Customer oriented CDISC implementation

Edelbert Arnold, Accovion GmbH, Eschborn, Germany
Ulrike Plank, Accovion GmbH, Eschborn, Germany

ABSTRACT

The Clinical Data Interchange Standards Consortium (CDISC) represents the most important development in data exchange within the pharmaceutical industry and between the industry and regulatory authorities. After recommendation by the Food and Drug Administration (FDA), this data standard has been rapidly implemented by regulatory bodies such as the FDA, contract research organizations (CROs), pharmaceutical / biotechnological companies and other data suppliers (e.g. laboratories, and users of electronic patient record outcomes). Depending on the scope and extent of internal standardizations (e.g. global data dictionaries, and standard analysis tools in SAS ®) companies may choose different concepts for implementation of CDISC standards.

The CRO Accovion (formerly Covidence) has chosen a fully integrated approach to the implementation of CDISC. This covers data collection (case report form and database design), a standard process for the generation of CDISC datasets, a mapping tool to transfer clinical data from other formats into a CDISC compliant structure, and standard SAS macros which are used for analysis. The main points outlined in this paper are as follows:

- Database set-up in the clinical data management system, adhering to CDISC-SDTM as much as possible
- Generation of datasets based on Study Data Tabulation Model (SDTM) and Analysis Dataset Model (ADaM)
- Integration of different data structures, e.g. from legacy studies
- Analysis output

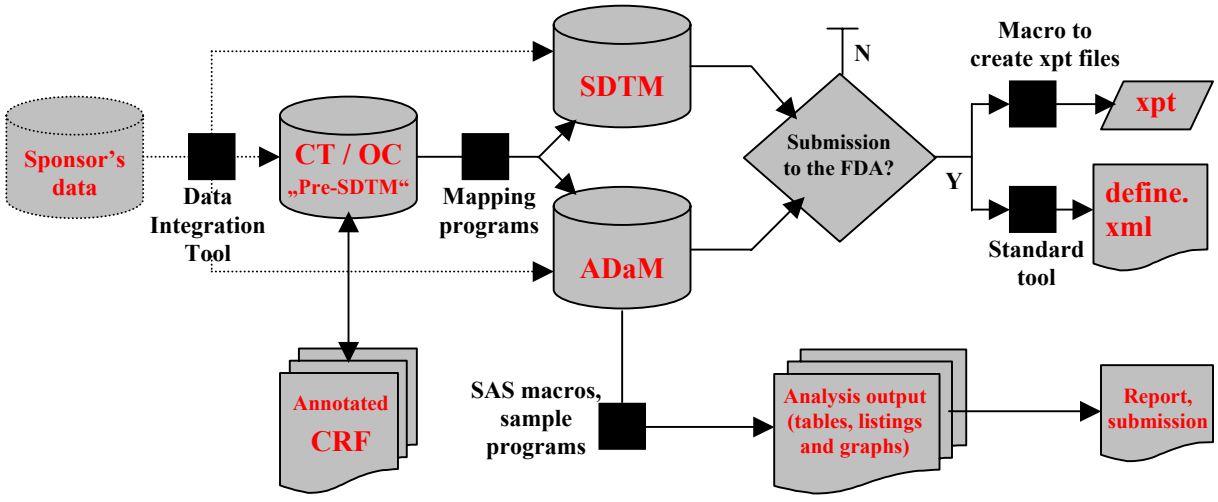
All parts of the implementation are investigated for overall efficiency, flexibility with respect to sponsor requirements, and compliance with FDA expectations.

INTRODUCTION

The set up of studies in the Accovion clinical database management systems Oracle Clinical (OC) and Clintrial (CT), adheres to CDISC-SDTM data structures as much as possible. Metadata libraries have been defined that allow efficient, compliant and rapid set up of study databases. This “Pre-SDTM” data structure forms our internal platform to feed our standard process for the generation of SDTM and ADaM datasets. Together with our set of standard analysis macros, this improves efficiency with a simultaneously reduced risk of errors.

Sponsor specific database structures can be transferred either directly to SDTM and ADaM or to the Accovion standard database structure that operates with our standard tools. Such data transformation is performed using a data integration tool, which can also be used for mapping data from various clinical trials into an integrated database.

The following graph represents the main aspects of the Accovion strategy:



A review of the current FDA expectations is helpful to identify the benefits of selecting an appropriate implementation strategy.

DELIVERABLES TO THE FDA

The FDA is the most influential drug agency and has been a key driver behind the development of CDISC standards. Following deliverables are expected by the FDA to be included in a submission:

- SDTM:
  - Datasets, currently still as SAS V5 transport files
  - define.xml (metadata and links)
  - Annotated case report form (annotated CRF)
- ADaM
  - Datasets, currently still as SAS V5 transport files
  - Documentation (metadata and links, similar to define.xml)
  - Dataset generation programs (currently recommended and reviewer specific, but should be available on request)
- Analysis output
  - Outputs in reports and summaries
  - Documentation (metadata and links)
  - Output generation programs (currently recommended and reviewer specific, but should be available on request)
- Optional
  - Supplemental documentation
  - Additional data listings or specific patient profiles

The FDA will not need data listings and patient profiles in the future. By uploading SDTM data into the FDA data warehouse and by using standard review tools the FDA is able to create its own data listings and patient profiles.

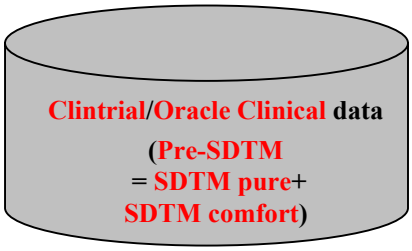
DATABASE SET-UP IN THE CLINICAL DATA MANAGEMENT SYSTEM (CDMS)

In order to set up a database for a new study (e.g. first study in a new drug project), it is recommended to follow the SDTM guidelines as closely as possible so as to adhere to the format expected by the FDA. However, depending on the technical constraints of the software environment, there are several solutions available. The following questions should be considered when determining a strategy for setting up a SDTM compliant database:

- How flexible is the underlying CDMS / Is a change or update of the CDMS envisioned?
- To what extent does the CDMS support metadata libraries and enforce their use?
- How flexible is the CDMS with respect to changes to the database (e.g. derivations) made at any time / What is the impact of late changes to ongoing data cleaning processes?
- How flexible and transparent is the CDMS in comparison to other tools (e.g. SAS) in terms of inclusion of more complex derivations/ How standardized are these more complex derivations?
- Are additional variables needed for data analysis tasks that are not defined in SDTM?

The vast majority of the SDTM Implementation Guide (SDTM-IG) can be easily followed in the CDMS. However, a fully integrated approach for implementation from data collection to data analysis and data integration requires a sophisticated strategy. Within Accovion, Clintrial and Oracle Clinical are available and the following strategy has been identified:

- Global libraries based on SDTM-IG were defined. These contain all SDTM variables and metadata as far as possible (i.e. the “SDTM pure” part).
- In addition, the global libraries contain variables that are beneficial for data analysis and reporting which are not required within SDTM (i.e. the “SDTM comfort” part).
- Structures from global libraries are copied to study level and adapted as necessary. These are made available for extraction in SAS.
- The CDMS does not (and cannot) contain a 100% compliant implementation of SDTM and therefore is a “Pre-SDTM”, because for some datasets and variables the realization in SAS seems to be more convenient or even necessary.



Our standard database setup will follow CDISC SDTM as long as these standards do not interfere with the sponsor’s requirements. This database structure builds the basis for further standard procedures of SDTM/AdaM generation and creation of analysis outputs. If the sponsor provides a different database setup, please refer to the chapter entitled “Integration of Different Data Structures”.

SDTM-PURE AND SDTM-COMFORT

The Accovion CDMS standard structure is composed of two categories: “SDTM pure” and “SDTM comfort”:

- SDTM pure: Variables which are requested and standardized by the SDTM Implementation Guide
- SDTM comfort: Variables which are not required by SDTM, but which are useful for the generation of ADaM

Examples:

- SDTM pure:
  - EG.EGDTC: Date/Time of ECG in ISO8601 format
  - DS.DSDECOD: Controlled Terminology for DSTERM (disposition event)
- SDTM comfort:
  - EG.EGDT: Date of ECG (numeric with format DATE9, used for analysis based on ADaM)
  - EG.EGDC: Date of ECG (character with format DD-MMM-YYYY, used for listings based on AdaM)
  - DS.DSDECODN: Numeric Code for Controlled Terminology for DSTERM (used for analysis based on ADaM)

DSDECOD	DSDECODN
ADVERSE EVENT	1
COMPLETED	2
…	…
NON-COMPLIANCE WITH STUDY DRUG	6
…	…
STUDY TERMINATED BY SPONSOR	13
…	…
OTHER	999

Note: If a selection for some terms is required, "DSDECODN IN (1,6,13)" is more convenient and less error-prone in terms of typing errors than "DSDECOD IN ('ADVERSE EVENT', 'NON-COMPLIANCE WITH STUDY DRUG', ‘STUDY TERMINATED BY SPONSOR’)".

During creation of the final SDTM, the SDTM comfort variables will be removed, but the SDTM pure variables will be retained.

Note, that CDMS specific system items are not further considered in this paper.

PRE-SDTM

As mentioned earlier, the Accovion global library does not totally reflect the CDISC-SDTM. Consequently the CDMS can be seen as a “Pre-SDTM”. If a submission to the FDA is planned, full CDISC-SDTM compliance is recommended. The remaining steps to reach full compliance with SDTM are performed using post-processing in SAS. For more details about these steps refer to the chapter entitled “SAS Post-Processing for Full-SDTM”.

SDTM AND ADAM IMPLEMENTATION

The FDA recommends submission of study data according to the CDISC Submission Data Standards (SDS), version 3.1. This guideline includes information relating to the Study Data Tabulation Model (SDTM). In the future, analysis datasets should be structured based on Analysis Dataset Model (ADaM) considerations, i.e. standardization of analysis datasets using the most current ADaM guidelines is recommended.

In this chapter all steps involved after setting up the database are described so as to achieve full SDTM compliance and to implement ADaM.

The following table compares SDTM to ADaM:

Topic	SDTM	ADaM
Version	1.1 (SDS V3.1)	General Considerations: 1.0, other models 0.x
Reviewer	Medical Reviewer	Statistical Reviewer
Requirements	– Standardized structure for • upload into FDA data warehouse • use of standard review tools	– Standardization as much as possible – Analysis friendly: All data for one table in one dataset (“One PROC away”)
Characteristics	– Domain concept, mainly vertical structure – No redundancy – CRF data and trial design data – Textual results	– Analysis oriented – Common group and stratum variables in each dataset – More derived data incl. flags – Numeric codes, if required for analysis

With both models, SDTM and ADaM, all activities related to study analysis benefit of a high level of standardization due to a higher efficiency and a lower risk of errors. Irrespective of the possibilities of standardization, there should still be space for study/project specific additions. Such space is available to a greater degree in ADaM. Even these specific additions should follow some general rules so as to facilitate their integration into the process. Sponsors can take advantage of both standardization and flexibility.

SELECTION OF A STRATEGY FOR SDTM AND ADAM IMPLEMENTATION

Starting from an existing CDMS, there are several possibilities to generate SDTM and AdaM datasets. If the CDMS is fully SDTM compliant (which will probably not be feasible in most companies) only AdaM datasets need to be created. The extent of SDTM compliance in the CDMS implementation may vary across companies. At Accovion the global library covers all SDTM variables and metadata as far as possible.

The most appropriate approach for SDTM and ADaM implementation needs to be selected. Two possible concepts seem to be obvious:

- Independent creation: CDMS→SDTM and CDMS→ADaM
- Consecutive creation: CDMS→SDTM→ADaM

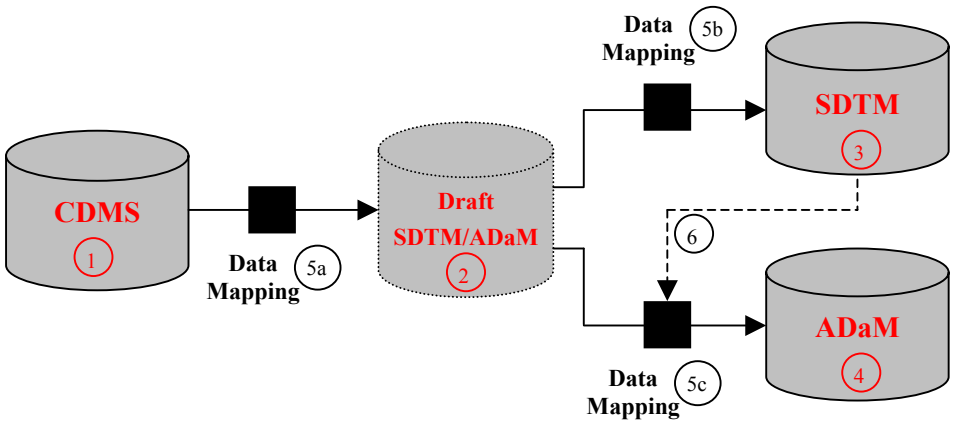
Other concepts may make sense in light of company specific requirements. These models were discussed in terms of the following:

- What is available in the CDMS?
- Will SDTM only be created for submissions?
- Can the workload for SDTM and ADaM be split usefully?
- Will calculations be performed twice, if variables exist in SDTM as well as in ADaM?
- What is the probability of inconsistencies occurring between CDMS, SDTM and ADaM?
- Will the input data for the generation of analysis data sets be available to the authority?
- What is the potential for standardization?

Accovion did not follow one of the concepts described above because of the following disadvantages:

- Independent creation: Derived values need to be calculated twice, if they exist in SDTM as well as in ADaM. This implies a risk of inconsistencies occurring between SDTM and ADAM, especially if the derivation is more complex.
- Consecutive creation: Variables already existing in CDMS, but are not needed in SDTM, have to be derived again for ADaM. This increases the risk of discrepancies occurring between CDMS and ADaM.

Consequently, Accovion selected a different strategy using Draft SDTM/ADaM as the interim step before creation of “Full-SDTM” and ADaM. The strategy is defined as follows, with the advantages and disadvantages outlined below:



1. CDMS is a “Pre-SDTM” database containing the major part of SDTM and variables needed to support ADaM generation.
2. In most cases the variables available in SDTM should also be stored in ADaM. Considering that, variables available in SDTM and ADaM should be only derived once. These derived variables are either empty in CDMS and have to be filled within this step or they have to be added. Draft SDTM/ADaM contains the structure necessary for SDTM (i.e. it has a vertical structure in many datasets) and additional variables only needed in ADaM. This approach is in line with the principles of Good Programming Practice and minimizes the risk of discrepancies occurring between SDTM and ADaM.
   Draft SDTM/ADaM is displayed with a dashed frame, because it does not have to be stored permanently as SDTM and ADaM.
3. To create full SDTM from Draft SDTM/ADaM datasets, in many cases only the deletion of variables not required in SDTM i.e. SDTM comfort variables is required. Sometimes, additional variables may need to be moved to supplemental datasets SUPP~. In exceptional circumstances due to practical reasons, new datasets may need to be created, especially if data is not linked with any CRF. However, in general all information required for ADaM should be available in Draft SDTM/ADaM.

Because of the few steps between Draft SDTM/ADaM and SDTM it seems useful to generate "Full-SDTM" generally.

4. Many steps have to be performed between Draft SDTM/ADaM and ADaM. Most SDTM comfort variables are stored permanently in ADaM. Additional derived variables need to be determined. In some cases additional records may need to be created, e.g. for endpoint records in findings datasets. For some of the datasets it may be useful to transpose data from the vertical to the horizontal so as to support the analysis. To achieve efficient "One PROC away" datasets, information has to be selected from different Draft SDTM/ADaM datasets and needs to be combined in one final ADaM dataset.
   In general, all data required for a specific table need to be in one analysis dataset. Due to differences between SDTM and ADaM in terms of grouping variables, managing the process from Draft SDTM/ADaM to ADaM can be quite challenging. Consequently, the relationships between the existence of ADaM and Draft SDTM/ADaM datasets are much more complex than those between SDTM and Draft SDTM/ADaM.
5. The three data mapping knots symbolize the underlying tools. All these tasks can be implemented in SAS. The potential for standardization is different for each of the three knots. For the paths from CDMS to Draft SDTM/ADaM (5a) and to SDTM (5b) a high degree of standardization can be achieved in the mapping programs so that only few study or project specific additions will be necessary. Most of the steps are only data extractions from the CDMS with very few mapping steps. The path to ADaM (5c) is more study specific, but also leaves a lot of room for standardization, at least on the project level, and preferably between projects.
6. The line from SDTM to ADaM is dashed, because it may be relevant in only very few cases. In general all necessary information for ADaM should be available in Draft SDTM/ADaM. However, there may be exceptions due to practical reasons. If new data sets are created for SDTM and this data is also needed for ADaM generation, then SDTM input is needed for ADaM.

The advantages and disadvantages of this approach are outlined below:

Question/aspect	Answer/estimation
What is available in CDMS?	"Pre-SDTM": High transparency in data flow for the reviewer
SDTM generation for non-submission studies?	Yes, but as SDTM generation is highly standardized by implementation in CDMS and standardization of work steps common to ADaM and SDTM, additional workload is minimized.
Useful split of SDTM and ADaM workload possible?	Workload is not completely independent, but separation into three mapping steps can be performed. Because of the dependencies between SDTM and ADaM, it is not recommended to separate the three steps.
Dual derivations, if variables common to SDTM and ADaM?	No, they are derived only once.
Probability of inconsistencies between CDMS, SDTM and ADaM	The risk between CDMS and SDTM is minimized if standard programs are used. Because there is no dual derivation for SDTM and ADaM no additional risk exists. Standard programs can be used to transpose findings datasets. There is a higher risk of errors for complex relationships between SDTM and ADaM, but this problem does not seem to be less than with other concepts.
Input data for analysis datasets available to the authority?	Even if the CDMS is not submitted, input data partly exists and the close relationship between the CDMS and the SDTM guidelines makes it easy for the reviewer.
Potential of Standardization	High

SAS POST-PROCESSING FOR FULL-SDTM

This involves two mapping steps, from CDMS to Draft SDTM/ADaM and then to SDTM. Starting from the "Pre-SDTM" implementation in the CDMS, necessary tasks for completion of SDTM are listed below as SAS post-processing together with the step at which they should be realized:

Step: CDMS→Draft SDTM	
Task	**Examples**
Adapt variable names (only Oracle Clinical)	Physical Examination (DOMAIN='PE'): TESTCD in Oracle Clinical → PETESTCD in Draft SDTM
Derivation of variables, which are empty in "Pre-SDTM"	Trial Arms dataset (no linkage to any CRF): A part of the variables will have to be filled
Derivation of variables, which do not exist in "Pre-SDTM"	Treatment start RFSTDTC and study/treatment end RFENDTC, baseline flags, ~DRVFL variables (Derivation flags, e.g. in the case of additional baseline records)
Step: Draft SDTM→SDTM	
Task	**Examples**
Delete SDTM comfort variables	EG.EGDT (Date of ECG, numeric), EG.EGDC (Date of ECG, character), DS.DSDECODN (Numeric Code for Controlled Terminology for DSTERM)

Notes:

1. 8 character SAS variable names cannot be stored in Oracle Clinical without major loss of functionality. Variables were set up according to CDISC classes (interventions, findings, …). Upon data extract the appropriate domain prefix is added to achieve CDISC SDTM compliant variable names.
2. The necessity of variable derivations in SAS post-processing may vary across projects and studies. In some cases the derivation is easily possible in the CDMS. In others it may be the better choice practically to derive it in the post-processing step.
3. The earlier described cases, that variables are moved to supplemental datasets SUPP~ or new datasets need to be created in the step from Draft SDTM to SDTM, are not listed, because they are expected to be rare.

ADAM IMPLEMENTATION

Until now the guidelines for ADaM are not as restrictive as those for SDTM. The CDISC Analysis Dataset Modeling Team has published version 1.0 of the General Considerations at the end of 2004. In this document the general content, structure and metadata for analysis datasets are described. Some draft models from the ADaM team give additional guidance including some specific statistical methodology. Among these models the one called “Subject-Level Analysis” (resulting in the ADSL dataset) is of great importance because of the following reasons:

- It can be the main tool for creation of “One PROC away” datasets as it includes variables needed in many analysis, e.g.:
  - Subject identifiers
  - Treatments, populations, strata
  - Important baseline and demographic characteristics
  - Factors that could affect the response

  Subject identifiers e.g. USUBJID and treatment variables will be needed in each project and study. In comparison, the list of other variables will be study or at least project specific.
- There can be several data sources (at Accovion mainly Draft SDTM/AdaM), e.g. EX for actual treatment, DM for demographic variables.
- Less data transformations are necessary.

Within the process, variables of the ADSL dataset can be generated after Draft SDTM/ADaM has been completed.

Some examples of additions in the step from Draft SDTM/ADaM to ADaM are listed below:

- In Interventions datasets: Overall treatment duration (calculated from RFSTDTC and RFENDTC)
- In Events datasets: Treatment emergent flag, study medication at the time of the event in cross-over studies
- In Findings datasets: Change from baseline, endpoint flags, endpoint records

In general the ADSL dataset should be created before other ADaM datasets. This allows the information to be merged to other datasets as ADAE (adverse events) or ADEF (efficacy). The order, in which the other ADaM datasets will be created, depends on existing relationships between ADaM datasets. If for example for each subject, analysis relevant treatment intervals are defined in one dataset, which may be needed in another dataset, then the ADaM dataset including the treatment intervals should be created first. This is related to the strategy that derivations should be performed only once, as this is more efficient and reduces the risk of discrepancies, especially in the case of complex derivation rules.

The input for each ADaM dataset will usually be Draft SDTM/ADaM datasets and pre-existing ADaM datasets, only rarely will full SDTM data also be used. The list of ADaM datasets is supposed to be shorter than the list of full SDTM datasets. For example the separation of supplemental qualifiers from the rest of the associated information does not need to be kept in ADaM and datasets can be combined. Efficacy information existing in several SDTM datasets could be combined in one.

There is still a lot of freedom available in the possible set-up of an ADaM structure according to the guidelines. However, defining an additional company wide standard is useful, at least on the project level the structure should be consistent. Having the same structure in all studies involved in a project can support any meta-analysis relevant to a submission.

TOOLS FOR SDTM AND ADAM GENERATION

SAS is the software, which can perform all steps from the CDMS to SDTM and to ADaM, regardless of the used strategy. With the Accovion strategy, there are two possibilities for implementing mapping steps from the CDMS to SDTM and ADaM:

- If one CDMS dataset leads to exactly one SDTM dataset and one ADaM dataset (without consideration of ADSL), all mapping steps can be performed in one SAS program with Draft SDTM/ADaM as a work data set. The ADSL variables must be added later.
- In more complex relationships between SDTM and ADaM or if the workload should be clearly separated into SDTM and ADaM, all of the Draft SDTM/ADaM datasets should be stored permanently. The steps to reach full SDTM and ADaM then need to be performed in separate programs.

The second possibility works for all studies, while the first possibility is rarely expected to work as a stand-alone. A mixture of the two could be useful.

To ease the process of SDTM and ADaM generation, Accovion has developed a program generator based on Excel sheets, which are used as input files for a macro system. There are two types of Excel sheets, which can be used as mapping tables from a source data structure to a target data structure:

- Dataset mapping table: Domain level metadata
- Variable mapping table: Variable level metadata

Dataset mapping table:

	SDTM/ADaM domain metadata information			
Studyid	Domain	Description	Structure*	Sort order**
1234	ADSL	Subject Level Key Information	SP	USUBJID
1234	ADAE	Adverse Events	E	USUBJID, AESEQ
1234	ADEF	Efficacy	F	USUBJID, EFTESTCD, EFANLTMN

* SP = Special Purpose, E = Events, F = Findings
** USUBJID = Unique Subject Identifier, AESEQ = Sequence Number, EFTESTCD= Efficacy Short Name, EFANLTMN = Analysis Time Window Number;

Main characteristics of the domain level mapping table:

- Tabular overview of datasets
- Generally one line per target dataset should be added.
- The order of datasets in the mapping table corresponds to the dataset generation order.

Variable mapping table:

	SDTM/ADaM variable metadata information						Source information			Mapping information			
Studyid	Domain	Variable	Label	Type	Length	For-mat	SAS library	Dataset name	Variable name	Task*	Transform (SAS code)	Com-ments	Exec. order
1234	ADSL	SEX	Sex	Char	1		SDTM	DM	SEX	no change			
1234	ADSL	TRTAN	ATN**	Num			SDTM	EX	EXTRT	macro	%trta(di,do)***	…	2

* task: no change, assign, decode, num to char, char to num, derive …
** ATN=Actual Treatment Group Number
*** di = input dataset, do = output dataset

Main characteristics of the variable mapping table:

- Tabular overview of source data, target data and the mapping information
- Several source datasets for one target dataset allowed
- In general there is one row per target variable.
  Exceptions:
  - Some calls of macros, which do not belong specifically to one variable, e.g. in the case of a pre-processing to merge datasets
  - Merge variables available in more than one dataset
- For some standardized tasks, e.g. “no change” or “num to char”, no further transformation rule is needed (see variable SEX in the table above).
- Some more complex tasks, e.g. “derive” and “macro”, need further information in the transform column:
  - Derive: The SAS code needs to be added
  - Macro: The macro call has to be added, refer to the variable TRTAN in the table above. The corresponding macro needs to exist.
- The order of execution can be specified, if order is important.
- These variable mapping tables, in an early non-completed stage, can be used as a specification for programmers.

⇒ The program flow can easily be followed, so these tables can serve as a program documentation.

For each study the domain level table has to be filled once, while each output dataset needs its own variable level table. These variable level tables need to be created one per Draft SDTM/ADaM dataset, one per SDTM and one per ADaM dataset. When using standard structures, existing (standard) tables can be re-used, while for project and study specific modifications the tables need to be adapted or new tables need to be created.

AUTOMATION OF DOCUMENTATION

In the chapter “Deliverables to the FDA” it was mentioned, that apart from SDTM and ADaM datasets the FDA expects data documentation in a standardized format, known as define.xml. From define.xml there will be links to datasets and the annotated CRF. This raises the question of how define.xml can be created.

Some thoughts about the possible strategy are outlined below:

- In general, a part of define.xml could be created directly from the datasets. These are the dataset and variable attributes. In addition, associated codelists may be added.
- SAS can be used to create XML documents. Existing XML tools may also be an option.
- Some additional information still needs to be included which cannot be extracted from the datasets, e.g. the role of the variable (e.g. identifier, qualifier or timing) and comments.
- Links to datasets in xpt-format and to annotated CRF have to be included in the define.xml.

Consequently, an efficient method of creating define.xml should be found. A tool for the generation of SDTM and ADaM datasets, which is based on the use of mapping tables developed at Accovion, has previously been described. These mapping specification tables can serve as specifications for the programmers, as input for the macro system and as documentation for a reviewer. In addition, the structure of these mapping tables is similar to the structure of domain and variable level metadata tables of define.xml. Therefore the contents of the mapping tables can be re-used for define.xml.

To demonstrate the process involved in transforming mapping tables to domain and variable level metadata, the tables previously described will be taken as the source:

Domain level metadata:

ADaM Datasets for study 1234						
Dataset	Description	Structure	Purpose	Keys	Location	Documentation*
ADSL	Subject Level Key Information	1 record per subject	Analysis	USUBJID	../1234/dds/adsl.xpt	SAP and/or adsl.sas
ADAE	Adverse Events	1 record per subject per event	Analysis	USUBJID, AESEQ	../1234/dds/adae.xpt	SAP and/or adae.sas
ADEF	Efficacy	1 record per subject per parameter per analysis time window	Analysis	USUBJID, EFTESTCD, EFANLTMN	../1234/dds/adef.xpt	SAP and/or adef.sas

*column NOT in define.xml standard structure, but can be added sponsor specific to link with Statistical Analysis Plan (SAP) and/or programs

Comparing the dataset mapping table previously described and the domain level table above shows that the columns “Dataset”, “Description” and “Keys” can be copied, and other fields easily filled with potential for some automation.

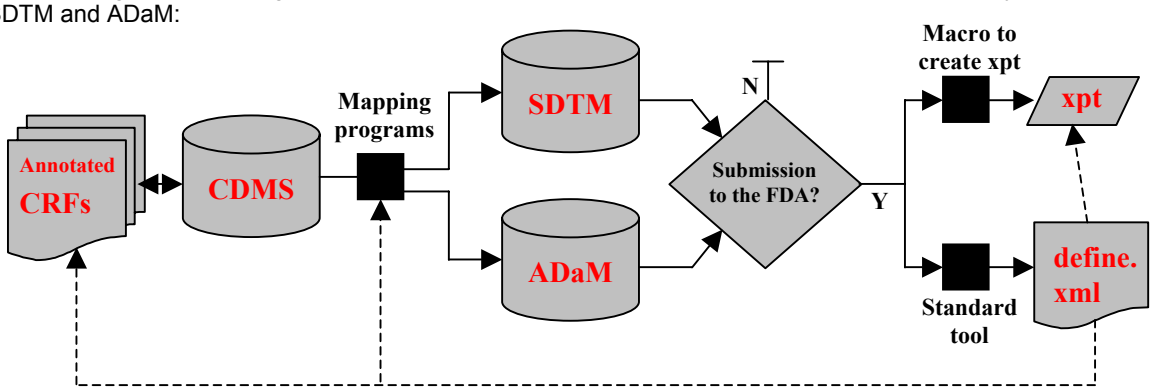
Variable level metadata:

Variable Metadata for Dataset ADSL – 1 record per subject						
Variable	Label	Type	Controlled Terms or Format	Origin	Role	Comment
SEX	Sex	Char		DM.SEX	Result qualifier	No change
TRTAN	Actual Treatment Group Number	Num	0 = ‘Placebo’ 1 = ‘Drug A’ 2 = ‘Drug B’ 3 = ‘Drugs A+C’ 4 = ‘Drugs B+C’	Derived from EX.EXTRT	Selection, result qualifier	For one subject either one or two treatments in EXTRT are possible

Comparing the variable mapping table to the variable level metadata shows that the columns “Variable”, “Label”, “Type” can be copied, and the other fields can be filled with potential for automation. If for example the dataset name is DM, the variable is SEX, and the task is ’no change’, then “Origin” can be filled with “DM.SEX”.

To create define.xml a standard tool can be developed, which collects all the necessary information from the mapping tables and other files. A check for consistency between the datasets and the contents of define.xml can also be integrated in the define.xml generation tool. This procedure can be used in all cases where SDTM and AdaM need to be implemented.

The following represents a graphical representation of the main aspects of the procedure used by Accovion to create SDTM and ADaM:



SDTM and ADaM will be created for all studies. Draft SDTM/ADaM is not shown. The tools used to create xpt files and define.xml only need to be used for submissions to the FDA. Dashed lines symbolize the links.

INTEGRATION OF DIFFERENT DATA STRUCTURES

The previous chapter was based mainly on a full service scenario, ranging from the CDMS to the SDTM and ADaM datasets, and including information required by the FDA. The scenario will now be considered of data structures being available, which have to be transferred into an FDA compliant structure. The same strategy can be used, however the possible interaction with the Accovion model needs to be discussed. Different data structures may be available from legacy studies at Accovion or from sponsors.

Depending on the available source data and the required target data structures an optimal strategy will be selected.

PURPOSE OF MIGRATION (TARGET DATA)

There are two possibilities why a migration may be necessary:

- Migration for meta-analysis (ADaM only)
- Migration of legacy data for submissions (SDTM and ADaM)

SOURCE DATA

There are three possible source data structures:

- Study specific CDMS structure
- SDTM
- Analysis datasets (ADS) in study specific structure

Some considerations with regard to the appropriate choice of source data in case more than CDMS data exists:

- If SDTM is available, generation of ADaM should be based on available SDTMs.
- If ADS is available, it should be used, if:
  - The ADS data structure is similar to the target ADaM structure
  - Complex derivations in ADS (e.g. per-protocol flags) should not be recalculated for SDTM/ADaM to avoid the risk of discrepancies.
  - ADS contains more than half of the target SDTM/ADaM
- If SDTM and/or ADS do not contain all the information needed for the target datasets, the required information should be extracted from the CDMS if possible.
- Backward mapping from ADS to SDTM is probably not as useful as forward mapping from CDMS

These issues provide some guidance for the selection of source data. The resulting scenarios are as follows:

Target Data Structure / **Source Data Structure**	**ADaM only**	**SDTM + ADaM**
CDMS	1. CDMS ⇒ AdaM 2. Accovion model: CDMS ⇒ “Pre-SDTM” ⇒ (SDTM) + ADaM	1. Independent model 2. Consecutive model 3. Accovion model: CDMS ⇒ “Pre-SDTM” ⇒ SDTM + ADaM
CDMS + SDTM	[CDMS] + SDTM ⇒ ADaM	[CDMS] + SDTM ⇒ ADaM
CDMS + analysis datasets (ADS)	[CDMS] + ADS ⇒ ADaM	1. ADS ⇒ SDTM + ADaM 2. CDMS ⇒ SDTM, ADS ⇒ ADaM
CDMS + SDTM + analysis datasets (ADS)	[CDMS] + SDTM + ADS ⇒ ADaM	[CDMS] + SDTM + ADS ⇒ ADaM

Note: [CDMS] means, the usage of CDMS is optional; (SDTM) means, it is not needed, but automatically created.

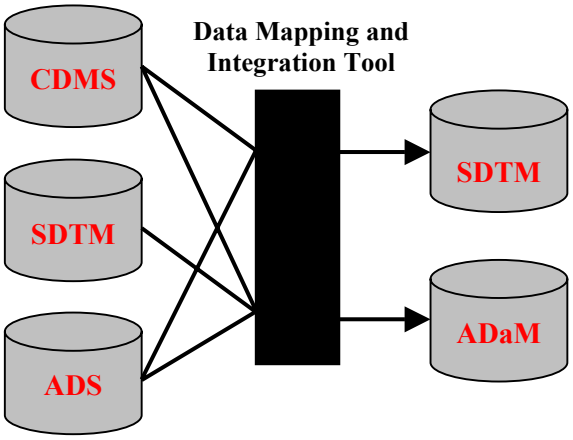
At Accovion the same macro system used for mapping from source to target datasets can also be used for the data migration. The position of the migration in the overall process is dependent on the strategy.

SELECTION OF A DATA MIGRATION STRATEGY

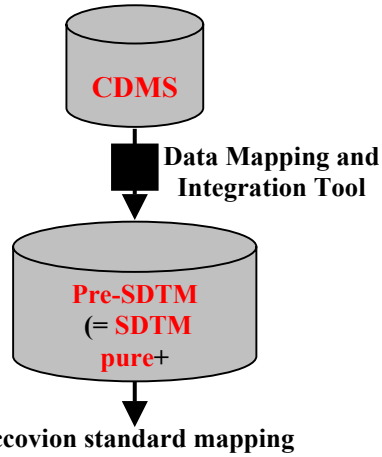
The following aspects should be considered when selecting an appropriate strategy:

- Duration of programming
- Risk of errors
- Risk of inconsistencies
- Validation effort
- Reusability of programs/potential of standardization

There are two possible scenarios, which are outlined below. Direct migration and migration to “Pre-SDTM”.



Direct Migration



Migration to “Pre-SDTM“

Migration to “Pre-SDTM” contains a much higher potential for standardization than direct migration. From “Pre-SDTM” to SDTM and ADaM a lot can be performed using standard mapping. This has advantages in terms of the points outlined above, with the exception of the one related to the risk of inconsistencies.

Conclusion:

1. If only CDMS exists: Mapping to “Pre-SDTM” is recommended.
2. If CDMS exists and the decision was made to map to “Pre-SDTM” disregarding existing SDTM and/or ADS datasets, two more points need to be taken into account:
   - Increased risk of discrepancies between pre-existing datasets and new SDTM/ADaM datasets
   - Increased validation effort to compare old and new SDTM, ADS and ADaM.

   Neither strategy is better than the other one.
3. If CDMS, SDTM and/or ADS exist and are used, it does not make sense to go back to “Pre-SDTM”.

For details on tools for integration, see the chapter entitled “Tools for SDTM and ADaM Implementation”.

ANALYSIS OUTPUT

The programming of analysis outputs should be performed based on analysis datasets. In addition to the guidance from the ADaM model, study specific specifications, e.g. SAP and planned outputs (i.e. tables, listings and graphs) need to be considered for the design of analysis datasets.

RELATIONSHIP OF ADAM DATASETS AND ANALYSIS OUTPUT

ADaM datasets should be set up based on the following prerequisites:

- Mock tables, listings and graphs defined based on the SAP
- ADaM guidance
- The principle of analysis-ready datasets: Each statistic in the table can be replicated by running a standard statistical procedure using the analysis dataset as input, e.g. PROC FREQ, PROC MIXED etc. with little or no data preparations (e.g. WHERE statements). An appropriate setup of an ADSL dataset supports this principle.
- Derive variables only once: This can be realized by merging ADSL with other ADaM datasets and reusing already derived ADaM variables in other ADaM datasets.

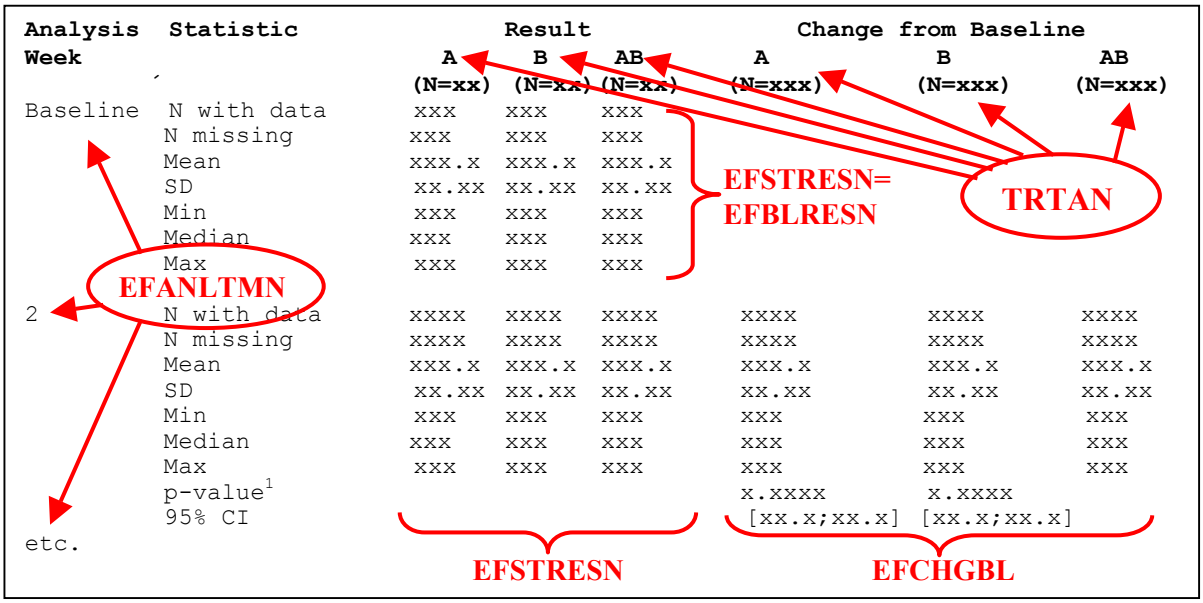
The use of ADaM dataset standards can support the potential for standard analysis programs. There is a benefit to using a standardized process in terms of the following advantages:

- Minimized programming effort
- Reduced risk of programming errors
- Less validation effort because of the use of validated standard macros
- Reuse of programs

Examples of using analysis datasets in tables:

1. Descriptive statistics for the change in the primary efficacy variable from baseline to endpoint (use selection WHERE EFTESTCD='PEF' for primary efficacy parameter and ITT='Y' for ITT subjects in dataset ADEF):

Table 1.1 Change from Baseline in Primary Efficacy Variable – ITT population

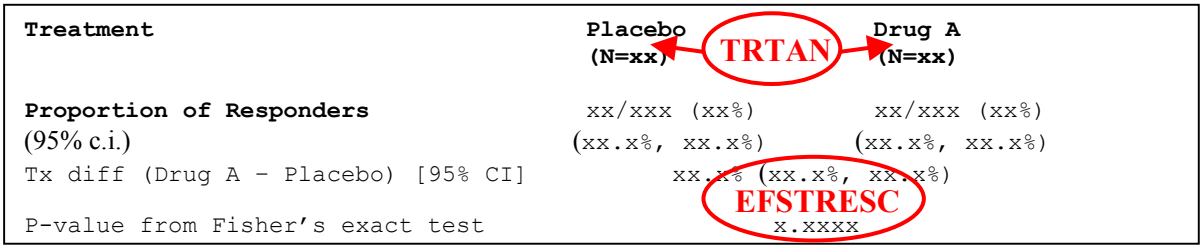


Note: [1] p-values from pairwise comparison AB versus A resp. B from t-test;
ITT = intention-to-treat; treatment groups as treated.

The variable names could be defined differently, according to the sponsor's needs. Instead of the underlying vertical dataset structure as in SDTM using EFTESTCD, there could also be a horizontal structure defined with one record per subject and analysis time window (EFANLTMN). Then EFBLRESN, EFSTRESN and EFCHGBL could be replaced with PEF_BASE, PEF_VAL and PEF_CHG for baseline values, current values, and the change from baseline in primary efficacy parameter.

2. Treatment success considering different study specific aspects (using the selection WHERE EFTESTCD='RSP' for the responder, WHERE EFANLTMN=999 for the endpoint selection and PP='Y' for the per-protocol subjects in the dataset ADEF):

Table 1.2 Proportion of Responders at Endpoint – Per-Protocol population



Note: Treatment groups as treated. Responder = …

The variable names could again be defined differently, and according to the sponsor's needs. Instead of the underlying vertical dataset structure as in SDTM using EFTESTCD, there could also be a horizontal structure defined with one record per subject. Then EFSTRESC = "Y"/"N" would be replaced with RESP = "Y"/"N".

STANDARD ANALYSIS TOOLS

At Accovion, in addition to SDTM and ADaM metadata standards, the standardization efforts are supported by:

- A set of 20 defined core tables (standard table layouts), e.g. for MedDRA coded terms by system organ class
- A set of standard output macros, e.g. macros for categorical analysis and inclusion of titles and footnotes

The majority of study tables can be grouped according to the set of core tables. This enables Accovion to build a library containing standard macros and sample programs. The macros are flexible enough to meet study specific needs, e.g. inclusion of p-values, presentation of percentages (N/Y?), etc.

DOCUMENTATION OF ANALYSIS PROCESSES

In the chapter entitled “Deliverables to the FDA” it was stated that in addition to outputs in reports and summaries, the FDA expects analysis documentation. The delivery of analysis programs is also recommended. There is currently no CDISC pre-defined structure for the analysis metadata.

Based on the recommendations in the ADaM General Considerations, the documentation could be structured as follows, using the two examples from the chapter entitled “Relationship of ADaM datasets and Analysis Output”:

Analysis-Level Metadata for study 1234				
Analysis Name	Description	Reason	Dataset	Documentation
Table 1.1: ef00001t.lst	Change from Baseline in Primary Efficacy Variable – ITT population	Prespecified in Protocol	../1234/dds/adef.xpt	SAP, Section X.Y and /or ef00001t.sas
Table 1.2: ef00002t.lst	Proportion of Responders at Endpoint – Per-Protocol population	Prespecified in Protocol	../1234/dds/adef.xpt	SAP, Section X.Y and /or ef00002t.sas

The document contains one entry per output with file identifier, title, and reason for creation listed. Furthermore the table provides links to ADaM datasets, related information in the SAP and the programs.

Additionally, information can be added, e.g. selection criteria for the ADaM dataset, used key variables, supporting SAS macros and key program codes. Other supplemental documentation can also be added. The content proposed in the above table is useful and in accordance with the current ADaM guidance. Sponsor specific changes can be taken into consideration with regard to reviewer’s preferences.

CONCLUSION

CDISC standards have a large impact on the clinical development process. They define the data structures in the clinical database management system, case report tabulations and analysis datasets. These standards affect several operative functions where clinical data has to be handled, i.e. mainly database and statistical programmers. Other functions use the available data structures for their daily work, i.e. CRF designers and data coordinators only work with CDMS, whereas statisticians work with analysis datasets. FDA reviewers use SDTM and ADaM. The approach to implementation depends to some extent on the company. However, some guidance has been given, and basic rules which should be followed are listed.

The available CDISC guidelines and the high potential of standardization increase the efficiency of programming and review. This allows the generation of standard datasets, standard analysis macros and standard review tools. Each company can benefit in the long-term from these possibilities, even if data is not submitted to the FDA. SDTM guidelines are already relatively stable, while ADaM guidelines are still under development. Both are expected to provide the best approach for the future.

At Accovion a modular approach is defined which is compliant with CDISC expectations. This process has been presented in detail in this paper. Data structures at several stages within a clinical trial have been described. These include setting up a clinical database management system, generating SDTM and ADaM datasets and creating study report outputs based on a high level of standardization. Furthermore the handling of studies with “Non-CDISC” structures has been outlined.

In addition to the advantages of possible standardization, Accovion’s modular approach provides flexibility to deal with study and project specific rules and future changes in SDTM and ADaM models. Sponsor specific needs can also be integrated into the process as required.

REFERENCES

CDISC quick link to models: http://www.cdisc.org/standards/index.html

SDTM V1.1: http://www.cdisc.org/models/sdtm/v1.1/index.html

- Case Report Tabulation Data Definition Specification (CRT-DDS, also called define.xml): http://www.cdisc.org/models/def/v1.0/index.html

ADAM: http://www.cdisc.org/models/adam/V1.0/index.html

- General Considerations V1.0: http://www.cdisc.org/models/adam/V1.0/ADaMGeneralConsiderationsV1.0.pdf
- Subject Level Analysis V0.2: http://www.cdisc.org/models/adam/V1.0/ADaMSubject-LevelV0.21.pdf
- Change From Baseline V0.2: http://www.cdisc.org/models/adam/V1.0/ADaMChangefromBaselineV0.2.pdf
- Categorical Data Analysis V0.2: http://www.cdisc.org/models/adam/V1.0/CategoricalDataV0.2.pdf

FDA Guidance: http://www.fda.gov/cder/regulatory/ersr/default.htm

- Study Data Specifications V1.1: http://www.fda.gov/cder/regulatory/ersr/Studydata-v1.1.pdf
- xpt Generation: http://www.sas.com/fda-esub

Susan Kenny and Michael Litzinger: “Strategies for Implementing SDTM and ADaM standards”, Paper FC03 at PharmaSUG 2005, http://www.pharmasug.org/2005/FC03.pdf

ACKNOWLEDGMENTS

The authors would like to thank all colleagues, who spent their time giving valuable assistance and input in the form of discussions, advice and review of the document.

CONTACT INFORMATION

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

Edelbert Arnold
Accovion GmbH
Helfmann-Park 2
D-65760 Eschborn (Frankfurt)
Germany
Phone (49) 6196-7709-321
edelbert.arnold@accovion.com