PhUSE 2014

Paper DH01

ADaM conversions: The good, the bad and the ugly

Carla Santillan, Business & Decision Life Sciences, Brussels, Belgium
Jessica Minkue Mi Edou, Business & Decision Life Sciences, Brussels, Belgium

ABSTRACT

A good, bad and ugly face were identified from three types of ADaM conversions experiences. We will share our experiences, strategies and solutions from:

- conversion of a pooled database;
- dictionary updates;
- conversion with SAP and TFLs as the only sources

The Good: creation of final ADaM datasets in which SDTM / ADaM traceability was accomplished. SDTM and ADaM metadata modified accordingly to each other's structure requirements. The Bad: Source derived datasets and SDTM metadata didn't match up due to differences in formats, SDTM codelist, dictionary versions, etc. The Ugly: incomplete and unclear source and specifications can lead to mistakes or inconsistencies that could not be found until the validation process, causing re work and delay. In addition source datasets derived values may need to be re-derived to fit within the ADaM structure. All three conversions had challenges and led to process improvements which will help in future conversion projects.

INTRODUCTION

In this paper we will briefly summarize our experiences from two types of ADaM conversions, ADaM conversion linear approach and conversion of a Pooled dataset using a parallel approach. The third experience was related to dictionary updates into the Pooled database, this was not a complete conversion, but an update. We will take separately each of them and describe our good, bad and ugly experiences. Explaining in each case the general or specific solutions we found.

ADAM CONVERSION (LINEAR APPROACH)

This conversion starts with a good understanding of the relationship between the analysis results, the analysis datasets and the SDTM domains. A path between an element and its immediate predecessor is established, always identifying the SDTM/ADaM traceability. See in fig 1 the flow chart that summarize the steps taken in an ADaM conversion.

The following input is required as source documentation: Protocol; Statistical Analysis Plan (SAP); Analysis specifications plus analysis program (if available); Clinical Summary Report (or statistical outputs) if available; raw data (SDTM structure or not); SDTM Define.xml if SDTM is available.

In this approach the raw datasets are converted into SDTM datasets which are used later on to create ADaM datasets. Original Analysis Data Sets (ADS) that are created as intermediate step before TFLs, normally are developed using raw and not SDTM datasets. All the available documentation related to the ADS like the SAP, datasets programs, formats, and specifications are used to establish traceability between SDTM/ADaM.

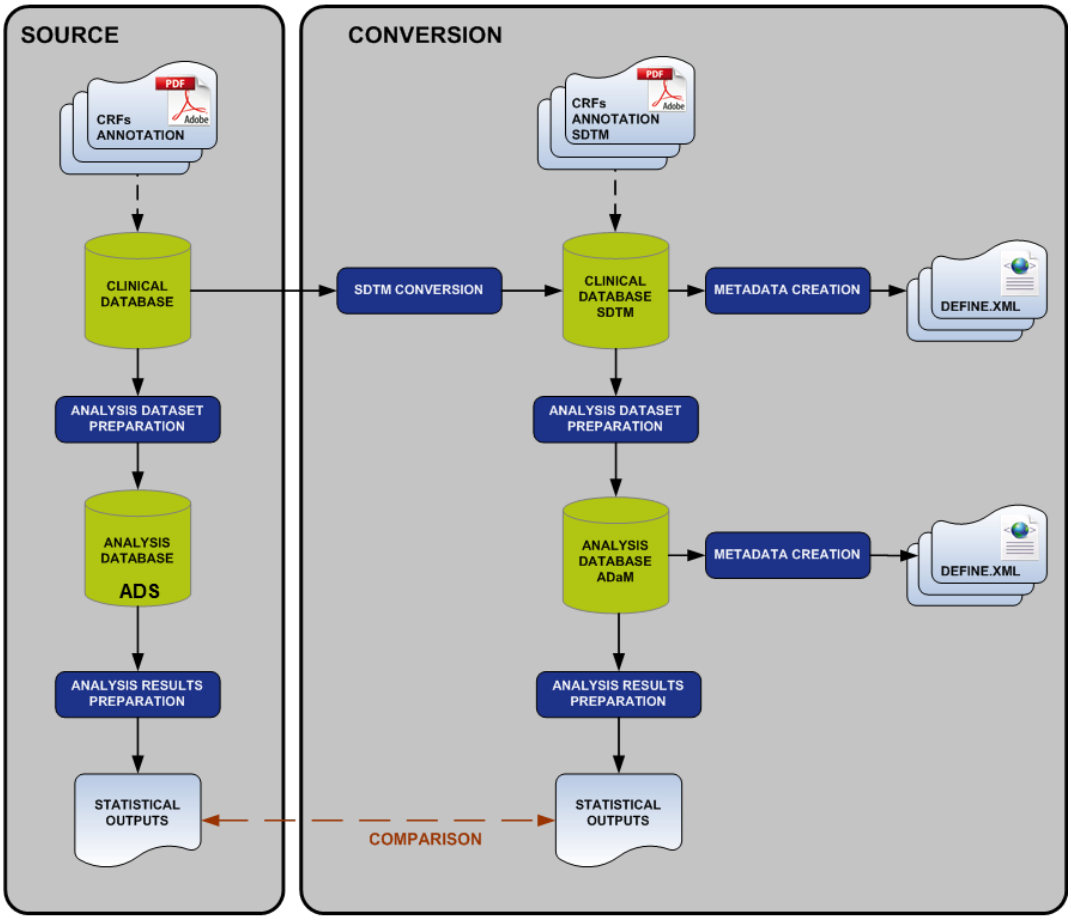


Fig 1: Conversion of original analysis datasets into ADaM datasets, linear approach

As a result of this conversion we obtain ADaM datasets, ADaM mapping (specifications), Define.xml, Reviewer's Guide, and the Statistical output or a subset of key tables can be generated using ADaM datasets. As validation a set of Key TFLs are re-created using ADaM datasets and compared with the TFLs created from the original ADS.

The creation of ADaM specifications follows ADaM IG rules and includes all relevant information per domain and per variable. The relationship between source (SDTM) and target (ADaM) datasets is placed in ADaM mapping. A selection of macros is used to read the ADaM mapping and create ADaM datasets and later on create ADaM Define.xml.

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)

If SDTM/ADaM traceability was properly addressed during the conversion then the below path, see fig 2, should be easily found by the FDA reviewers:

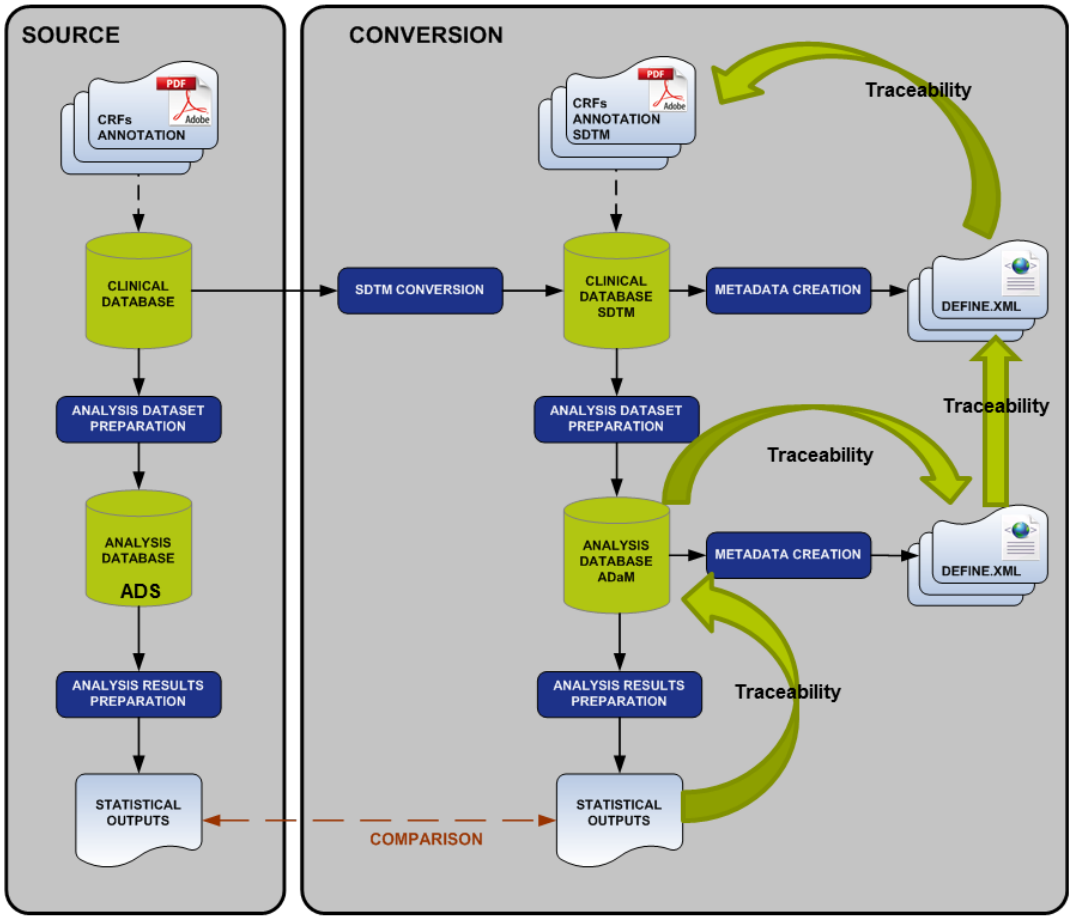


Fig 2: Conversion of original analysis datasets into ADaM datasets, linear approach – Traceability goes from Statistical Outputs (TFLs) => Analysis Database ADaM => ADaM Define.xml => SDTM Define.xml => CRF.

If any source documentation is unclear the traceability and the ADaM conversion process may be jeopardized.

CASE STUDY: CONVERSION WITH SAP AND TFLS AS THE ONLY SOURCES

CONVERSION PROCESS

In this study case we converted legacy studies into ADaM standard for an FDA submission. To perform the task we had the raw data in SDTM format, the protocol, the statistical analysis plan, the original statistical outputs and some of the analysis programs. The conversion of the raw clinical data into SDTM standards was done by BDLS SDTM team.

For this ADaM conversion, the linear process explained in previous section was used. The following steps can summarize the conversion: creation of the mapping, creation of the ADaM datasets using the SDTM database and validation by reproducing the statistical outputs.

Creation of mapping: The original statistical outputs were utilized to define the type of ADaM variables and datasets needed. We had to ensure that all the tables, figures and listings can be reproduce from the new ADaM datasets. The computational algorithms defined per each derived ADaM variable were then established using the protocol and SAP and based on the SDTM database content.

Creation of ADaM datasets: ADaM datasets were created using SDTM database and some automated BDLS macros. Eventually there were some additional programming for derived variables (variables not copied from SDTM database).

Validation: The purpose of the validation is to make sure that the computational algorithms implemented in the mapping, and the subsequent programming are accurate. In this case the validation involved the comparison of the statistical

outputs produced using the ADaM datasets against the original statistical outputs provided by the sponsor and produced using their original ADS.

THE GOOD:

With this conversion we were able to achieve a good traceability between the statistical outputs, the ADaM datasets and the SDTM domains.

Following this type of conversion that uses SDTM datasets as source, traceability between SDTM and ADaM is ensured, but there is a risk of not being able to reproduce the same results.

The fact that BDLS was also responsible for the SDTM conversion helped in the process. Issues with SDTM domains were resolved rapidly. If information which was valuable to us was missing from a SDTM domain, we were able to identify quickly if it was possible to include this information in one of the domain, or if we had to think of another solution to resolve the problem. Since computational algorithms are based on SDTM database, we had to make sure that the SDTM team included in their databases, before these were final, all the information necessary for the ADaM team.

Having the ADS in ADaM structure facilitated the addition of this legacy study as part of an Analysis Pooled database.

THE BAD:

Decisions taken at the sponsor's data management level impacted the conversion process. Some categorical variables were updated with new combination of values, probably for clarity and consistency with newer studies. Therefore, some variables ended up having fewer categories than they had initially. As a result, the new tables did not match with the original ones.

The coding dictionary versions were also updated. SDTM used newer dictionaries than the one from the original analysis, leading to differences at validation.

Finally, due to restrictions in SDTM standards some values reported originally were not available for analysis in the SDTM datasets. Key variables in SDTM domains cannot be empty, if in the original analysis some missing values were summarized in a table, these values were not available anymore with the new source, SDTM. Let's take the example of a categorical variable, RESPONDER that was represented in a summary table. For all the 18 subjects in the study, the information was missing. For non-missing values this information is represented as a TESTCD in a SDTM domain. In SDTM, a record is created only if there is a value to report. Which means that the information of subjects having a missing value for the variable RESPONDER is nowhere in SDTM. Then, it is not possible to re-create the variable RESPONDER as originally with all the values set to missing in an ADaM domain due to traceability issues with SDTM.

All the above discrepancies between the original tables and the one reproduced with the ADaM datasets that were not resolved due to traceability lost between SDTM and ADaM were well documented and communicated to the sponsor. None of them represented any significant change in the interpretation of the original analysis results.

THE UGLY:

It is not uncommon in legacy studies for documentation to be incomplete or missing, and for this study the specifications for Analysis Data Sets were not available. Consequently, ADaM specifications were created using the information found in the SAP and the protocol. Therefore, at validation we found that the rules specified in the SAP were not necessarily interpreted in the same way in the ADaM specification as they were for the original analysis. For example, from our interpretation of the SAP, for a specific baseline calculation we included only patients having a certain number of symptoms episodes, whereas in the original tables all subjects were included in the same calculation.

In the previous example, the mistake was easily found and corrected. However others were not so simple, and we had to try multiple solutions before having the proper result. One of the reasons was that the study was conducted more than 10 years ago and, the analysis data and output were handled by different companies. Given the age of the study, the datasets were handled by people who were not longer involved in the project and we were not able to have all the answers to our questions when discrepancies were found at the validation step. It was difficult to identify if the reason for the mismatch was because the derivation of the variable was not correct, a rounding problem, or the statistic used was not exactly the same. A trial-and-error approach with different derivations were needed before finding the proper one. In addition to this, as the original ADS used to produce the original statistical outputs were not

available, it was difficult to identify the problem immediately. Another fact that slow down the process was that the listings were in pdf format, so the comparison with the ADaM datasets had to be done manually.

ADAM CONVERSION (PARALLEL APPROACH)

In this approach the original ADS content is used as a source for ADaM creation, while in the linear approach SDTM dataset was used as source.

From the original analysis datasets and specifications from the sponsor, ADaM specifications and ADaM datasets are generated. The original ADS were modified to have ADaM structure, taking into account in this transformation the traceability between the new ADaM datasets and the converted SDTM datasets.

The following input is required: Protocol; Statistical Analysis Plan (SAP); Analysis specifications plus analysis program (if available); Analysis dataset (Sponsor structure) plus specifications; Clinical Summary Report (or statistical outputs) if available.

The following output is generated: ADaM datasets, ADaM mapping (specifications), Define.xml, and Reviewer's Guide.

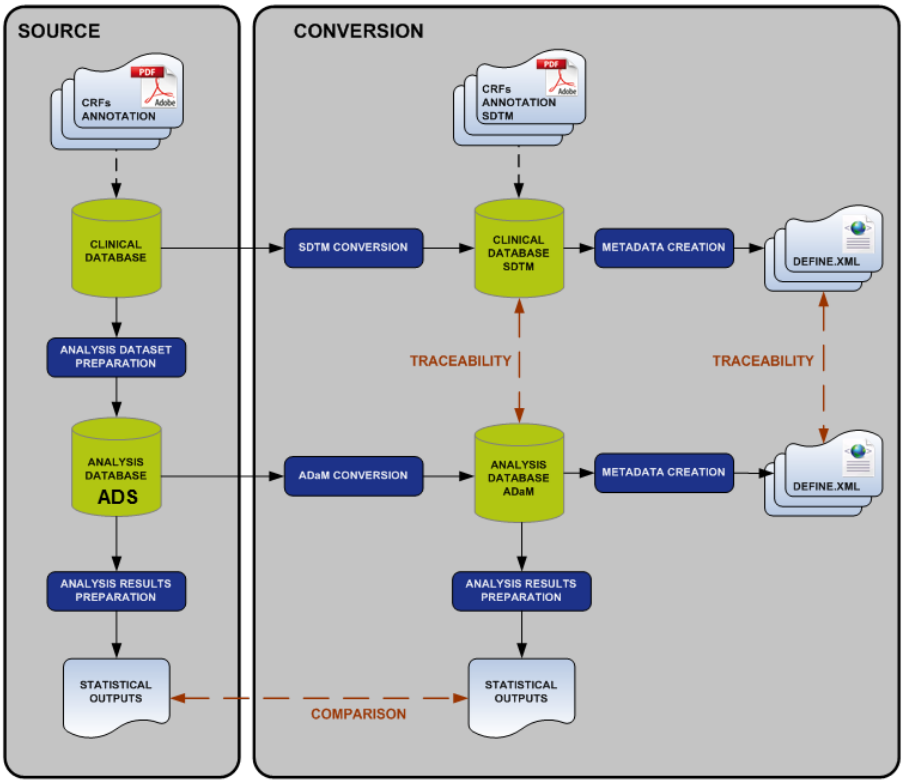


Fig 3: Conversion of original analysis datasets into ADaM datasets, parallel approach

As validation a set of Key TFLs are re-created using ADaM datasets and compared with the TFLs created from the original ADS.

CONVERSION IN A POOLED DATASET

Originally 12 source trial analysis datasets were pooled together into a common structure by the sponsor, this pooling process involved transformation of data values in order to reconcile the trial peculiarities. These pooled analysis datasets then served as source for the conversion to ADaM datasets.

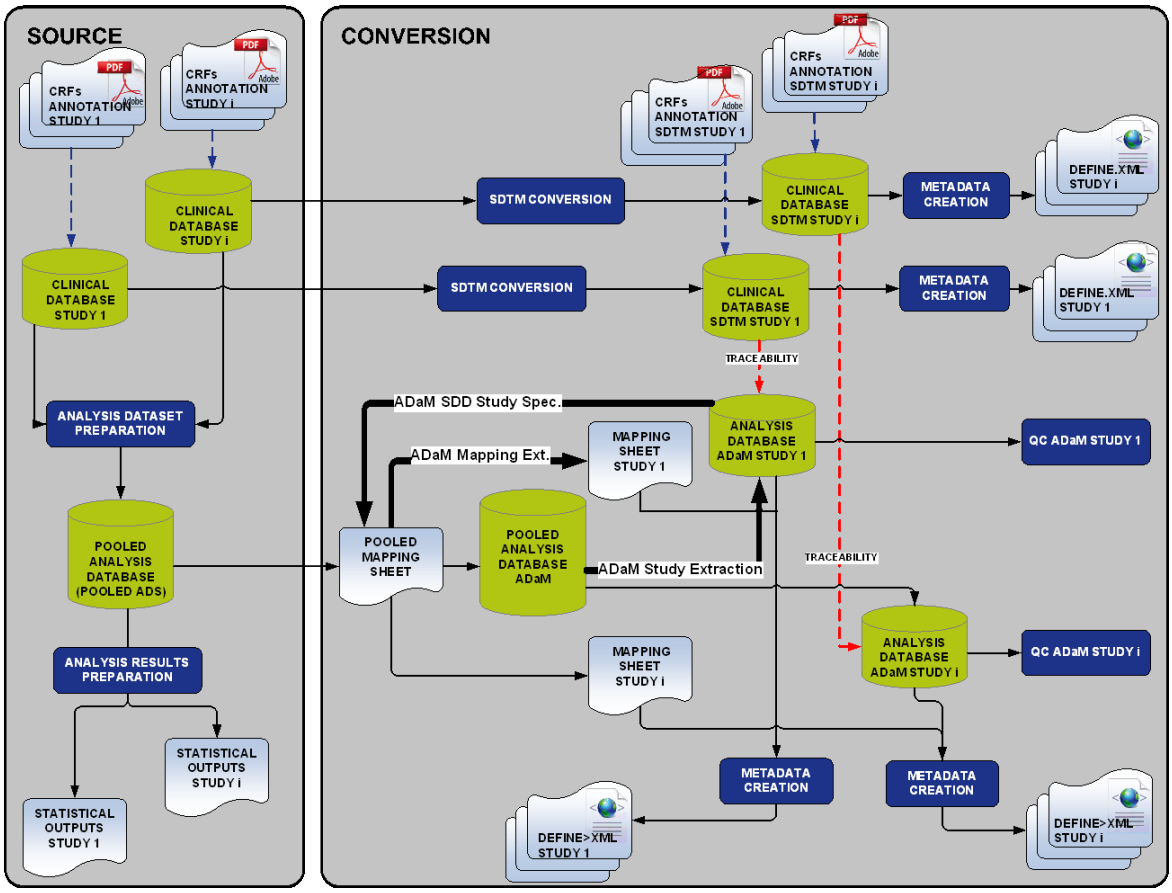


Fig 4: Conversion of pooled analysis into ADaM datasets, parallel approach and extraction process.

The conversion performed can be summarized in the following steps

Creation of ADaM Pooled Mapping: Based on sponsor's derived pooled analysis datasets and their metadata, a detailed 'pooled mapping' workbook was created, describing all conversion specifications from the source analysis datasets to the target ADaM analysis dataset, see fig 5.

DOMAIN PREFIX	VARIABLE NAME	VARIABLE LABEL	VARIABLE TYPE	VARIABLE LENGTH	ROLE	CDISC CT	COMP. ALGORITHM	EXTERNAL DICTIONARY	ORIGIN (name of ADAM dataset or derived)	COMMENTS	DOMAIN PREFIX	VARIABLE NAME	SOURCE FORMAT	CONDITION TYPE for h or i or	Transpose RE VARIABLE	Transpose RE VALUE
ADSL	STUDYID	Study Identifier	char	20	Identifier				DM					h		
ADSL	STUDYOR	Trial Number Origin	char	20	Identifier		ADSL.STUDYOR		Derived		ADS1	SOURCE_VAR1		d		
ADSL	STUDYTP	Individual Study type	char	40	Selection	STUDYTP	ADSL.STUDYTP		Derived		ADS1	SOURCE_VAR2		c		
ADSL	STUDYTPC	Individual Study type (C	char	3	Selection	STDYTPC	ADSL.STUDYTPC		Derived		ADS1	SOURCE_VAR3		c		
ADSL	USUBJID	Unique Subject Identifie	char	20	Identifier				DM	The USUBJID variable has a fixed format: 'XXXX-YYYY-ZZZZZZ', where 'XXXX' indicates the 4-digit compound code, 'YYYY' the 4-digit study code and 'ZZZZZZ' the 6-digit patient code	ADS1	SOURCE_VAR4		d		
ADSL	SUBJID	Subject Identifier for th	char	20	Identifier				DM		ADS1	SOURCE_VAR5		d		
ADSL	SITEID	Study Site Identifier	char	200	Identifier				DM		ADS1	SOURCE_VAR1		c		
ADSL		Center Number	char	8	Analysis											

INSTRUCTIONS_ADAM | B. ADaM table metadata | D. ADaM variables metadata | G. ADaM VLM | I. Codelist metadata | J. Computational algorithm | L. XML parameters ...

Fig 5: Pooled ADaM Mapping (ADaM specifications).In green ADaM details that will be included in the Define.xml, in gray Source information and information on how each variable was mapped into an ADaM variable.

Creation of the Pooled ADaM Datasets: The programming of the ADaM datasets are created based on the pooled mapping and the source pooled analysis datasets using BDLS macros.

Study Extraction – Individual ADaM Datasets: ADaM datasets are extracted from the ADaM pooled datasets and XPTs created. Empty domains and variables are not included.

Study Extraction – Individual ADaM specifications: The study specific mapping is extracted from the pooled mapping (specifications), containing ONLY the variables used in the study that are not empty. Study specific computational algorithms should be updated manually per study.

Reviewer's guide contains study specific long computational algorithms and a subset of definitions from the pooled specifications.

Finally Define.xml (and pdf) are created per each study.

NOTE: any change in the Pooled impact all items produced downstream, this is why extractions can start only when Pooled ADaM datasets and mapping are final.

Pooled mapping contains references to domains, variables, codelist, value-level metadata and computational algorithms that could be involved in one of the studies involved in this ADaM Pooled Database.

We created a set of 40 "pooled" domains that contained data from 12 studies, with some domains having no records for a number of specific studies.

In order to limit to a minimum the need for manual input for each extraction and avoid many errors induced by repetition, BDLS have created a series of SAS ® programs combined with SAS-produced Excel sheets. Therefore we have SAS macros designed to:

- Extract individual ADaM study datasets from the ADaM pooled datasets
- Extract individual ADaM mapping from the ADaM pooled mapping
- Identify variables mapped/not mapped in SDTM for each study
- Track the updates in specifications

The pooled mapping had one extra column per study extracted, to make this file "extraction-oriented". The extraction process was reduced to updates of SAS macro parameters and (re)running SAS programs (along with some copy/paste-special operations in Excel).

DOMAIN PREFIX	VARIABLE NAME	VARIABLE LABEL	VARIABLE TYPE	VARIABLE LENGTH	ROLE	USER DEFINED FORMAT	CDISC CT	COMP. ALGORITHM	EXTERNAL DICTIONARY	(name of ADAM dataset	COMMENTS	MAIN PRE	VARIABLE NAME	SOURCE FORMAT		STUDY 0001	STUDY 0002	STUDY 0003	STUDY 0004	STUDY 0005	STUDY 0006
ADAE	AESDT	Start Date of Adverse Event	num	8	Timing	DATE9.		ADAE.AESDT		Derived		AE	AEOND		c	YES	YES	YES	YES	YES	YES
ADAE	AESTM	StartTime of Adverse Event	num	8	Timing	TIME8.		ADAE.AESTM		Derived		AE	AEONT		c	YES	YES	YES	YES	YES	YES
ADAE	AESTDTC	Start Date/Time of Adverse Event	char	20	Analysis					AE		AE	AEONY		d	YES	YES	YES	YES	YES	YES
ADAE		Start Date Imputation Flag	char	3	Analysis		DATEFL														
	AESDTF							ADAE.AESDTF		Derived		AE	AEONIMP/AEONYMD		d	YES	YES	YES	YES	YES	YES
ADAE	AEEDT	End Date of Adverse Event	num	8	Timing	DATE9.		ADAE.AEEDT		Derived		AE	AEEND		c	YES	YES	YES	YES	YES	YES
ADAE	AEENDTC	End Date/Time of Adverse Event	char	20	Analysis					AE		AE	AEEND		d	YES	YES	YES	YES	YES	YES
ADAE		Analysis End Date Imputation Fla	char	3	Analysis		DATEFL														
	AEDTF							ADAE.AEDTF		Derived		AE	ENDYM		d	YES	YES	YES	YES	YES	YES
ADAE	AEENRF	End Relative to Reference Period	char	20	Analysis		AEENRF			AE		AE	AEEND		f	YES	YES	YES	YES	YES	YES
ADAE	BGLU	Blood Glucose Value	char	25	Analysis		BLOODGLU	ADAE.BGLU		Derived		AE	BGLU		f	YES	YES	YES	YES	YES	YES
ADAE	BGLUN	Blood Glucose Value (N)	num	8	Analysis		BLOODGLN	ADAE.BGLUN		Derived		AE	BGLU		c	YES	YES	YES	YES	YES	YES
ADAE	AETERM	Reported Term for the Adverse E	char	200	Analysis					AE		AE	AEMN		c	YES	YES	YES	YES	YES	YES
ADAE	AESYM	Adverse Event Symptom	char	40	Analysis		AESYM	ADAE.AESYM		Derived		AE	SYMAE		f	NO	NO	NO	#N/A	NO	NO
ADAE	AESYMN	Adverse Event Symptom (N)	num	8	Analysis		AESYMN	ADAE.AESYMN		Derived		AE	SYMAE		c	NO	NO	NO	#N/A	NO	NO
ADAE	AELLT	Lowest Level Term	char	200	Analysis					AE		AE	AELLT		c	YES	YES	YES	YES	YES	YES
ADAE	AELLTCD	Lowest Level Term Code	char	10	Analysis					AE		AE	AELLTC		c	YES	YES	YES	YES	YES	YES
ADAE		Action Taken With Study Treatm	char	40	Analysis		AEACN														
	AEACN									AE		AE	AEACTA/AEACTE		d	YES	YES	YES	YES	YES	YES

Fig 6: ADaM mapping showing extra columns per each study used for extraction of Study specific mapping.

The only manual operation left to the statistical programmer was the update of the computational algorithms (if necessary) in the pooled ADaM mapping AND the update (if necessary) of the study-specific algorithms, along with the update of the reviewer's guide.

QC was done using BDLS Data Model Compliance Checker (DMCC) to check ADaM metadata as well as ADaM data validation checks to ensure consistency between the target datasets and ADaM Define.xml. The QC was combined with specific programmed SDTM/ADaM consistency checks. The automated validation process was completed by a final quality control for release including a check-list for a manual review supported by a QC content tool on source and target, and a 100% manual QC on a random sample.

CASE STUDY, CONVERSION PARALLEL APPROACH

This was a conversion for a set of studies from the same therapeutic area, containing approximately 40 pooled analysis datasets from 12 studies into Pooled ADaM datasets and its posterior extraction. As validation the client recreated TFLs using final ADaM datasets per each study.

Mapping decisions, clarifications of ADS specifications and source data issues were documented and reported to the sponsor periodically.

THE GOOD:

SDTM/ADaM traceability was ensured due mainly to a good communication between the two teams

In this type of conversion as only the structure of the source analysis datasets is changed, the risk of not being able to reproduce TFLs is low or null. As BDLS was in charge of SDTM and ADaM conversion, this facilitated the communication between the two teams and sped up the solution of any inconsistency. Although the development of a Pooled dataset is a long process, as processes mature, data extraction for each study became very fast, and consistency checks across studies were no longer needed.

SDTM variables that were re-used in ADaM, such as VISIT, ACTARM, ARM were derived consistently in SDTM across studies, therefore codelist were re-usable in ADaM due to its consistency across studies. Non subject-oriented datasets (e.g. IPV data, dictionaries, etc.) were added as extra datasets, but not as part of SDTM.

Consistently with the conversion using the linear approach, all derived variables not mapped in SDTM but used in derivation rules in ADaM were, when necessary and possible, added in SDTM. We ensured that all SDTM questions were addressed before SDTM was final and before the QC of ADaM datasets. In case of derived variables from a Data Collection Module System (DCMS), or ADS variables generated from standard macros from the sponsor, the SDTM variable was not available. Therefore this variable had to be redefined as being derived in ADaM and not as an SDTM variable. BDLS and the sponsor worked together to identify a proper derivation rule.

The creation of a comment file with all our questions from and to the sponsor, facilitates to keep a track of all questions answers and decisions

THE BAD:

Derivation rules were not always well documented, this caused difficulties to assess traceability. For example Pooled derivation rules did not always include a general/individual rule applicable to all studies. Some studies were added into the pooled in different times and derivations were not properly updated.

For each study extracted, the study specific derivations have to be updated manually after the extraction.

SDTM limitations affected the conversion, and we found raw variables that were not available in SDTM, but were mentioned in ADS specifications. Some of them were not allowed to be included in SDTM, because they were derived variables from a Data Collection Module System (DCMS). For these cases we redefined them with the help of the sponsor. In other cases some raw variables were not mapped in SDTM in the same domain, but found in another domain. E.g. VISITNUM, VSDTC not mapped in AE or CM.

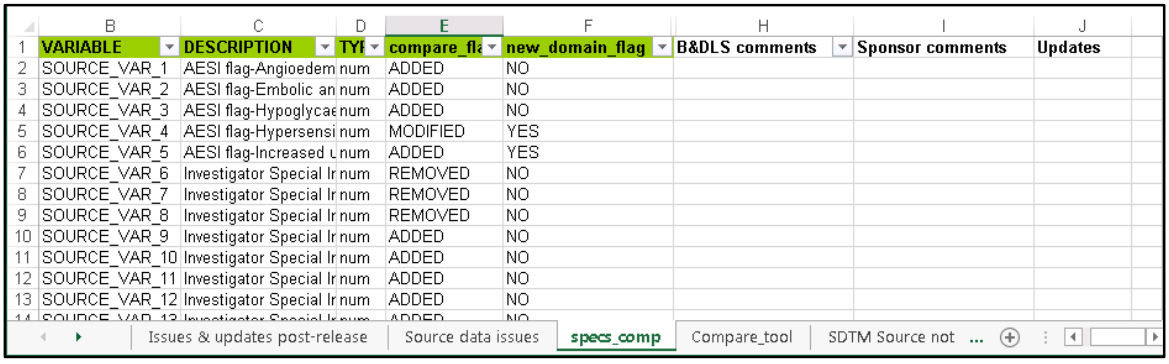
SDTM changes in codelist or maximum length affected directly ADaM, e.g. comments bigger than 250 char were split by SDTM. Different codelist values were used for the same variable name in different studies e.g. AESER “Yes”, “No” versus “Yes”, “No”, “No, but significant”.

And due to traceability between SDTM/ADaM, timelines for ADaM datasets were influenced by SDTM timelines.

THE UGLY:

The structure of the pooled original ADS datasets changed with the inclusion of new studies. These changes impacted the pooled mapping, pooled datasets and extractions. At variable level the 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. A change in the content of the source implied a re-definition of the conversion rule applied to the variable. Therefore, 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

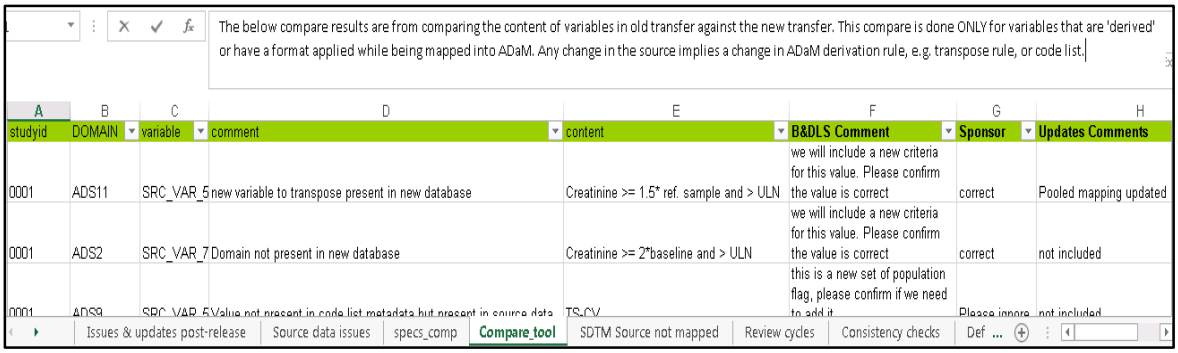
before. The change in content was compared for variables which had a codelist applied, variables that were transposed or derived during the ADaM conversion.



	B	C	D	E	F	H	I	J
1	VARIABLE	DESCRIPTION	TYI	compare_fla	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 u	num	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			

Issues & updates post-release | Source data issues | specs_comp | Compare_tool | SDTM Source not ...

Fig 7: Results of a compare of Original ADS metadata, included in the communication file to the sponsor.



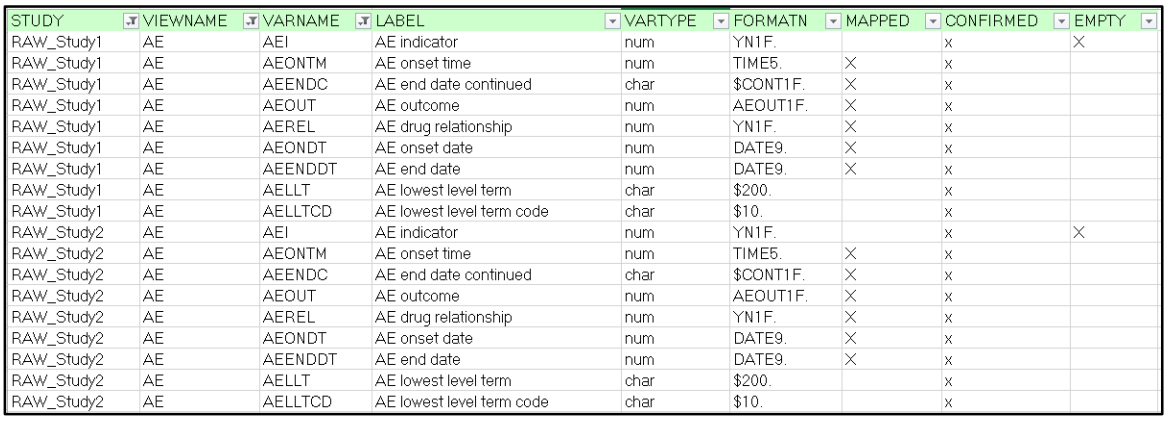
The below compare results are from comparing the content of variables in old transfer against the new transfer. This compare is done ONLY for variables that are 'derived' or have a format applied while being mapped into ADaM. Any change in the source implies a change in ADaM derivation rule, e.g. transpose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

Fig 8: Results from compare content in variables being “derived”, “has a CL applied”, “transposed” during ADaM conversion, included in the communication file to the sponsor.

To ensure traceability with SDTM, a good knowledge of what was mapped from RAW into SDTM and in which study, was essential. Considering the amount of studies involved we needed to create a tool to identify per study which variable was mapped in SDTM.



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	

Fig 9: List of RAW variables per study per domain from each SDTM metadata, specifying if the variables was mapped into SDTM, and if it was empty.

We noticed that some variables were present in the specifications, but were empty in the original ADS, BDLS confirmed with the sponsor a complete list of empty variables per study before starting the conversion.

In other cases derived variables from a Data Collection Module System (DCMS) were used as ADS key variables. As they were derived, they were not available from SDTM, and were not included in ADaM. Each case was investigated separately. Together with the sponsor either a new derivation rule was established or an explanation in the Reviewer's guide added, or the variables was considered not relevant and therefore ignored in ADaM.

Changes in raw datasets structure for new studies were not immediately communicated to the ADaM team. Due to the good communication with the SDTM team we were informed of any change before SDTM was final.

CASE STUDY, DICTIONARY UPDATES

MedDRA ® and WHO Code variables were updated to a newer dictionary version than the original. These updates were implemented in 9 of the 12 studies part of the pooled database explained in the above section. As ADS source we received from the sponsor only the Pooled datasets affected by dictionary updates.

The Pooled mapping was updated accordingly, and all computational algorithms mentioning dictionary variables or its content, (like --LLT, --HLT, etc.) were consistently modified. 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.

SDTM dictionary version had to be identified because not all studies in SDTM had the same dictionary version. This was clarified in the Reviewer's guide of each study.

THE GOOD:

Once we identified which changes we could expect in the new transfer were related to dictionary updates, the same procedure was applied to the other studies. As this was a mechanical process, it was sure, simple and fast.

Reviewer's guide updates were very similar in each study. Once the changes required had been established for the first study, these could be applied quickly across the remaining studies.

THE BAD:

During our initial compare of the source, old ADS versus new ADS with dictionary updates, we found some new variables that shouldn't be the result of a dictionary update. This was at the sponsor side, while re-running general pooled macros that were updated for newer studies part of the pooled. This generated new derived variables not used in old studies analysis, but now part of the Pooled.

Due to new variables added into the Pooled for new studies, BDLS macros failed when re-running the poled ADaM datasets. ADaM mapping had to be updated, and new macros were developed to detect NEW variables not mapped previously, that were present in the new original ADS.

A list of changes per dataset per study was checked with the sponsor. The origin of each difference was identified, and final changes were confirmed with the sponsor before converting the original ADS datasets to the ADaM pooled for its posterior extraction.

THE UGLY:

There was not really an ugly face here, and the process went fast after the second study and we received several transfers per each study on a frequent basis. To support this, the folder structure where we stored source and target datasets for old studies had to be updated, and we kept track of the original ADS used and final version of ADaM sent. By doing so, we were able to compare dataset versions. A discrepancy report was created to list differences found, detailing their origin and resolution.

CONCLUSIONS:

This type of conversions is very useful when the original analysis datasets are not created in ADaM standard. With the parallel approach we ensured reproduction of TFLs and consistency between the studies part of the pooled. With the linear approach SDTM/ADaM traceability is demonstrated, but the risk of not reproducing identical results is greater.

The source pooled ADS should be stable and specifications clear, for the creation of draft mapping and draft ADaM datasets. The sponsor's feedback on draft mapping and draft ADaM datasets is crucial to address all details up front.

In the case of a pooled dataset, ideally Pooled ADaM datasets should be finalized before the extractions can start and after SDTM is final for all studies included in the pooling. It is important to remember that any change in the pooled dataset may affect and influence the process upstream.

If new studies are added to the pooled ADS, some checks should be performed at each new transfer. Content of variables should be checked to identify any that may have changed during the conversion due to a derivation, a transposition or a codelist being applied.

During a dictionary update, ideally the SDTM dictionary version is the same as in ADaM. If not there are specific ADaM variables that can be used to include an updated version of the dictionary. The process should be relatively straightforward, and easy to implement, but we did encounter some issues due to the lack of knowledge of how the sponsor updated their pooled datasets.

Good communication is needed between SDTM, ADaM team members and the sponsor, during the whole process and prompt reactions from each side is necessary. All decisions should be easily tracked if they are properly documented in a communication file.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Carla Santillan
Business & Decision Life Sciences
141, Rue Saint-Lambert
1200, Bruxelles
Belgium
Work: 0032 2 774 1100
Fax: 0032 2 774 1199
Email: carla.santillan@businessdecision.com