PhUSE 2012

Paper CD01

ADaM or SDTM? A Comparison of Pooling Strategies for Integrated Analyses in the Age of CDISC

Joerg Guettner, Bayer Pharma AG, Wuppertal, Germany
Alexandru Cuza, UCB Biosciences GmbH, Monheim, Germany

ABSTRACT

With the two CDISC Standards, Study Data Tabulation Model (SDTM) and Analysis Data Model (ADaM), there are basically two different pooling strategies of single study data possible for an integrated analysis. The first one is to pool the SDTM data sets and to derive the analysis data sets from the pooled SDTMs. The second one is to pool the single study analysis data sets and make all needed derivations for the integrated analysis during the pooling process. This paper compares the two approaches on the basis of actual experiences with submissions in two companies following the two different strategies. As most companies are in a transition phase at the moment from legacy, i.e. non-CDISC standards, to CDISC standards, these submissions contain also single studies based on legacy standards.

INTRODUCTION

In the past each sponsor built integrated databases for submissions based on his own standards. There were usually only few analysis data sets or even derived variables. Most derivations were done on the fly in the analysis programs creating the tables, figures and listings. Traceability was not taken care of.

Nowadays this has changed completely with CDISC. There are now two different standards available:

- Study Data Tabulation Model (SDTM)
- Analysis Data Model (ADaM)

Traceability is of major priority now. But as most companies are in a transition phase from legacy standards to the CDISC standards, studies in legacy standards need to be included often also in the integrated analysis together with studies already in one of the CDISC standards.

The goal is clear as going back to legacy standards should be no option. An integrated database using the ADaM standard has to be created. But how can this be done? With the two different CDISC standards there are basically two different strategies of creating such an integrated database:

1. The first strategy is to pool all the single study SDTMs and to derive the analysis datasets from the pooled SDTMs.
2. The second one is to pool the single study analysis datasets and to make all needed derivations during the pooling process.

These two approaches are compared on the basis of actual experiences with submissions in two companies following the two different strategies.

STRATEGY BASED ON SDTM (UCB SUBMISSION)

The data flow from collection to analysis is illustrated in Figure 1.
The collected data are entered by Data Management into actually ONE database or legacy databases in the past. The ONE database was implemented 2009-2010 in order to have a unique database for data management activities, unify the several legacy ones being in place over time and be closer to SDTM standards.

After data management locked the database SDTM panels are created. For the study level analysis ADaM analysis datasets (ADS) are created according to statistical analysis plan (SAP).

For integrated analyses the study SDTM are pooled together and consolidated regarding controlled terminology (e.g. several versions over time might have been used in different studies) and other CDISC SDTM specifications.
Based on the pooled SDTM based on integrated SAP (ISAP) ADaM based ADs are created.

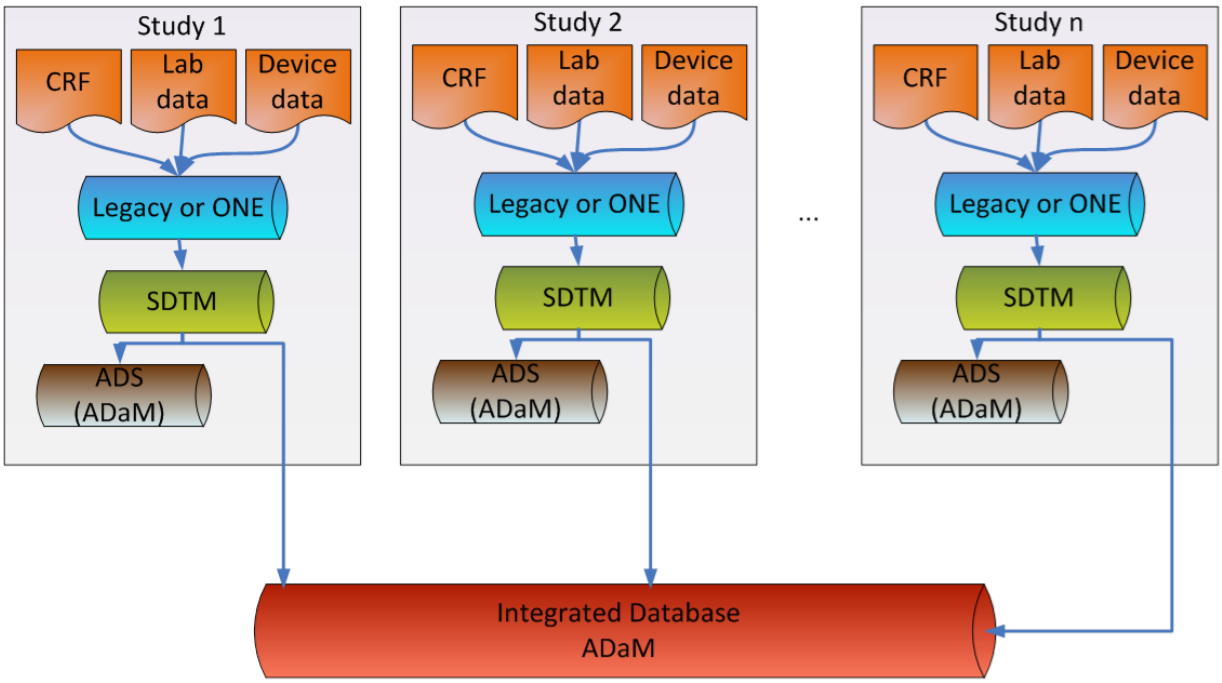

Figure 1: UCB Data Flow

EXPERIENCES WITH UCB SUBMISSION

Thirteen Phase III studies with placebo-controlled plus open label extension and a Phase IV study were integrated. Four studies were reported in a time before CDISC standards were introduced; therefore it was needed to create SDTM for them first. These studies having raw data in three different legacy structures, an increased amount of time was needed to map to SDTM.

SDTM were pooled and controlled terminology consolidated to be consistent over all data. Together with the four studies mentioned above (legacy structure) this was the second time/resources consuming part.

Based on ISAP ADaM analysis datasets were created out of the SDTM pool. Due to the high level of standardization in SDTM and with ADaM guidelines (CDISC and internal) in place, this was easier than expected.

STRATEGY BASED ON ADaM (BAYER SUBMISSION)

The planned data flow from the captured data to the integrated database is displayed in Figure 2.

From all the collected data (electronic data capture (EDC), laboratory data and device data, like ECG) of a single study data management creates a database called Operational Acquisition Database (OAD). The OAD contains all variables that are necessary to extract SDTM and the SDTM supplemental datasets, but also additional variables, like variables containing a code (e.g. the MedDRA code of AEDECOD (Dictionary-Derived Term) or a numerical representation of variables, like AEOUT (Outcome of Adverse Event)). Therefore the SDTMs and the SDTM supplemental data sets can be easily extracted from OAD.

But also statistics can create their analysis datasets (ADS) using OAD. Especially there is no need to add variables containing a code to SDTM variables as these variables are already in the database. Statistics adds all variables and derives new records as needed for their analyses and required by ADaM.

For the integrated database just the ADS of the single studies needs to be pooled. All needed derivations of variables or records are done during the pooling of each study.

The integrated database contains almost all information collected in the OAD as well in the ADS. If a new record needs to be added to the integrated database (e.g. to harmonize derivations), the original record of the single study stays in the integrated data base. To assure traceability to single study results there are two flags in the database, marking which records were used by the single study and which by the integrated database.

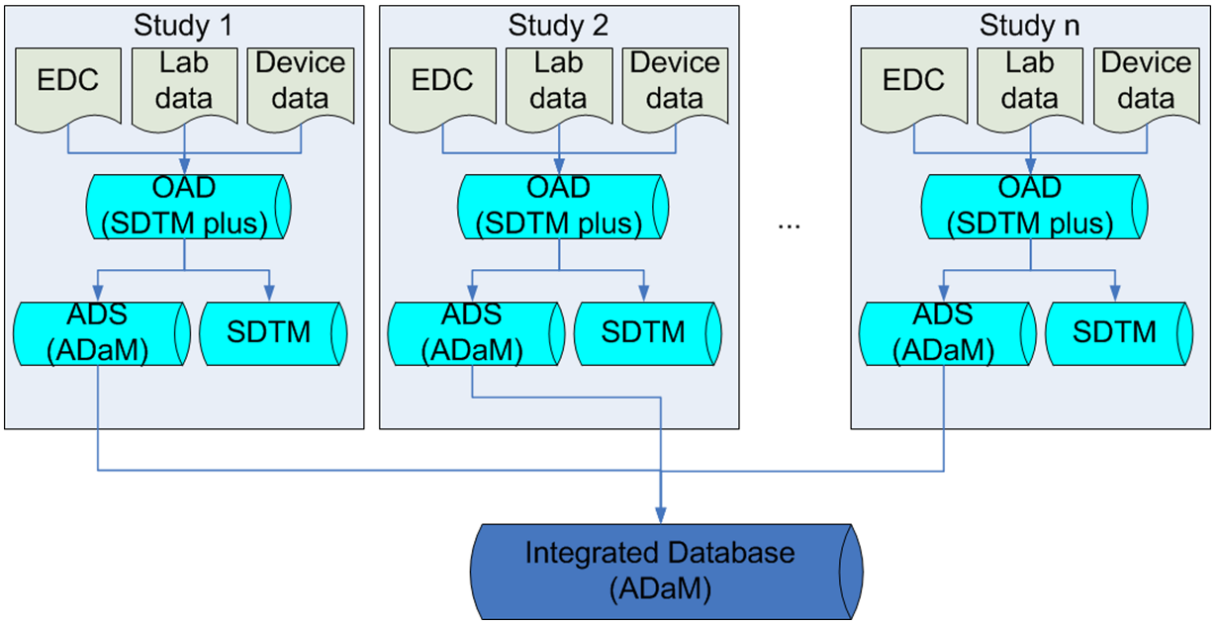

Figure 2: Planned Data Flow

The actual data flow for the Bayer submission differs quite a lot from the theoretical one and is displayed in Figure 3. This submission was done together with an external partner and included six phase I and phase II trials (including one extension trial) and the two pivotal trials.

All the phase I and phase II trials were conducted by different contract research organizations (CRO) using their own CRO legacy standards. Also the two pivotal trials used a legacy company standard.

As this project was just the first indication for submission and all other indications of this compound already use SDTM and ADaM based standards, it was decided to create ADS based on ADaM. Unfortunately it was also decided to build the integrated database based on the legacy CRO DM data (instead of SDTMs) and for the pivotal trials on

the ADaM ADS. The SDTMs for all trials were built in parallel to the pooling process and the ADS creation of the pivotal trials to save time in the submission process.
The parallel development of SDTM on the one side and the ADS and integrated database on the other side made reconciliation necessary.

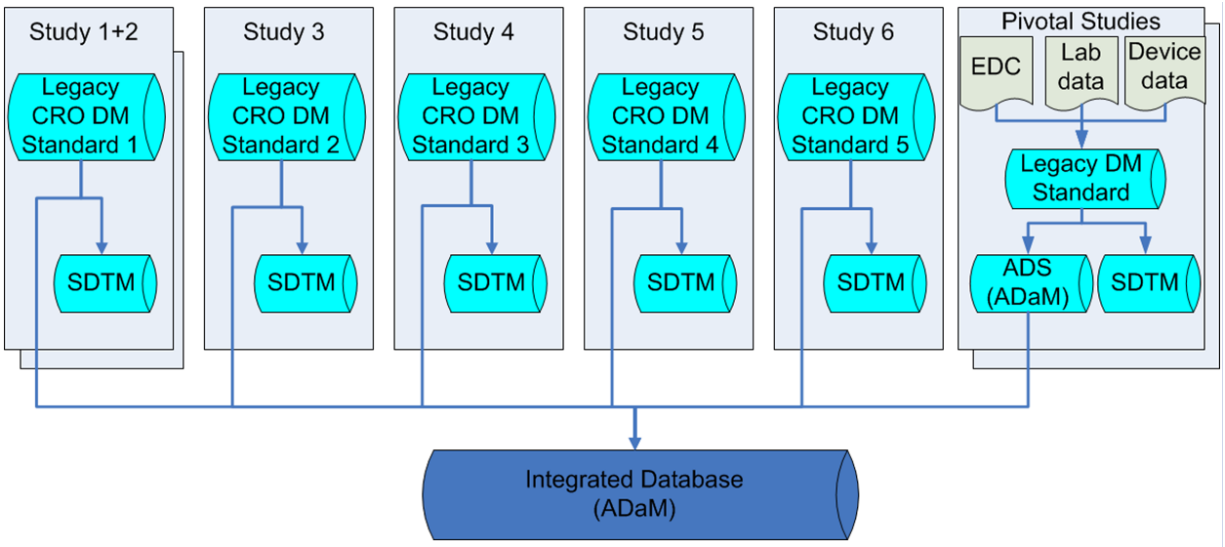

Figure 3: Actual Data Flow

EXPERIENCES WITH BAYER SUBMISSION

The first part of building the integrated database was to pool all the phase I and II studies using different legacy CRO data management standards. There were only two out of six studies using a comparable standard. It was necessary to bring first all studies to the OAD, i.e. SDTM like, structure and then to derive all analysis variables and records. The derivation of analysis variables and additional records was primarily done with macros that could be used on all studies after the first harmonization part (i.e. converting to SDTM like structure). This was a very time and resource consuming process. Especially the first harmonization part took a lot of time as a lot of remapping was done to assure the use of CDISC controlled terminology and the Bayer standard codelists.
After the experience with the phase I and II trials, the second part of building the integrated database was surprisingly easy. The statistical analysis plans of the two pivotal trials as well as of the integrated summaries of efficacy and safety were well harmonized. Therefore it was quite easy to pool these trials. There was only one exception. During the validation process it was recognized that for one parameter different tests were used in the pivotal trials. To assure traceability to the single study data and results, this was solved by adding additional records to the integrated database with a harmonized parameter.

COMPARISON OF THE ADVANTAGES AND DISADVANTAGES OF BOTH STRATEGIES

Both strategies have their own prerequisites and advantages, but also disadvantages.

REQUIREMENTS

The strategy based on SDTM has no special requirements. But of course it is of advantage, if studies are using the same raw data standard, e.g. the ONE database.
For the strategy based on ADaM it is best to have standards for the data, the metadata, the codelists and the derivations. If all these standards are all the same for the single trials and the integrated database, it is very easy to pool as only variables and records for the integrated database need to be added.

TIME AND RESOURCES

As all derivations (ADaM and for the integration) can be done directly on SDTM like a single study, only limited time and resources are needed for the pooling strategy based on SDTM.
The pooling of ADaM has the advantage that many derivations are already available in the single studies. I.e. there is no need to derive variables or records again for the integrated analysis as long as all studies use the same rules for derivation. Therefore only very limited time and resources are needed for the strategy based on ADaM.
Legacy studies increase the amount of time and resources needed in both approaches. Especially for the approach based on ADaM much more resources and time were needed for studies using legacy standards.

TRACEABILITY TO SINGLE STUDY DATA

The strategy based on SDTM is easily traceable from the integrated analysis datasets (ADaM) back to the single study SDTMs as all datasets are within the CDISC standard.
In theory also the Bayer strategy would be traceable from the integrated analysis datasets (ADaM) back to the single study SDTMs. But as the SDTMs were created parallel to the analysis datasets this is only possible after a reconciliation of the analysis datasets with the SDTMs.
Traceability is sometimes very challenging for studies using legacy standards, e.g. when data are shifted from a horizontal to a vertical structure.

TRACEABILITY TO SINGLE STUDY ANALYSIS

For the strategy based on SDTM it is very difficult to have traceability from the integrated database back to the single study results, if different algorithms were used for derivations.
In the approach based on ADaM all records coming from the single study remain unchanged in the integrated database. If a change to such a record is necessary (e.g. due to a different derivation algorithm) an additional record is added and both records are flagged with the information, if the record was used by the single study analysis or the integrated analysis. Therefore differences between single study and integrated analysis are easy to identify.
Studies using legacy standards are not traceable from integrated to single study analysis in both approaches.

COMPLEXITY OF TRANSFER PROGRAMS AND VALIDATION

For the strategy based on SDTM the complexity of transfer programs and therefore validation is low as the reusability of code is high.
As only few derivations are needed the complexity of transfer programs and validation is also low for the strategy based on ADaM.

For studies using legacy standards remapping to controlled CDISC terminology is needed in both approaches. So the complexity of transfer and validation is much higher.

HARMONIZATION OF ALGORITHMS AND DERIVATIONS

In the strategy based on SDTM no harmonization is needed as the integrated ADaM panels are directly created out of SDTM.
In the case that there were different derivation algorithms used in the single studies, there is a need to harmonize these algorithms for the strategy based on ADaM. Identifying that different algorithms were used without looking at the source code is very difficult. At Bayer different derivation algorithms are usually identified by comparing the single study SAPs and the different versions of the project SAP that were used to create the single study SAPs.

ADDITIONAL TASKS

Reconciliation of controlled terminology is needed due to the fact that different ages of SDTM controlled terminology are being usually integrated in both approaches.
Because of the decision to develop the SDTMs in parallel to the analysis datasets, it was necessary to do reconciliation between the SDTMs and the analysis datasets of the integrated database of the Bayer submission.

CONCLUSION

The strategy based on SDTM and the strategy based on ADaM are both acceptable pooling strategies. They have their own prerequisites and advantages, but also disadvantages. Unfortunately there is no strategy that is always preferable to the other. It depends always on the situation. If you live in a well standardized environment (i.e. standards for data, metadata, derivations and analysis) and if you are willing to spend the time and resources to integrate legacy studies, the strategy based on ADaM is be a good choice as it ensures traceability to the single study results. In less standardized environments the strategy based on SDTM is recommendable as integrating legacy studies is not that time and resource consuming.

ACKNOWLEDGMENTS

We would like to thank Karl-Stephan Neufeldt (UCB) for his valuable input and contributions towards this paper.

CONTACT INFORMATION

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

Joerg Guettner
Bayer Pharma AG
Aprather Weg 18a
42113 Wuppertal, Germany
Work Phone: +49 202 36 5221
Fax: +49 202 36 8477
Email: joerg.guettner@bayer.com
Web: www.bayer.com

Alexandru Cuza
UCB Biosciences GmbH
Alfred-Nobel-Str. 10
40789 Monheim, Germany
Work Phone: +49 2173 48 1191
Fax:
Email: alexandru.cuza@ucb.com
Web: www.ucb.com