

Business & Decision Life Sciences

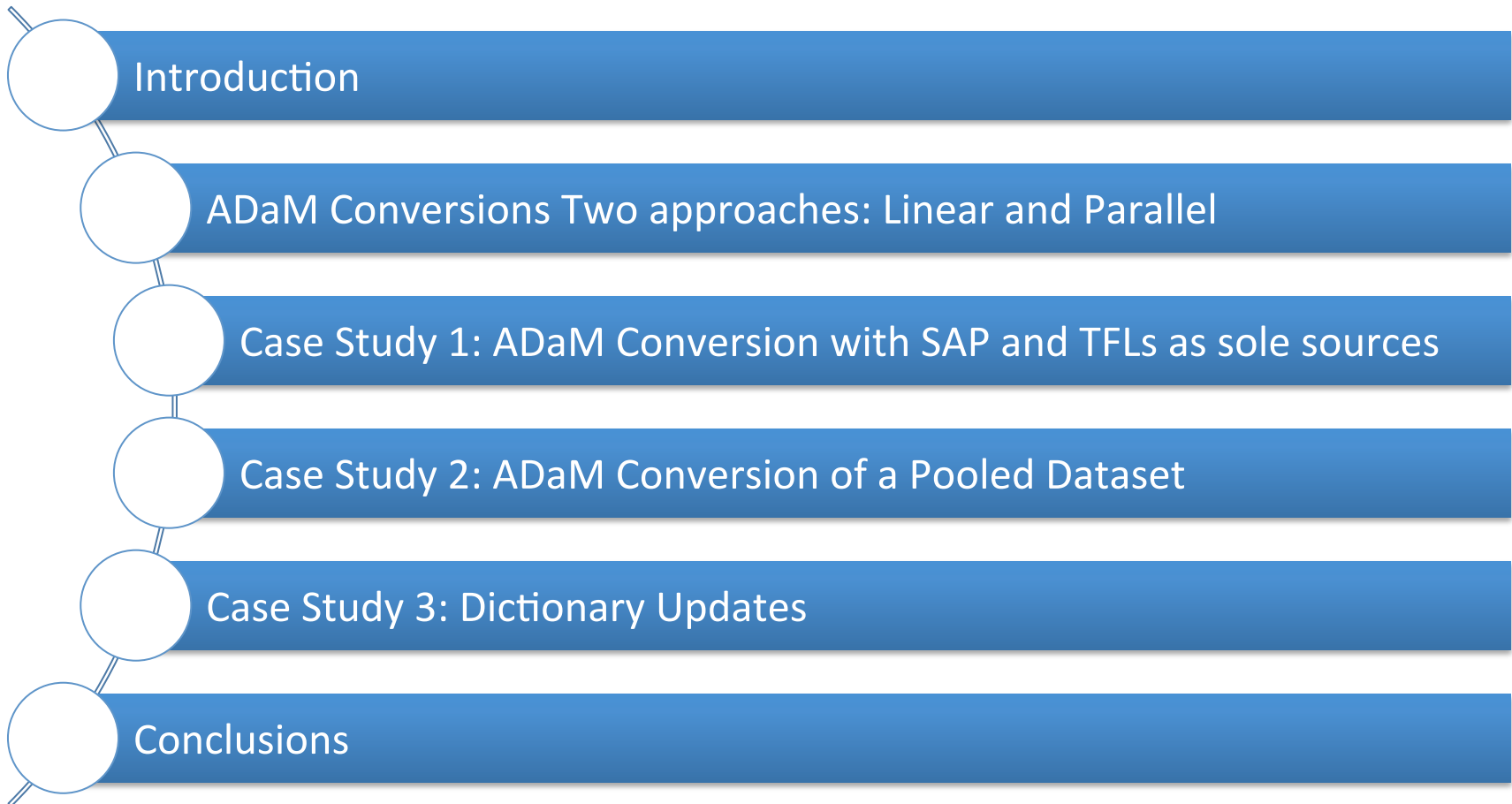
ADaM Conversions: The good, The bad and the Ugly

Carla Santillan, Jessica Minkue

18 September 2014

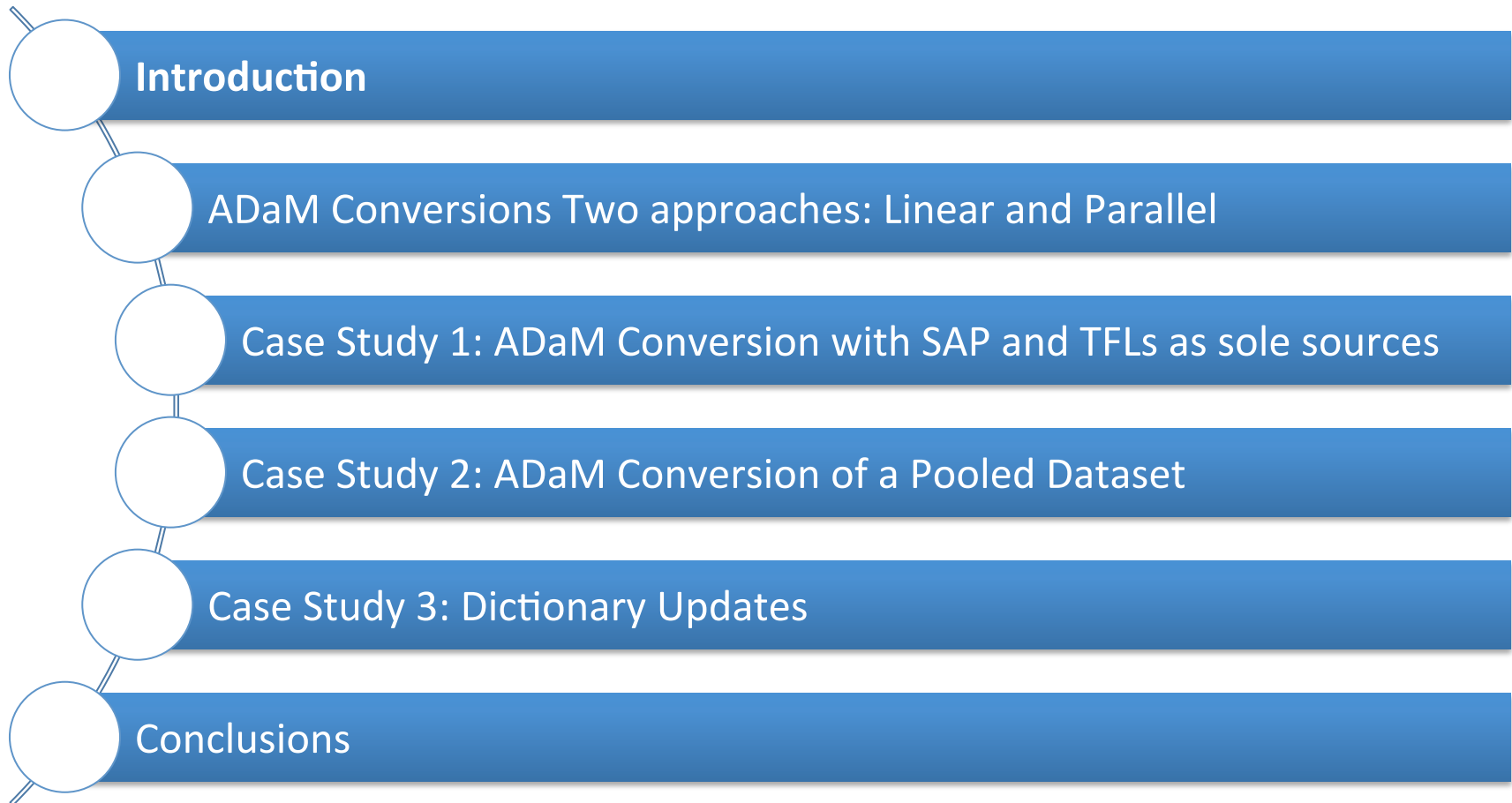


ADaM Conversions: The Good, the Bad and the Ugly



ADaM Conversion:

The good, the bad and the ugly



Introduction



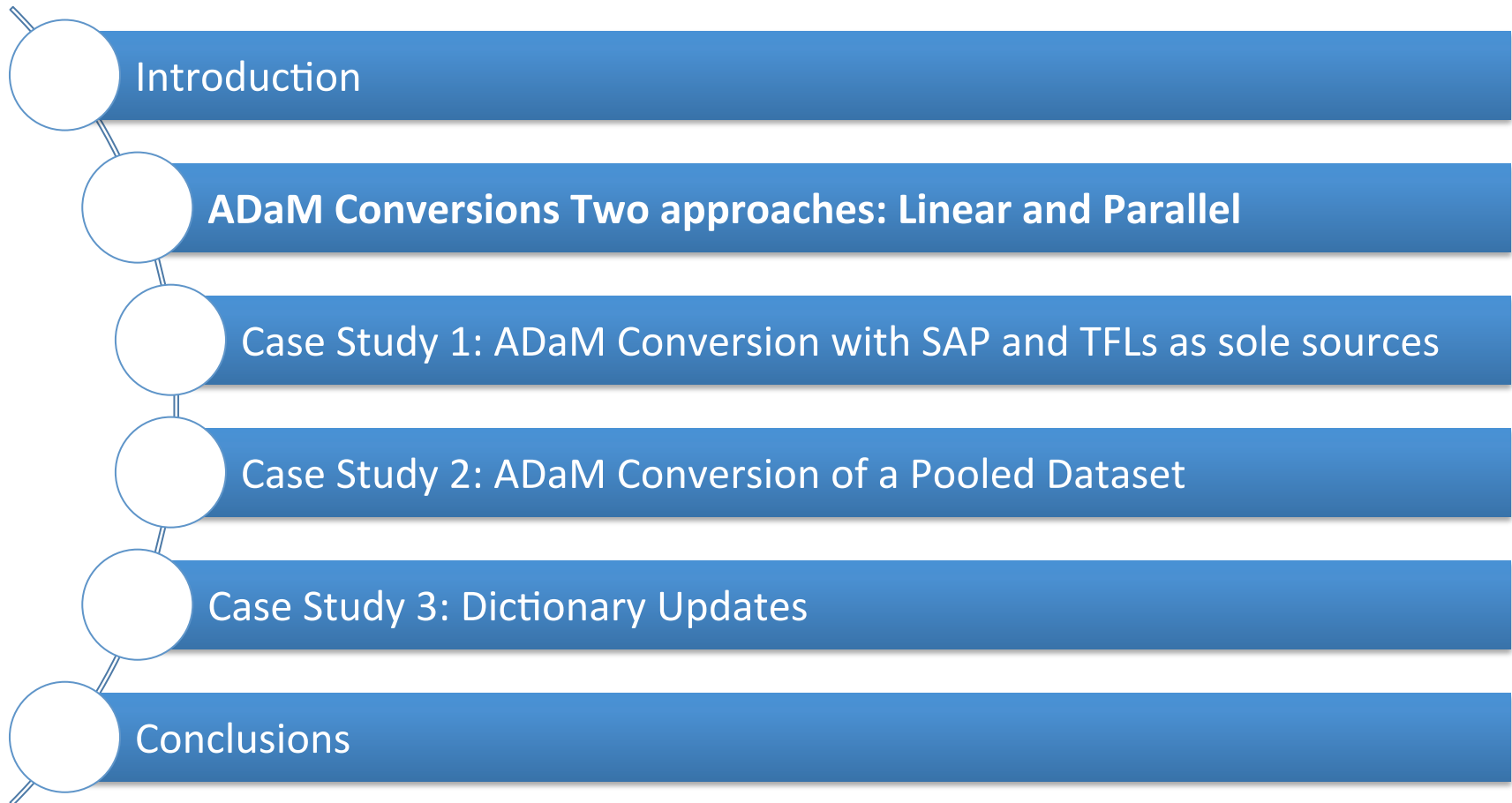
Business & Decision Life Sciences' experiences in ADaM conversion projects applying:

- Linear approach
- Parallel approach for Pooled datasets
- Dictionary updates in a Pooled database

Objective

Describing the good, bad and ugly experiences gained during the execution of these projects

ADaM Conversions: The Good, the Bad and the Ugly



ADaM Conversions Two approaches: Linear and Parallel



Good understanding of the relationship between

- Analysis results
- Analysis datasets
- SDTM domains

The **core** of any ADaM conversion is the SDTM/ADaM traceability, this can be found at two levels:

- **Metadata traceability**: finding a relationship between an analysis result and analysis dataset(s), or a relationship of the analysis variable to its source dataset(s) and variable(s)
- **Data point traceability**: finding the predecessor record(s)

ADaM Conversions Two approaches: Linear and Parallel



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%)

• 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 DSSTDC

Gender SEX 1 = male, 2 = female

Date of birth BRTHDTC Age AGEU years (Age is automatically calculated when screen is saved and closed)

Height cm

Weight kg

Waist circumference cm

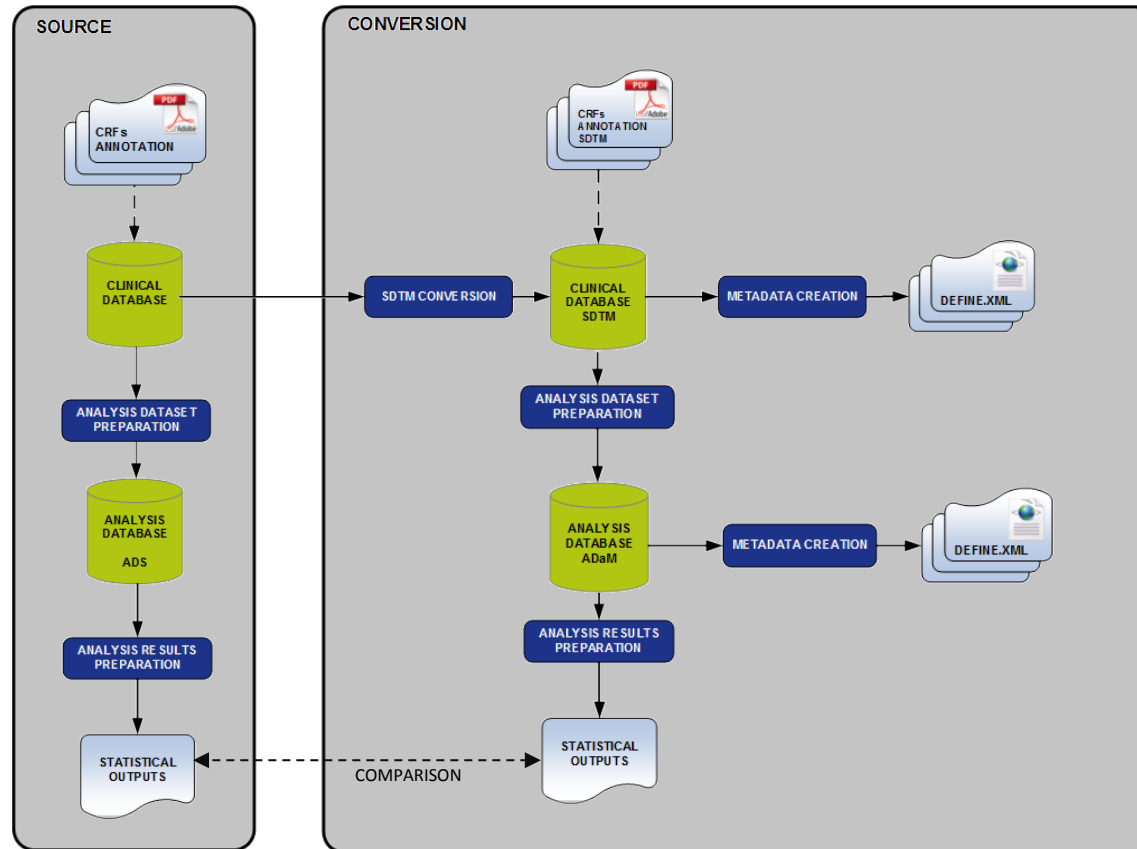
VSORRES / VSORRESU where VSTESTCD = "HEIGHT", "WEIGHT", "WAIST"

	STUDYID	USUBID	SUBID	BMI	BMIGR1	BMIGRIN	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	25.50361502 <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	30.968888131 >=30 kg/m**2			2 >=30 kg/m**2	3
7	9999-0001	9999-0001-000006	000006	39.687163916 >=30 kg/m**2			2 >=30 kg/m**2	3
8	9999-0001	9999-0001-000007	000007	25.82646281 <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	28.87613787 <30 kg/m**2			1 25-30 kg/m**2	2
12	9999-0001	9999-0001-000011	000011	29.372387383 <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	32.718619889 >=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

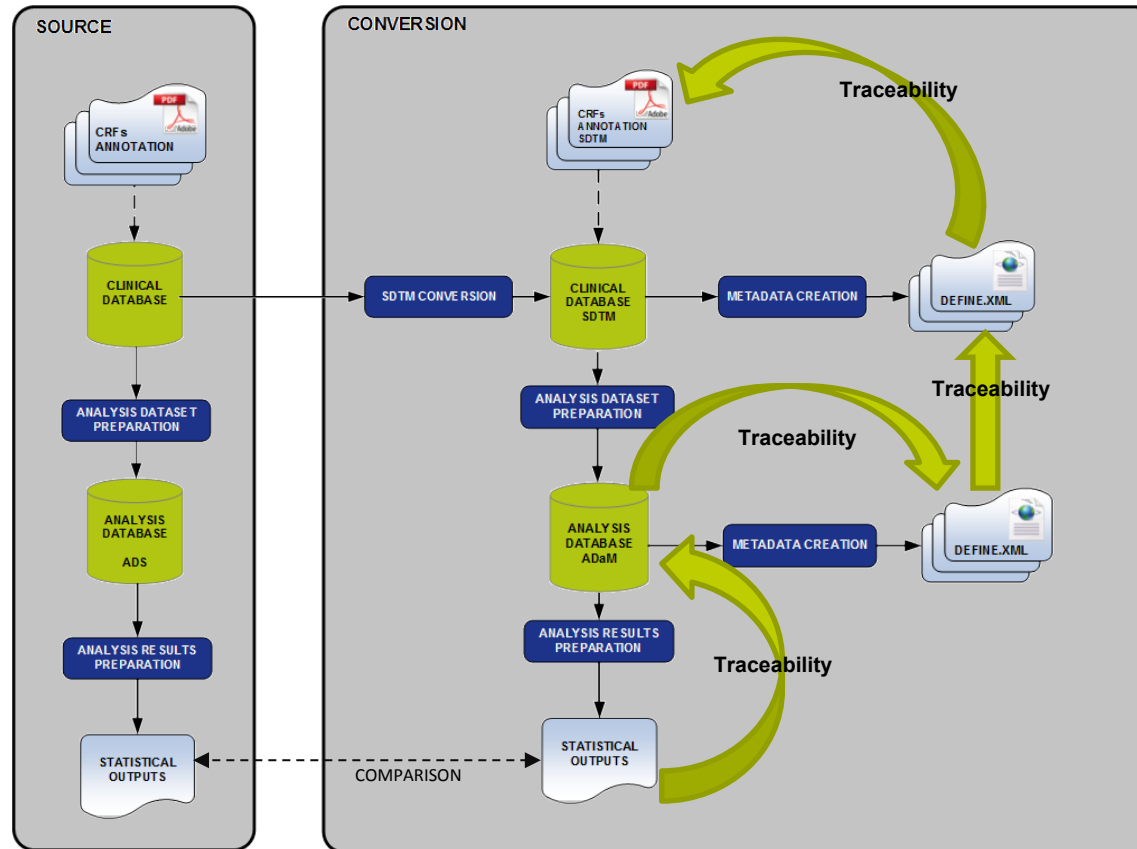
Variable	Label	Type	Controlled Terminology	Origin	Role	Comment
STUDYID	Study Identifier	text	Protocol, CRF Page 1	Protocol	IDENTIFIER	The STUDYID variable has a fixed format: 'XXXX-YYYY', where 'XXXX' indicates the 4-digit compound code and the 'YYYY' the 4-digit study code.
DOMAIN	Domain Abbreviation	text	DOMAIN	Assigned	IDENTIFIER	
USUBID	Unique Subject Identifier	text	Protocol	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	Assigned	TOPIC	
VSTEST	Vital Signs Test Name	text	Derived	Derived	SYNONYM QUALIFIER	
VSPOS	Vital Signs Position of Subject	text	POSITION	CRF Page 12	RECORD QUALIFIER	
VSORRES	Result or Finding in Original Units	text	Derived, CRF Page 9, 11	Derived, CRF Page 9, 11	RESULT QUALIFIER	
VSORRESU	Original Units	text	VSRRESU	CRF Page 9, 11	VARIABLE QUALIFIER	
VSTRESN	Character Result Finding in Std Format	text	Derived	Derived	RESULT QUALIFIER	
VSTRESN	Numeric Result Finding in Standard Units	float	3.1	Derived	RESULT QUALIFIER	

Variable	Label	Type	Controlled Terminology	Computational Algorithm or Method	Origin	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.
USUBID	Unique Subject Identifier	text			DM	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.
SUBID	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	
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	

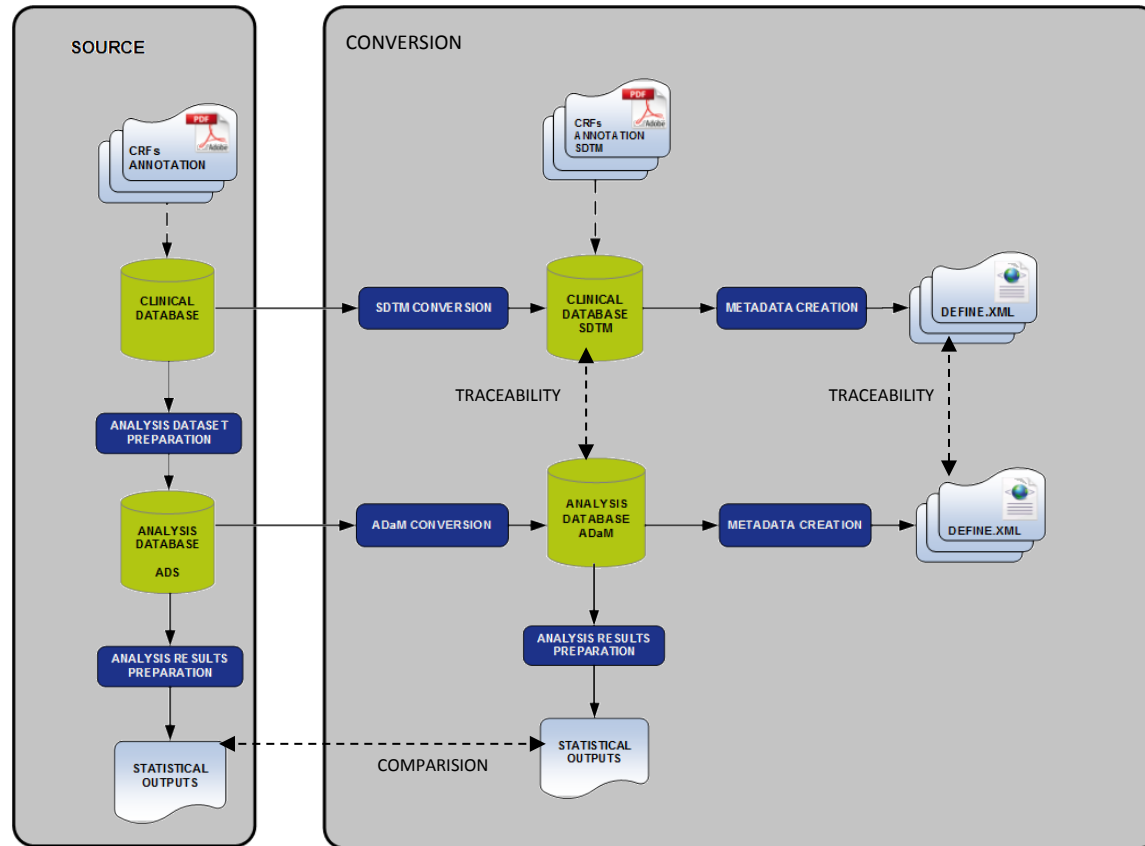
ADaM Conversions Two approaches: Linear and Parallel



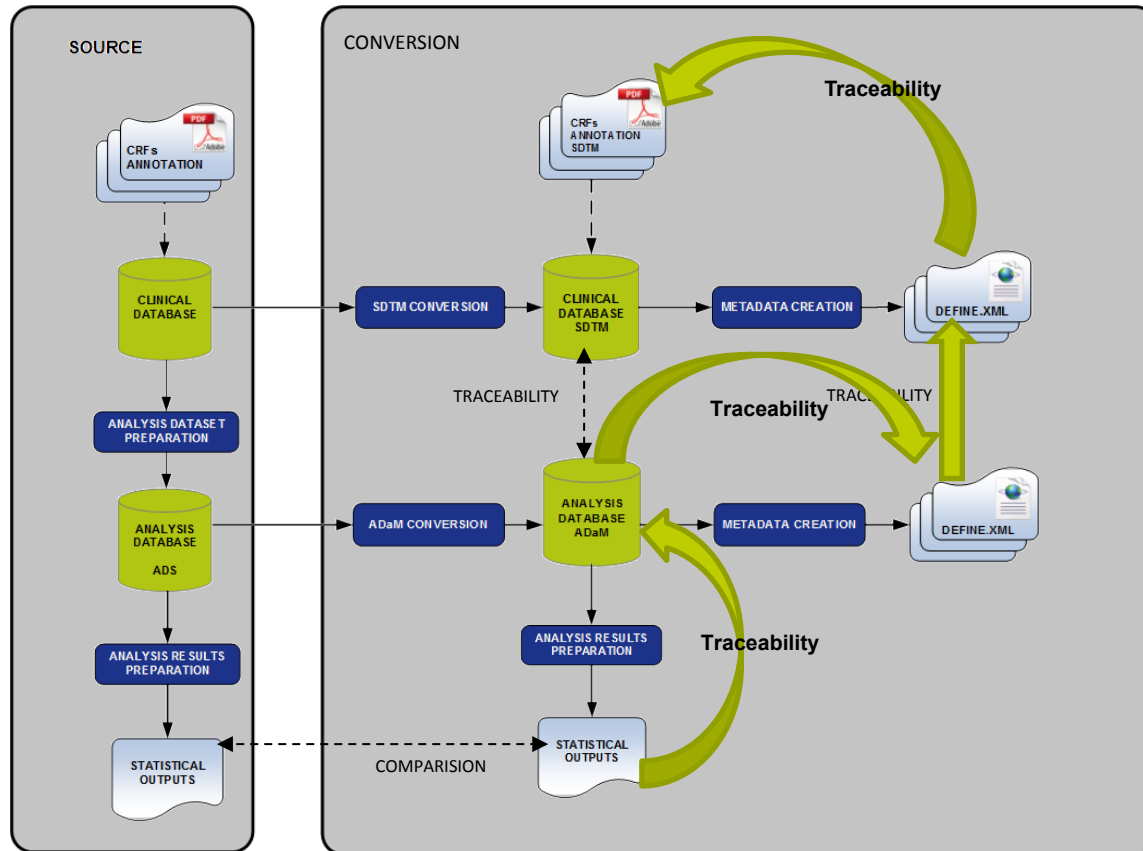
ADaM Conversions Two approaches: Linear and Parallel



ADaM Conversions Two approaches: Linear and Parallel

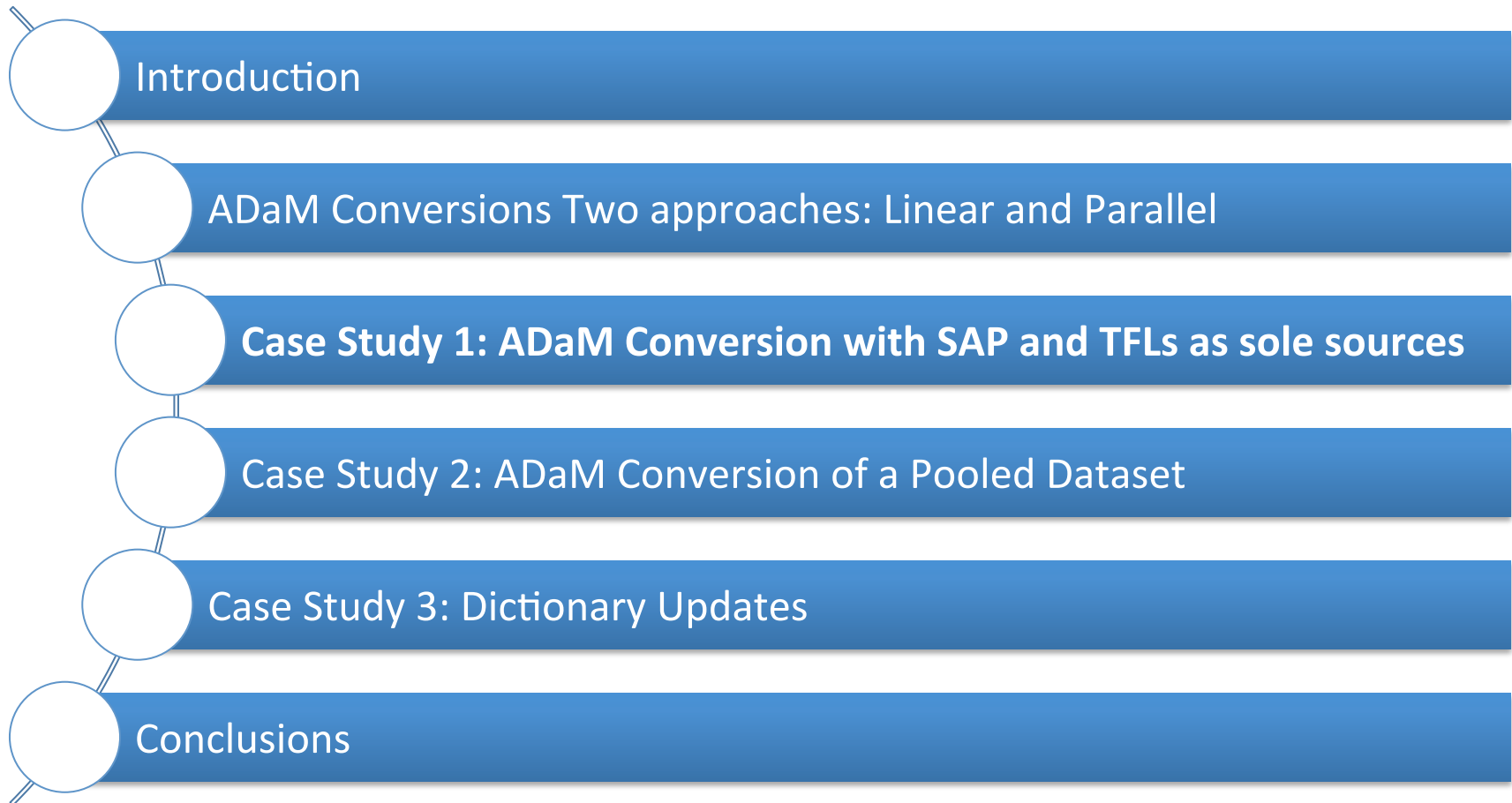


ADaM Conversions Two approaches: Linear and Parallel

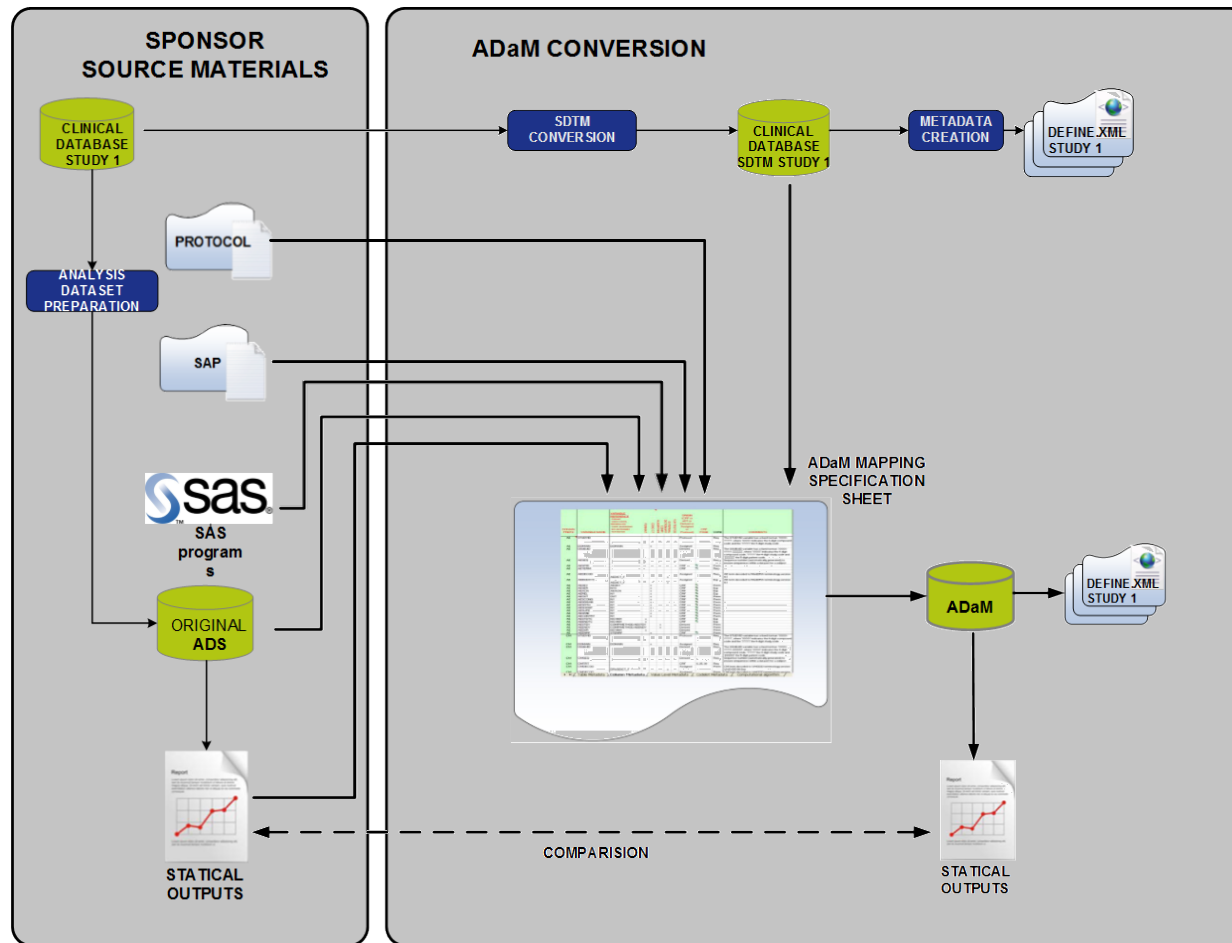


ADaM Conversion:

The good, the bad and the ugly



Case Study 1: ADaM Conversion with SAP and TFLs as sole Sources



Case Study 1: ADaM Conversion with SAP and TFLs as sole Sources



T
H
E
G
O
O
D



- **Good Traceability** between the statistical outputs, the ADaM datasets and the SDTM domains
- Since BDLS was responsible for the **SDTM Conversion**, issues with SDTM domains were resolved rapidly
- Having the ADS in ADaM structure facilitated the addition of this legacy study as part of an **Analysis Pooled Database**

Case Study 1: ADaM Conversion with SAP and TFLs as sole Sources



T
H
E
B
A
D



- Hidden differences with SDTM found during validation:
 - Different Coding Dictionary Versions in SDTM database
 - For consistency with newer studies, some categorical variables were updated with new combination of values
- Due to SDTM standards restrictions some values used for analysis were not available in the SDTM datasets
 - Key variables in SDTM domains cannot be empty.
→ In the initial analysis missing values were summarized in a table, these values were not available anymore in the SDTM datasets

Case Study 1: ADaM Conversion with SAP and TFLs as sole Sources



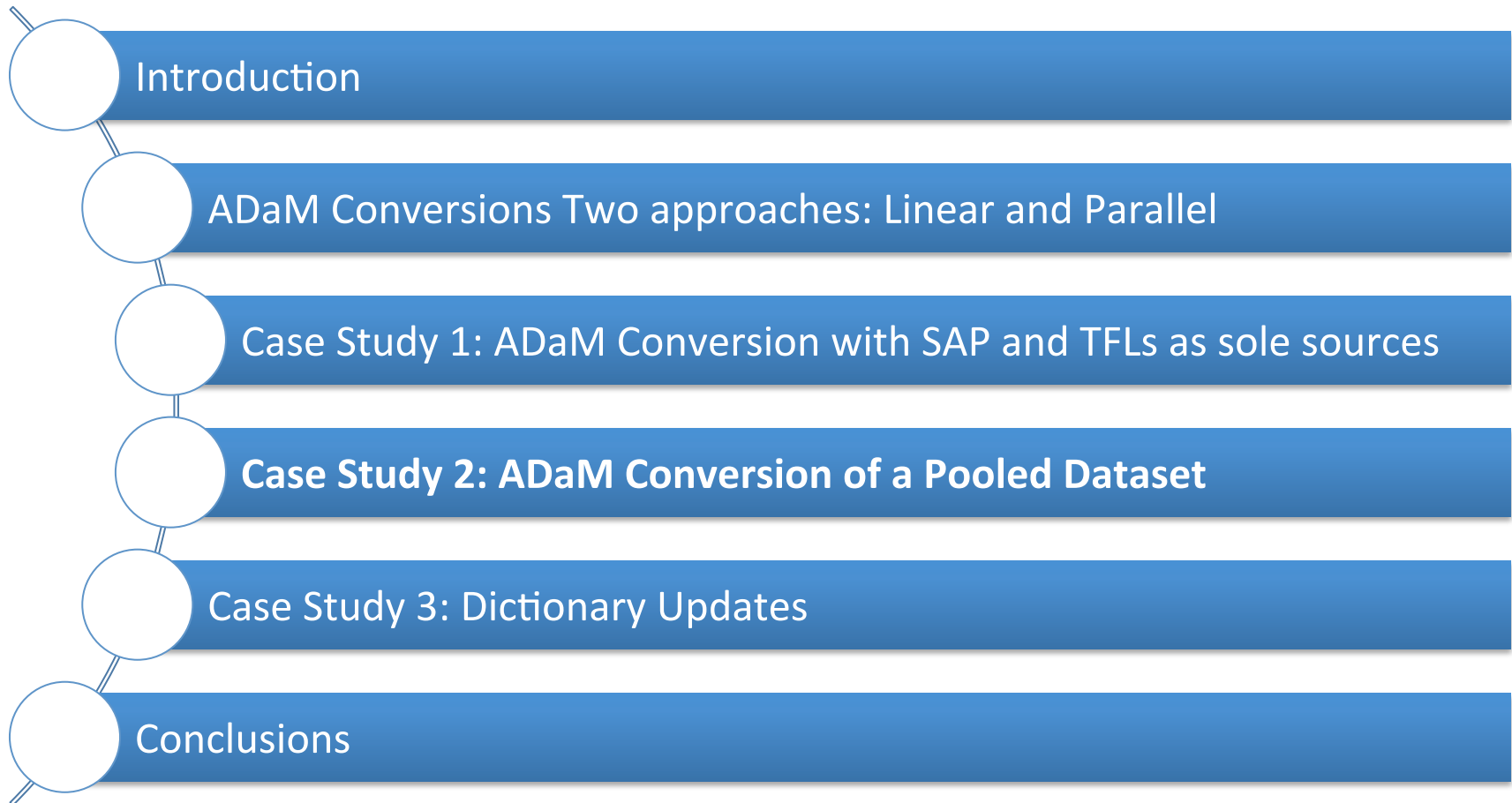
T
H
E
U
G
L
Y



- ADaM specifications created from SAP and protocol.
 - Rules specified in in Protocol and SAP were differently interpreted in the same way as in the original analysis.
- Multiple solutions were tried before having the proper result
- QC: manual comparisons were needed
 - **ADS** not always available
 - **Listings** in pdf format
- Resolving all these issues was time consuming

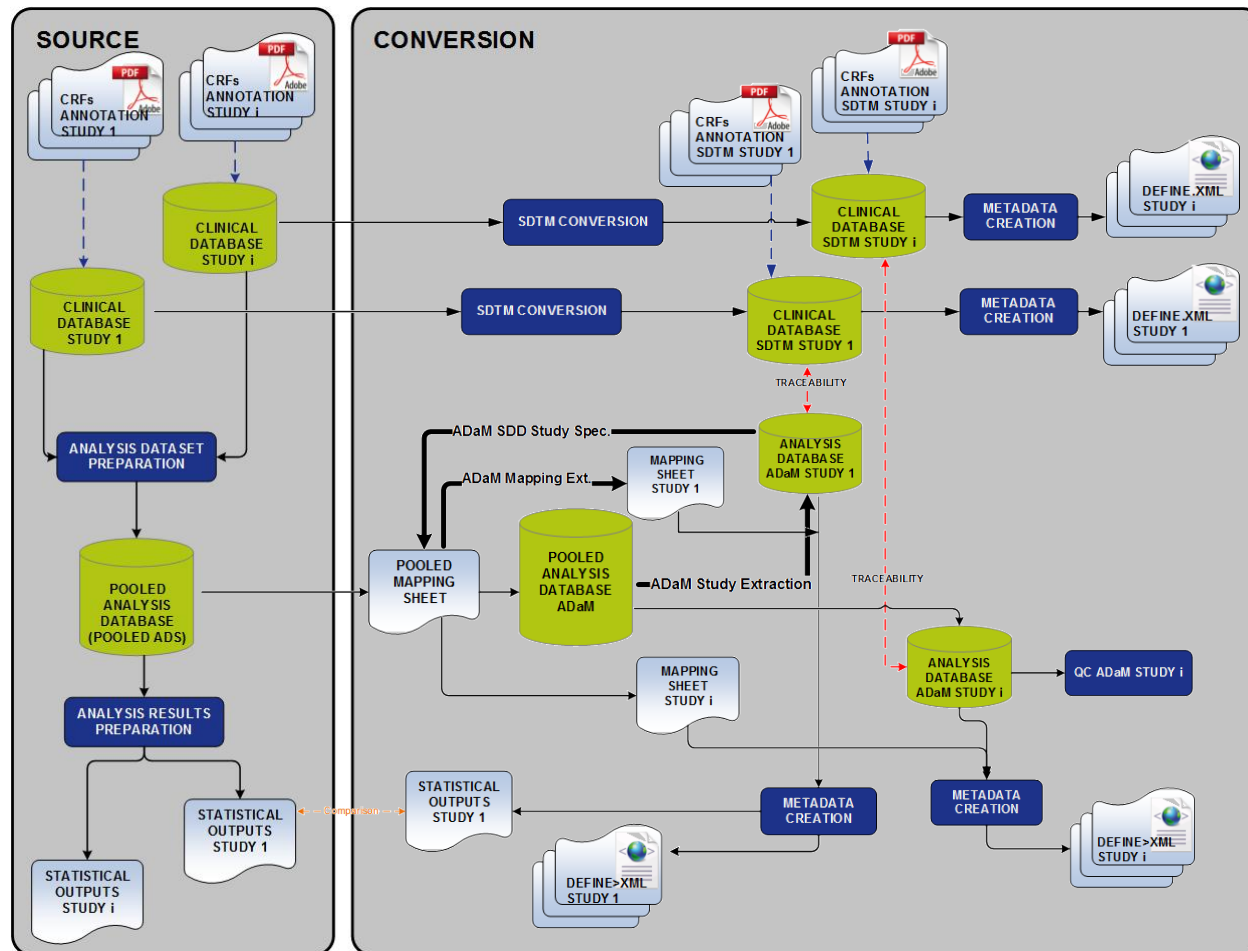
ADaM Conversion:

The good, the bad and the ugly



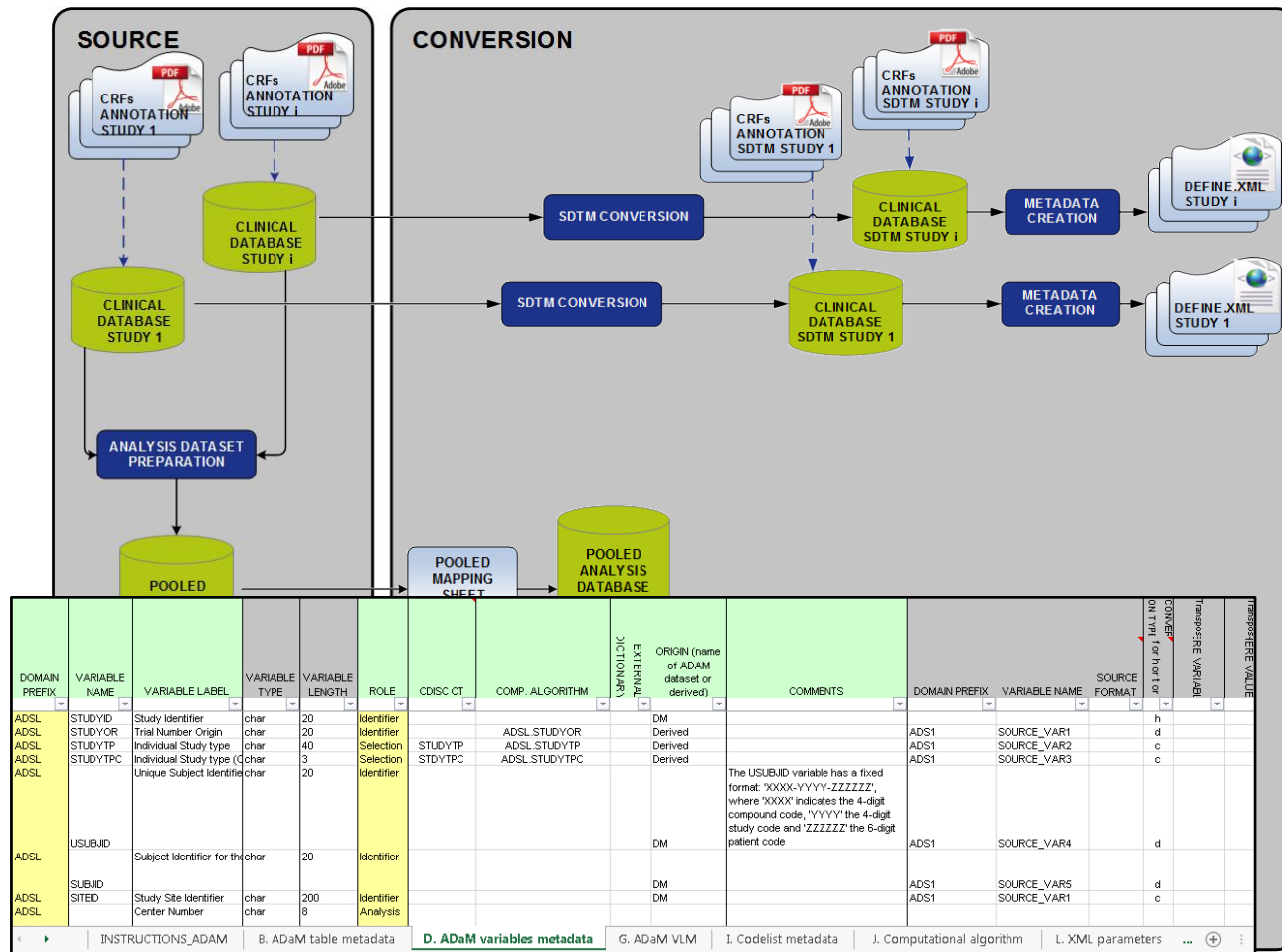
Case Study 2:

Adam Conversion of a Pooled Dataset



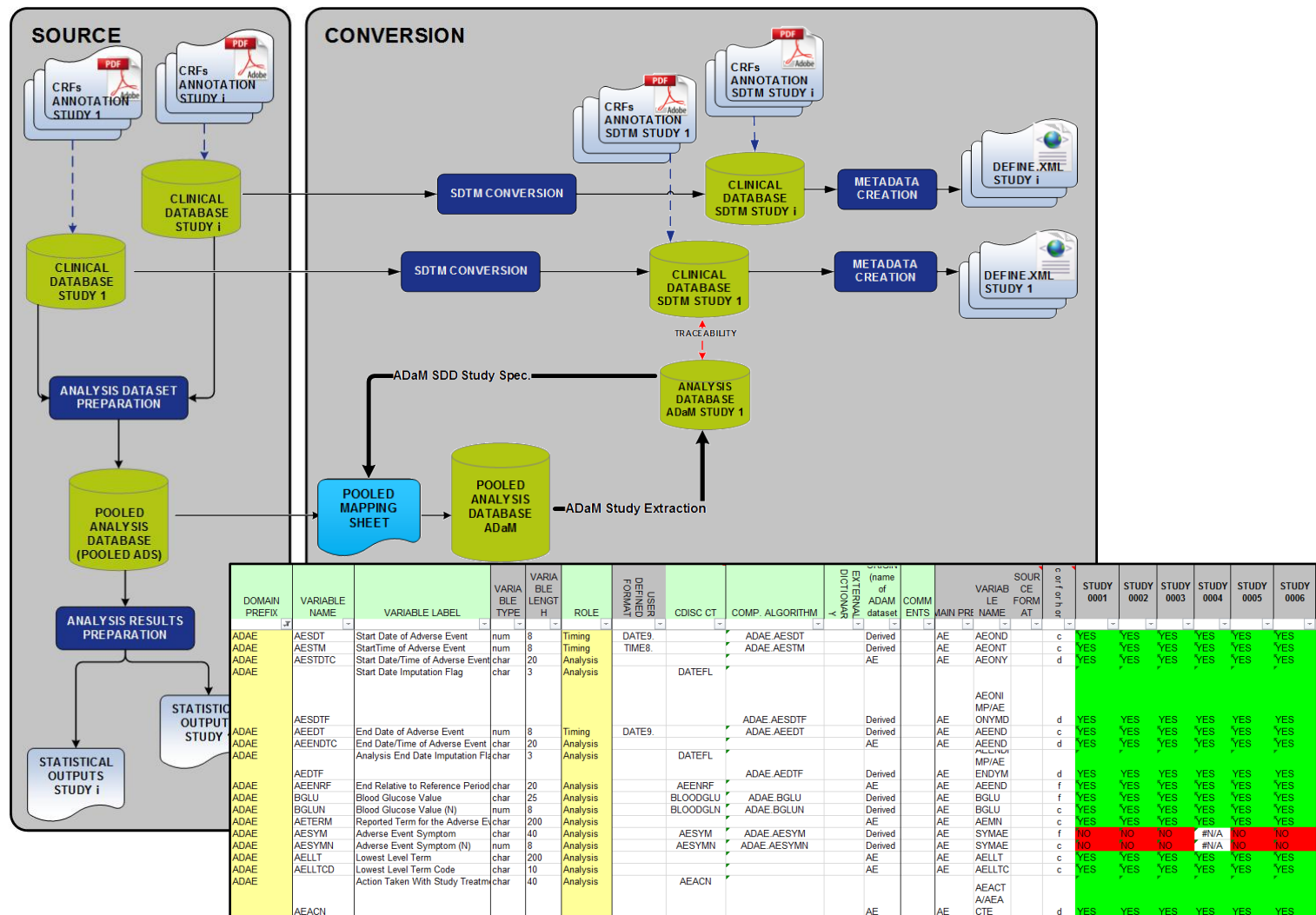
Case Study 2:

Adam Conversion of a Pooled Dataset



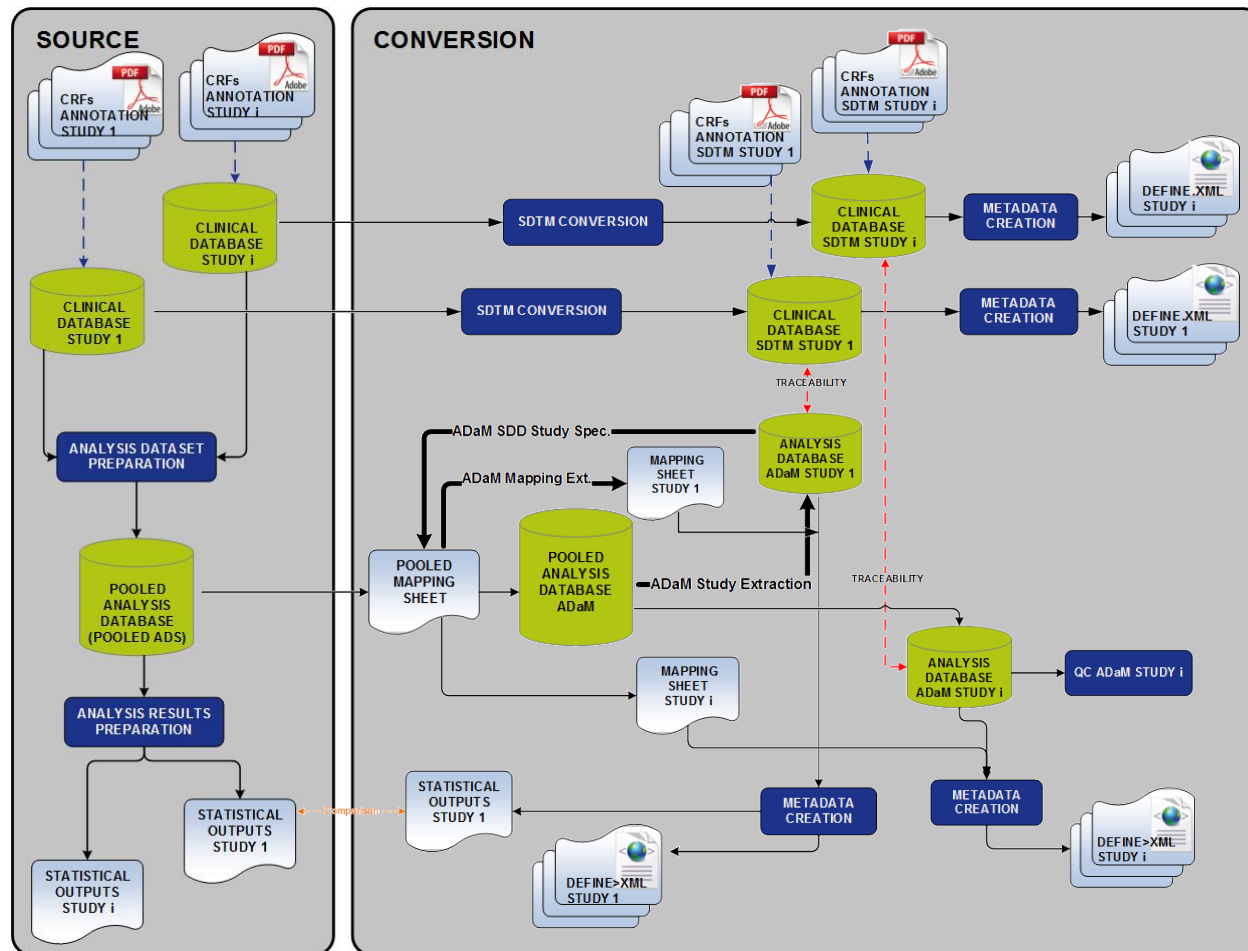
Case Study 2:

Adam Conversion of a Pooled Dataset



Case Study 2:

Adam Conversion of a Pooled Dataset



Case Study 2:

Adam Conversion of a Pooled Dataset



T
H
E
G
O
O
D



- BDLS was in charge of the SDTM and ADaM conversion, this **facilitated the communication** between the two teams and accelerated the problem solving
- With a **mature Pooled ADaM conversion process**, data extraction for each study is fast, and consistency checks across studies are no longer needed
- **Codelist were re-usable** in ADaM due to its consistency across studies.
- A **comment file** with all questions from and to the sponsor, facilitates the tracking of all answers and decisions

Case Study 2:

Adam Conversion of a Pooled Dataset



T
H
E

B
A
D



- Not well documented **Derivation Rules** caused difficulties to define the traceability.
- **Studies were added into the pooled database** at different time points, while derivations were not properly updated.
- **SDTM limitations** affected the ADaM conversion
 - some raw variables were not available in SDTM, but were mentioned in ADS specifications.
 - Specifications had to be redefined in cooperation with the sponsor.
- Due to traceability between SDTM/ADaM, **timelines** for ADaM datasets were influenced by SDTM timelines

Case Study 2:

Adam Conversion of a Pooled Dataset



T
H
E

U
G
L
Y



- **Structure of the pooled original ADS datasets** changed with the inclusion of new studies.
 - Impact on the pooled mapping, pooled datasets and extractions
- **Source variables** which had some manipulation during the conversion (e.g. variable derived, with a format or transposed) were **sensitive to changes in its content** at each transfer.
 - New compare tools to check changes in dataset structure between the original ADS received.

Case Study 2:

Adam Conversion of a Pooled Dataset



- BDLS developed **new compare tools** to check changes in dataset structure between the original ADS received, like number of records, empty, new, and missing variables that were present before
- Specifications** were **compared**: Results of a compare were included in the communication file to the sponsor

	B	C	D	E	F	H	I	J
1	VARIABLE	DESCRIPTION	TYI	compare_flg	new_domain_flag	B&DLS comments	Sponsor comments	Updates
2	SOURCE_VAR_1	AESI flag-Angioedem num		ADDED	NO			
3	SOURCE_VAR_2	AESI flag-Embolic an num		ADDED	NO			
4	SOURCE_VAR_3	AESI flag-Hypoglycae num		ADDED	NO			
5	SOURCE_VAR_4	AESI flag-Hypersensi num		MODIFIED	YES			
6	SOURCE_VAR_5	AESI flag-Increased unum		ADDED	YES			
7	SOURCE_VAR_6	Investigator Special Ir num		REMOVED	NO			
8	SOURCE_VAR_7	Investigator Special Ir num		REMOVED	NO			
9	SOURCE_VAR_8	Investigator Special Ir num		REMOVED	NO			
10	SOURCE_VAR_9	Investigator Special Ir num		ADDED	NO			
11	SOURCE_VAR_10	Investigator Special Ir num		ADDED	NO			
12	SOURCE_VAR_11	Investigator Special Ir num		ADDED	NO			
13	SOURCE_VAR_12	Investigator Special Ir num		ADDED	NO			
14	SOURCE_VAR_13	Investigator Special Ir num		ADDED	NO			
◀ ▶ Issues & updates post-release Source data issues specs_comp Compare_tool SDTM Source not mapped Review cycle								

Case Study 2:

Adam Conversion of a Pooled Dataset



- ADS content was compared:**

Results from content compare for variables being “derived”, “with a CL applied”, “transposed” during ADaM conversion, were included in the communication file to the sponsor

The below compare results are from comparing the content of variables in old transfer against the new transfer. This compare is done ONLY for variables that are 'derived' or have a format applied while being mapped into ADaM. Any change in the source implies a change in ADaM derivation rule, e.g. transpose rule, or code list.

A	B	C	D	E	F	G	H
studyid	DOMAIN	variable	comment	content	B&DLS Comment	Sponsor	Updates Comments
0001	ADS11	SRC_VAR_5	new variable to transpose present in new database	Creatinine >= 1.5* ref. sample and > ULN	we will include a new criteria for this value. Please confirm the value is correct	correct	Pooled mapping updated
0001	ADS2	SRC_VAR_7	Domain not present in new database	Creatinine >= 2*baseline and > ULN	we will include a new criteria for this value. Please confirm the value is correct	correct	not included
0001	ADS9	SRC_VAR_5	Value not present in code list metadata but present in source data	TS.CV	this is a new set of population flag, please confirm if we need to add it	Please ignore	not included

Issues & updates post-release | Source data issues | specs_comp | **Compare_tool** | SDTM Source not mapped | Review cycles | Consistency checks | Def ...

Case Study 2:

Adam Conversion of a Pooled Dataset

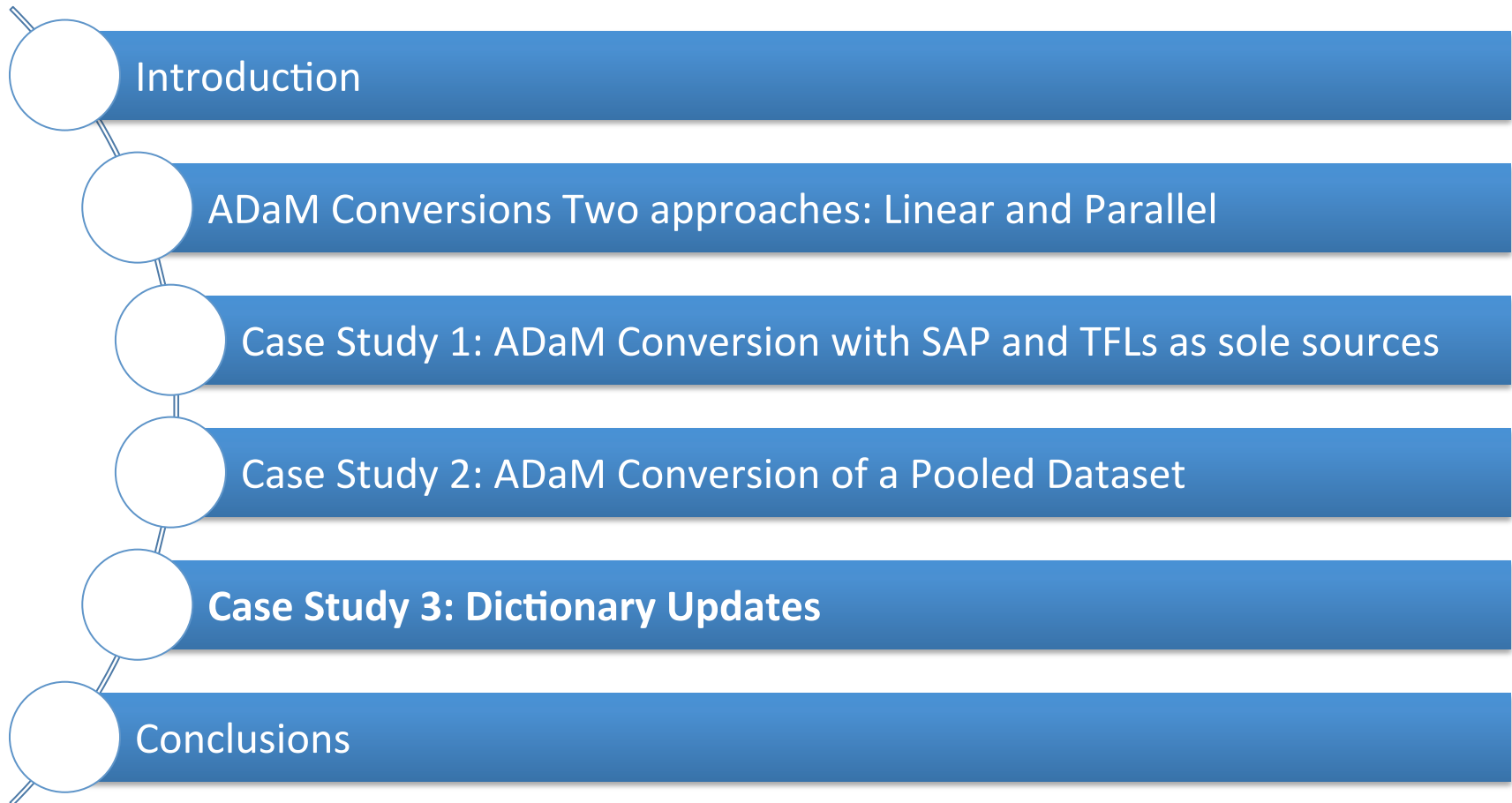


- To ensure traceability with SDTM, a good knowledge of **what was mapped from RAW into SDTM** and in which study, was essential
- List of RAW variables per study** per domain from each SDTM metadata, specifying if the variable was mapped into SDTM, and if it was empty

STUDY	VIEWNAME	VARNAME	LABEL	VARTYPE	FORMATN	MAPPED	CONFIRMED	EMPTY
RAW_Study1	AE	AEI	AE indicator	num	YN1F.		x	X
RAW_Study1	AE	AEONTM	AE onset time	num	TIME5.	X	x	
RAW_Study1	AE	AEENDC	AE end date continued	char	\$CONT1F.	X	x	
RAW_Study1	AE	AEOUT	AE outcome	num	AEOUT1F.	X	x	
RAW_Study1	AE	AEREL	AE drug relationship	num	YN1F.	X	x	
RAW_Study1	AE	AEONDT	AE onset date	num	DATE9.	X	x	
RAW_Study1	AE	AEENDDT	AE end date	num	DATE9.	X	x	
RAW_Study1	AE	AELLT	AE lowest level term	char	\$200.		x	
RAW_Study1	AE	AELLTCD	AE lowest level term code	char	\$10.		x	
RAW_Study2	AE	AEI	AE indicator	num	YN1F.		x	X
RAW_Study2	AE	AEONTM	AE onset time	num	TIME5.	X	x	
RAW_Study2	AE	AEENDC	AE end date continued	char	\$CONT1F.	X	x	
RAW_Study2	AE	AEOUT	AE outcome	num	AEOUT1F.	X	x	
RAW_Study2	AE	AEREL	AE drug relationship	num	YN1F.	X	x	
RAW_Study2	AE	AEONDT	AE onset date	num	DATE9.	X	x	
RAW_Study2	AE	AEENDDT	AE end date	num	DATE9.	X	x	
RAW_Study2	AE	AELLT	AE lowest level term	char	\$200.		x	
RAW_Study2	AE	AELLTCD	AE lowest level term code	char	\$10.		x	

ADaM Conversion:

The good, the bad and the ugly

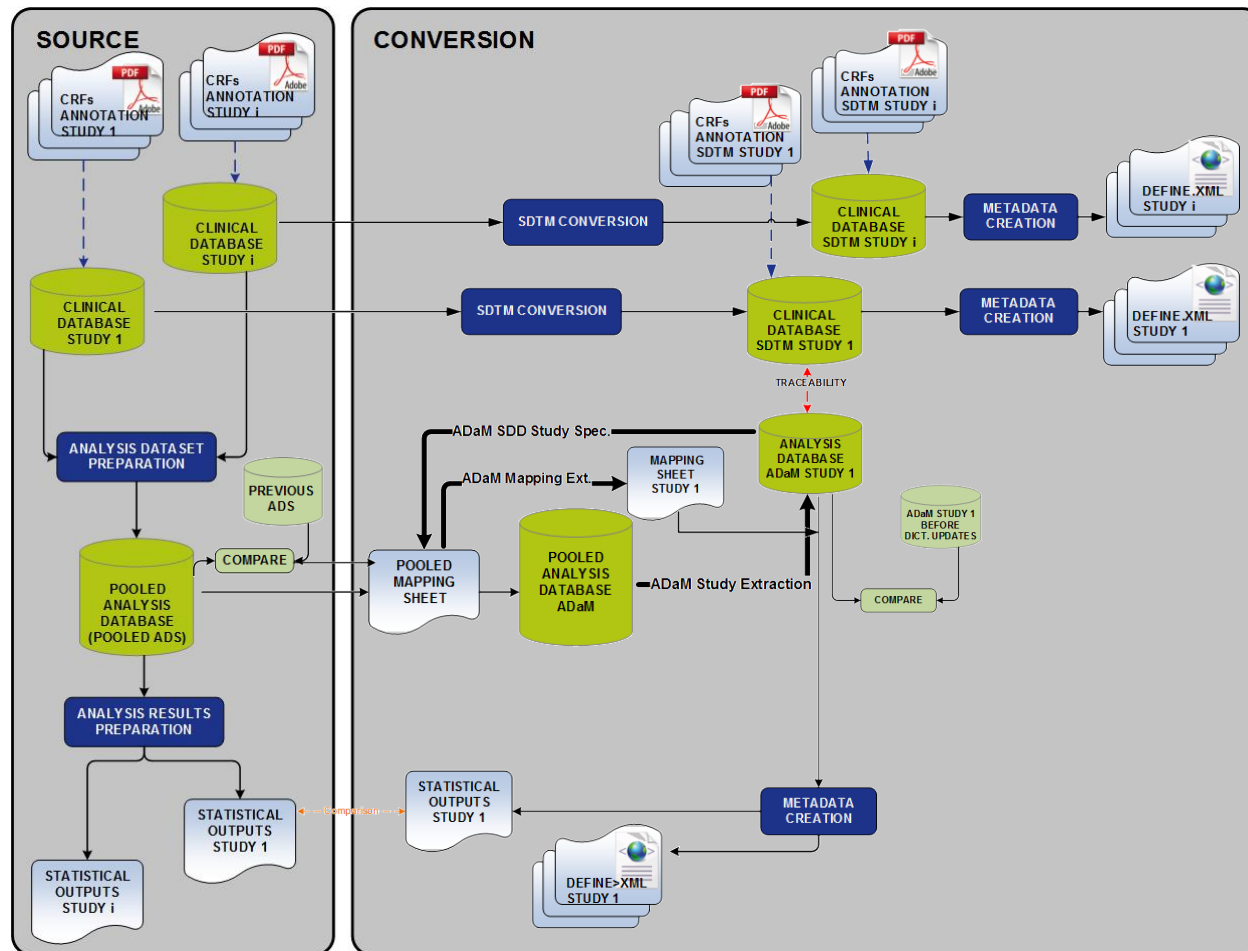


Case Study 3: Dictionary Updates



- MedDRA and WHO-Drug Code variables were updated to a newer dictionary version than the original. Updates were implemented in some studies as part of the pooled database
- As ADS source we received from the sponsor only the Pooled datasets affected by dictionary updates
- Pooled mapping was updated accordingly
- ADaM pooled datasets were regenerated, and each study was extracted
- Finally a comparison between the old and the new ADaM datasets of each study was performed to identify that the only changes were related to dictionary updates

Case Study 3: Dictionary Updates



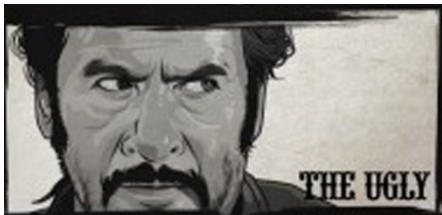
Case Study 3: Dictionary Updates



Once the **changes required had been established** for the first study, it was a mechanical process, it was sure, simple and fast



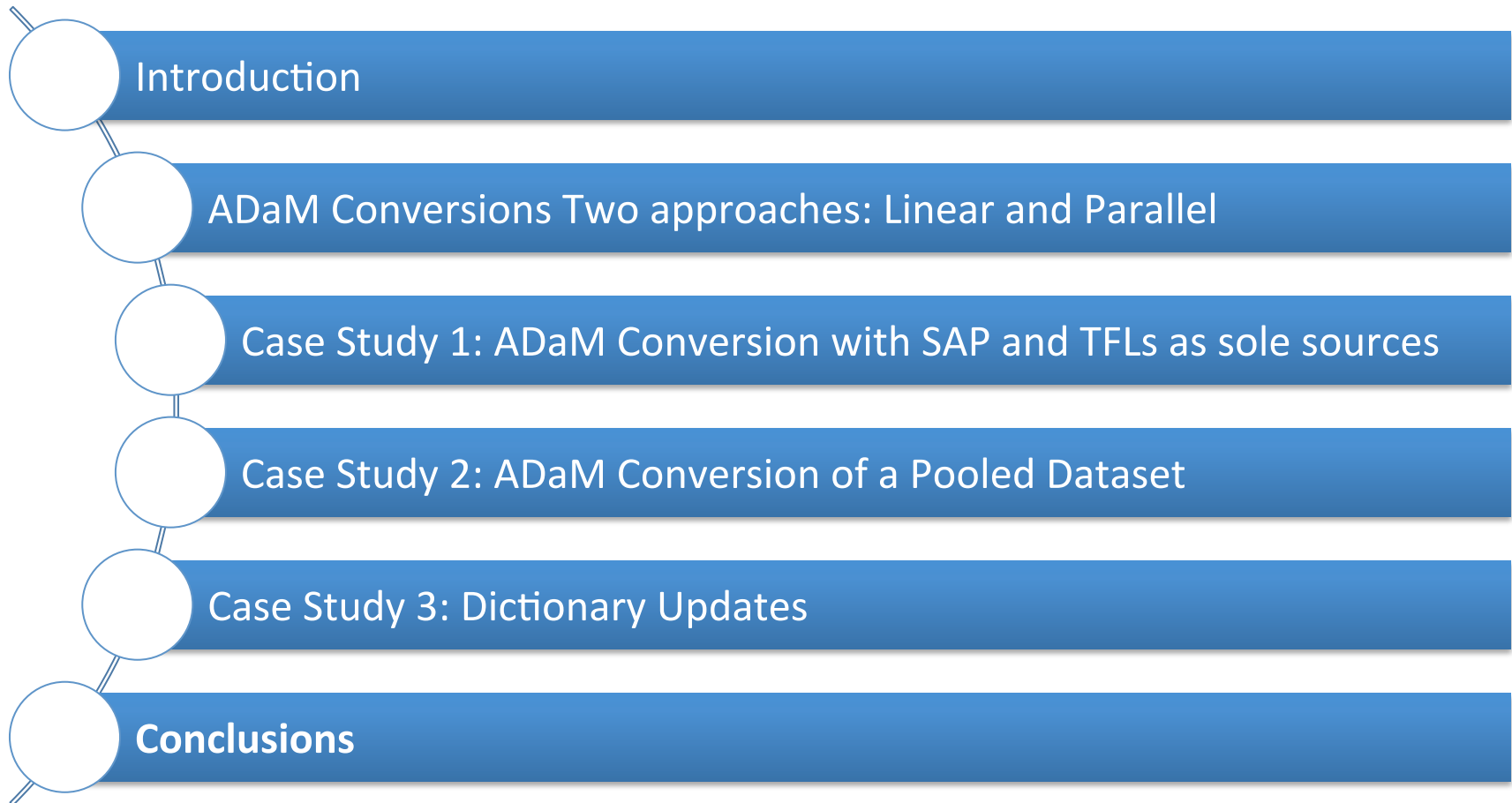
New ADS variables were found, not related to a dictionary update. Derived variables used in new studies but not used in old studies analysis were present. This caused BDLS macros failed when re-running the poled ADaM datasets



Several transfers per each study on a frequent basis. We kept track of the original ADS used and final version of ADaM sent. A discrepancy report was created to list differences found, detailing their origin and resolution

ADaM Conversion:

The good, the bad and the ugly



Conclusions



- **Parallel approach:**
 - ensured reproduction of TFLs
 - consistency between studies within a pooled database
- **Linear approach:**
 - SDTM/ADaM traceability is demonstrated
 - Bigger risk of not reproducing identical results to the initial analysis
- **Source pooled ADS** should be stable and **specifications** should be clear
- SDTM datasets & pooled ADaM datasets should be **finalized** before the individual study extractions can start

Conclusions



- Dictionary update: SDTM and ADaM uses the **same dictionary version**. Specific ADaM variables are available
- **Good communication** is needed between SDTM, ADaM team members and the sponsor, during the whole process
- **Sponsor's feedback** on draft mapping and draft ADaM datasets is crucial to address all details upfront
- All decisions can be easily tracked if they are properly documented in a **communication file**

Thank you for your attention.

Brussels, Belgium, 18 September 2014





Carla Santillán | Project Manager | Statistical Projects carla.santillan@businessdecision.com
Jessica Minkue Mi Edou | Statistical Programmer | jessica.minkue@businessdecision.com

Business & Decision Life Sciences
Sint-Lambertusstraat 141 rue Saint-Lambert
B-1200 Brussels
T: +32 2 774 11 00
F: +32 2 774 11 99
lifesciences@businessdecision.com
<http://www.businessdecision-lifesciences.com/>