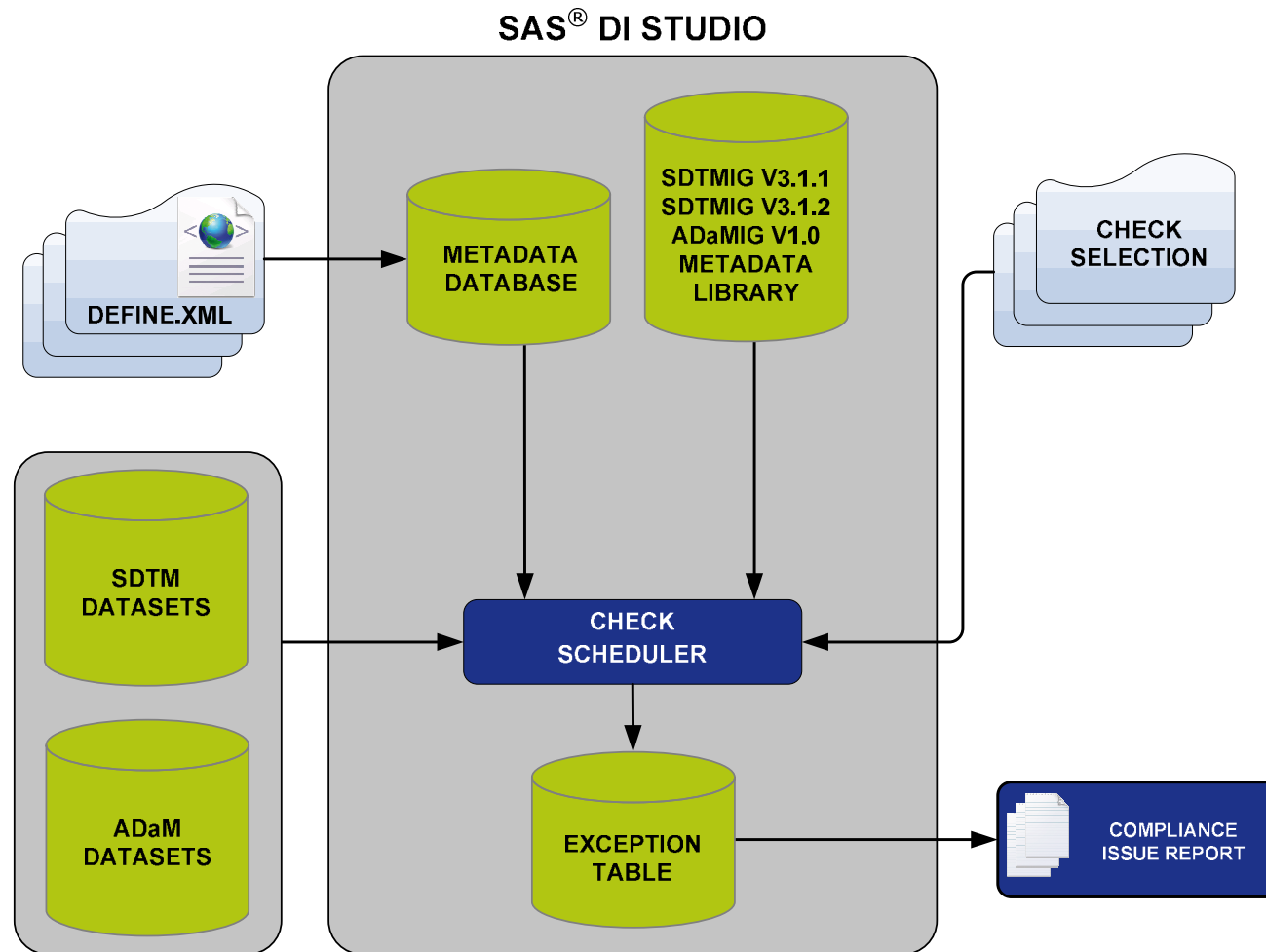
We have created an expanded & enhanced list of checks

- 154 WebSDM™ 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

QC Tier 1: Application Flowchart



QC Tier 2: Manual QC

- **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 Croatia (17x) India (135x) Israel (21x) Italy (15x) Malaysia (72x) Netherlands (33x) Poland (11x) Romania (24x) Slovakia (62x) Thailand (24x) Ukraine (89x)
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: 55.708 Minimum: 24 Maximum: 78 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 >=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

QC Tier 3: Data Steward

- **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		.

Communication Topics

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

Topics

1

Introduction

2

ADaM Conversion

3

Quality Control

4

Challenges & Conclusion

Challenges

- Understanding original analysis datasets and computational algorithms
 - Understanding derived variables in original ADs
- This type of project is always on critical path for a submission
 - Short timelines
 - Large team

Conclusion

- 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
- In an ideal world, analysis datasets are created from SDTM datasets, thereby ensuring 100% traceability



LIFE SCIENCES

**Thank you
for your attention**

Florence Somers
Manager Statistical Programming
Business & Decision Life Sciences

Tel +32 2 774 11 00
Fax +32 2 774 11 99
Mobile +32 491 15 37 47
florence.somers@businessdecision.com

Sint-Lambertusstraat 141
1200 Brussels

www.businessdecision-lifesciences.com



DRIVE YOUR PERFORMANCE®