PhUSE 2011

Paper CD07

SAS Drug Development as a platform governing data standards specification and clinical data exchange between pharmaceutical organizations

Mark Lambrecht, SAS, Tervuren, Belgium
Peter Van Reusel, Business & Decision Life Sciences, Brussels, Belgium

ABSTRACT

An innovative set-up is described here that consists of upfront clinical study data specifications in SAS Drug Development® (SDD) using CDISC standard data models. This set-up enables sponsors to automate the exchange of specifications and instructions to their internal and external partners. A central data standards library and a metadata repository, centrally accessible in SDD, are not only used to generate study specifications, but also to enable the verification of the study data and metadata against the specifications. The processes that govern exchange of study specifications and the subsequent execution of validation and verification SAS programs are enabled by automated and fully secured workflows available in SAS Drug Development.

INTRODUCTION TO CLINICAL DATA EXCHANGE

A major challenge in setting up clinical trials and the ongoing data collection is the management of clinical data specifications in use for any given trial. The Clinical Data Interchange Standards Consortium (CDISC) has developed open industry standards that have been a major step forward to harmonize the different clinical submission and analysis data formats. According to a well-published business case for the CDISC standards (Rozwell, 2009), half of the possible savings by introducing clinical data standards can be generated when imposing them at study start-up as opposed to harmonizing data at the end of the trial period and right before submission to authorities. A major stumbling block in realizing this benefit is the development of an operational mechanism to manage and define the exact data standard elements in use for any given trial (the so-called clinical metadata) in a controlled manner. This paper describes a solution to this problem by the creation of a Data Standards Library in SAS Drug Development that includes all data elements and specifications for all trials, a Study Metadata Repository, and the development of standard processes that allow to exchange and control the quality of the exchanged metadata. The solution enables the different stakeholders during a clinical trial to produce standards-adherent clinical data of the highest quality. This approach guarantees diverse important process characteristics such as patient data blinding, maintaining an intact trail between the upfront defined study metadata and the clinical data gathered during trial execution, and providing high-quality standards-adherent data to statistical programmers, statisticians and medical staff. The stakeholders involved in this process are the pharmaceutical sponsor and their data management groups, one or more Contract Research Organizations (CRO's), and possible other implicated partners (laboratory data providers, PK data providers). SAS Drug Development acts as the platform governing and enabling all the required process steps, and ensuring that proper controls are put in place at each step of the sequence.

SAS DRUG DEVELOPMENT AS THE ENABLING PLATFORM

SAS Drug Development is a leading clinical software solution that provides a secure, global access to a centralized clinical information repository for all authorized clinical development team members. SAS Drug Development has three major features that make it the ideal vehicle for managing the exchange of clinical data standards between the different stakeholders;

- The platform has a secure repository to store and generate clinical information in a regulatory compliant manner
- The data and metadata stored in the repository is easily but securely accessed in a web browser interface by an authenticated user role in a format that is ready to be processed by the SAS® execution engine. SAS Drug Development is tightly integrated with key SAS execution engine modules such as advanced

analytics procedures, a CDISC-standard processing engine (SAS Clinical Standards Toolkit) and the reporting engines in Base/SAS

- SAS Drug Development provides flexible yet rigidly controlled process workflows that enable the different described stakeholders to control the data and metadata at appropriate moments during the process. These process workflows can be consuming the stored data and libraries, and can be defined to automatically trigger SAS programs as a result of specific actions, such as the upload of new clinical data by a CRO into the system.

We have developed a solution in SAS Drug Development that centers around 5 key functional areas for clinical data standards exchange (see **Figure 1**): a Data Standards Library, Study Specification, Study Metadata Repository, a published Comparison/Validation process workflow and an Issue Tracking Mechanism. We will now describe and detail each of the 5 functional areas.

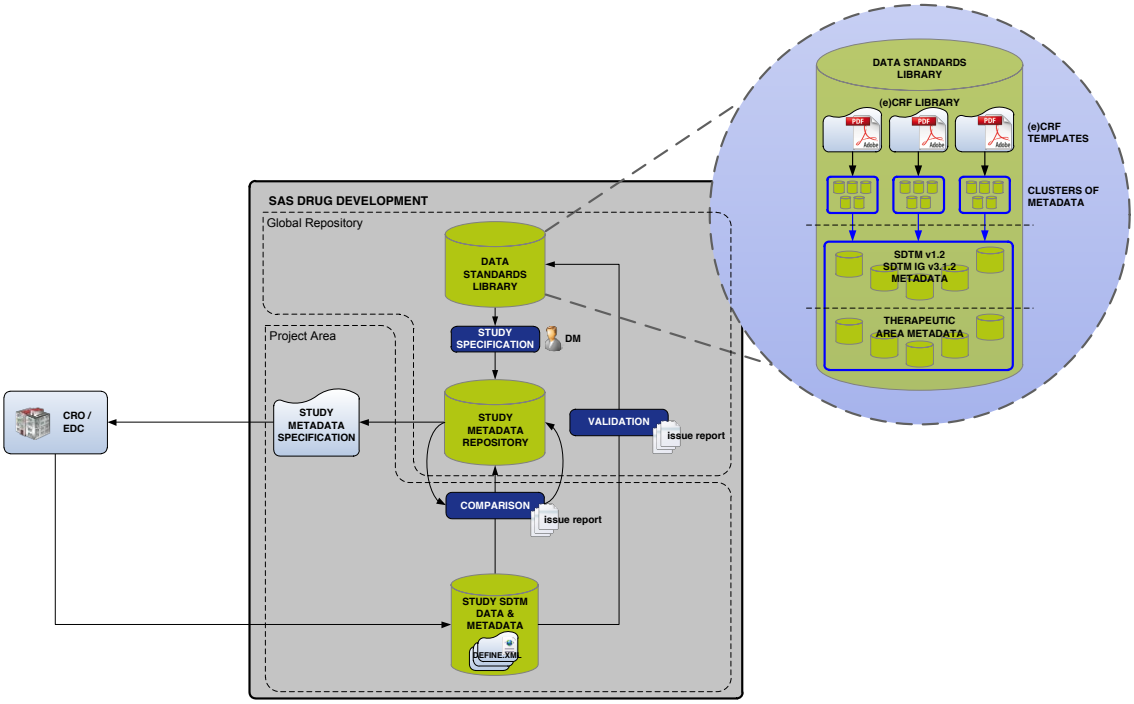

Figure 1 : Conceptual representation of the standards exchange platform enabled in SAS Drug Development

DATA STANDARDS LIBRARY

In SAS Drug Development, the Data Standards Metadata Library serves as a major building block of an integrated data (submission) solution and is stored in an area of the repository that is globally accessible (see **Figure 2)**. The Data Standards Metadata Library is envisaged as a CDASH compliant (e)CRF Library concept linked to the corresponding five-level SDTM metadata clusters (domain metadata, variable metadata, value level metadata, controlled terminology, computational algorithms). This Data Standards Library consists of information objects in SAS Drug Development repository and therefore it is tightly secured, under change management control and only modifiable by a data standards librarian. This set-up provides a 'pick and select' solution for building submission-ready study specifications as from study start.

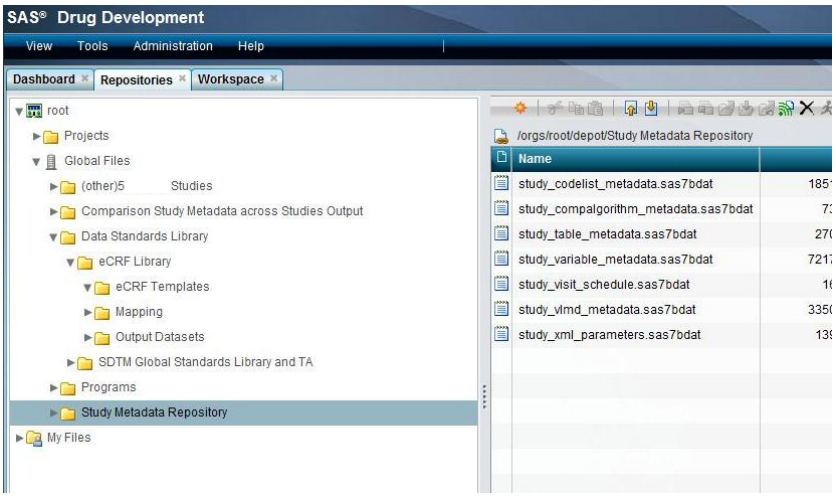

Figure 2 : SAS Drug Development stores the Data Standards Library and Study Metadata Repository in a globally accessible SDD area

The Data Standards Library consists of 3 distinct modules (**Figure 3**): a first module harbors the standard eCRF templates (eCRF Library). The (e)CRF library contains all standard (e)CRF modules having both a CDASH and a SDTM annotation. The (e)CRF modules focus on standardized content with respect to variables and code lists used and pre-printed text.

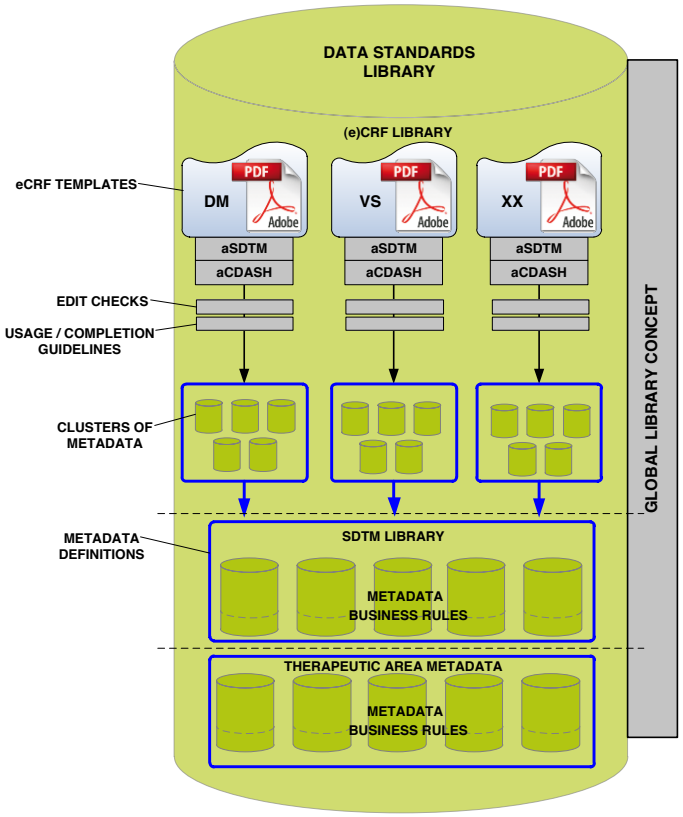

Figure 3 : The Data Standards Library consists of the eCRF Library, the SDTM Library and the Therapeutic Area Metadata.

The second module contains the SDTM Metadata Definitions (SDTM Library). The metadata library component is developed based on the CDISC SDTMIG V3.1.2 definitions, the (e)CRF template specific metadata extracts and

sponsor-defined business rules. The five levels of SDTM compliant metadata include metadata at domain level, variable metadata, code lists/terminology, value level metadata and computational algorithms.
A third module has the therapeutic area specific elements (Therapeutic Area Library).
Clearly, a set of business rules is needed to make sure that this metadata is defined and changed according to predefined guidelines. These rules are enforced by execution of specific SAS programs and user interfaces in SAS Drug Development that allow a data librarian to investigate, check and cross check the metadata against any new additions before using or uploading anything to the Data Standards Library.

STUDY SPECIFICATION AND STUDY METADATA REPOSITORY

In the Data Standards Library, the metadata clusters corresponding to the (e)CRF templates contain all standard domain structures, variables, code lists, value level metadata and computational algorithms.

Each unique (e)CRF template links to a cluster of relevant metadata in order to facilitate building the study specifications.
To build the specifications of a study the user needs to provide the necessary information to allow the selected metadata of the study to be copied from the Data Standards Library into the Study Metadata Repository by a 'pick and select' module implemented in the central clustered metadata model.

The user can then complete the visit schedule. This visit schedule lists the (e)CRF templates selected for building the study specifications of a study. The (e)CRF templates are assigned to visits and time points. The clusters of metadata attached to the selected (e)CRF templates are copied to the Study Metadata Repository.

Beside this visit schedule, study parameters are provided for each study. The study parameters are part of the define.xml that is extracted from the Study Metadata Repository.

The study specifications also require the trial design data. This includes the domains:
- Trial Arms (TA),
- Trial Elements (TE),
- Trial Inclusion/Exclusion Criteria (TI),
- Trial Summary (TS),
- Trial Visits (TV).

With this input and after running the study specifications build program in SAS Drug Development the Study Metadata Repository is populated (see **Figure 4**) with:
- Metadata selected by the study specifications build program,
- Visit schedule and selected (e)CRF templates,
- Study Parameters needed for the creation of the define.xml,
- Trial Design information.

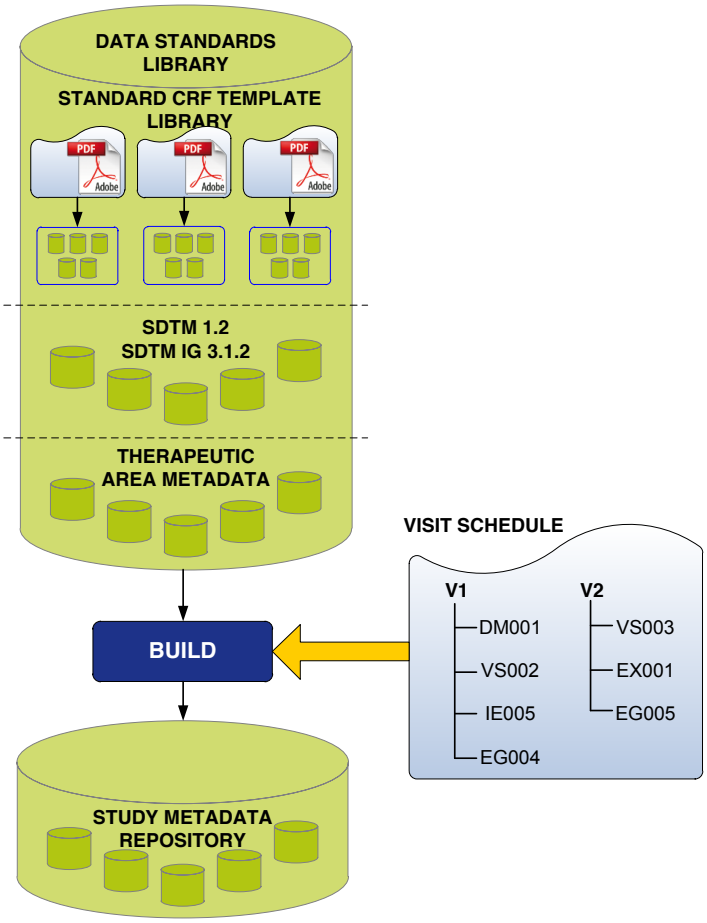

Figure 4 : The Study Metadata Specification build adds the study metadata to the Study Metadata Repository

In addition to a globally accessible area where the Data Standards Library is stored, SAS Drug Development has the ability to set up specific project or study areas and it is here that the study metadata, the associated study reports, and later on during the process, the study data is stored. As a result of this project-centric organization of the data, project- or trial-specific user groups can be defined in the system that has access to one or more trials as required.

A Study Specifications Build Program allows the user to provide the study specifications for a study. Study Specifications required to generate study metadata are Visit Schedule (combined with (e)CRF Templates), Study Parameters and Trial Design Domains. The Visit Schedule can be modified according to structure of the study.

The Study Metadata Repository consists of:

- Metadata selected by the study specifications build program,
- Visit schedule and selected (e)CRF templates overview,
- Study Parameters needed for the creation of the define.xml,
- Trial Design Information.

A Study Metadata Repository Maintenance Program aids the user in monitoring and editing the Study Metadata Repository.

Reports, document and other informative deliverables can be extracted from the Study Metadata Repository and sent to the external partner. This information enables the external partner to build or refine the study database. The reports that can be created out of the Study Metadata Repository are:

- Define.xml,
- Define.pdf,
- Print (e)CRF template,
- Visit Schedule,
- Data Governance Study Report,
- Study Metadata Comparison Report (across Studies and across Metadata versions),
- Metadata Listings,
- Trial design datasets.

To illustrate the relationship of the different libraries, and more detailed, how study specifications are generated in SAS Drug Development, we have developed a number of business process workflows that guide the different user roles alongside the described process. This is the topic of the next paragraph.

DEVELOPMENT OF SPECIFICATION AND VERIFICATION PROCESS WORKFLOWS IN SAS DRUG DEVELOPMENT

SPECIFICATION OF STUDY METADATA

A user belonging to the group that manages the data standards, often a data manager of the pharmaceutical company or an external CRO, provides the visit schedule, the study parameters and the trial design domains (if appropriate). When done, the user can close a task in the workflow that automatically activates a SAS program that generates the study metadata. In addition, a series of validation programs are triggered: programs that compare the generated study metadata with the Data Standards Library and a set of programs to compare the generated study metadata with those of other studies of the same therapeutic area or project (checking for metadata consistency). A sample report containing such a comparison across studies is seen in **Figure 5**.

Comparison Report against the Global Library / Level : TABLE / Column : STRUCTURE

Check	Dataset	Structure	ABC_0001	Global Standards
(all)	(all)	(all)	(all)	(all)
OK	AE	One record per adverse event per subject	X	X
OK	DM	One record per subject	X	X
OK	EX	One record per constant dosing interval per subject	X	X
OK	PC	One record per analyte per plannend time point number per time point reference per visit per subject	X	X
OK	SUPPAE	One record per IDVAR, IDVARVAL, and QNAM value per subject	X	X
Difference	VS	One record per vital sign measurement per visit per subject per position	X	

Figure 5 : Sample report comparing a study with the Data Standards Library.

EXCHANGE WITH CRO AND UPLOAD OF CLINICAL DATA INTO SAS DRUG DEVELOPMENT

When the data manager is satisfied with the generated study metadata, it can be released to the CRO or to the group that is responsible for operationally capturing the clinical data. When the data or trial manager releases his task, the CRO organization receives an automated e-mail message that alerts them for the arrival of new study metadata.

Later on, during clinical trial execution, the CRO can iteratively upload the clinical data, adherent to the data standards dictated by the study metadata, and this is controlled by a second workflow. The objective of the second workflow is to put in place proper SDTM (CDISC) validation checks and controlling whether the provided study data adheres to the pre-defined study metadata.

SDTM VALIDATION CHECKS

To ensure that the data provided by the external partner is consistent with the specifications sent by the pharmaceutical sponsor, a comparison with the Study Metadata Repository is performed. For this comparison all 5 levels of metadata and the Trial design Data will be compared. The output of this comparison will be provided in reports and stored in a database for metrics reports. The comparison of the study received from the external partner will be compared against the Study Metadata Repository (see **Figure 6**):

- Comparison of the study metadata with the Study Metadata Repository,
- Comparison of Trial Design Datasets received from the external partner and the Trial Design Datasets provided by the data manager of the trial sponsor.

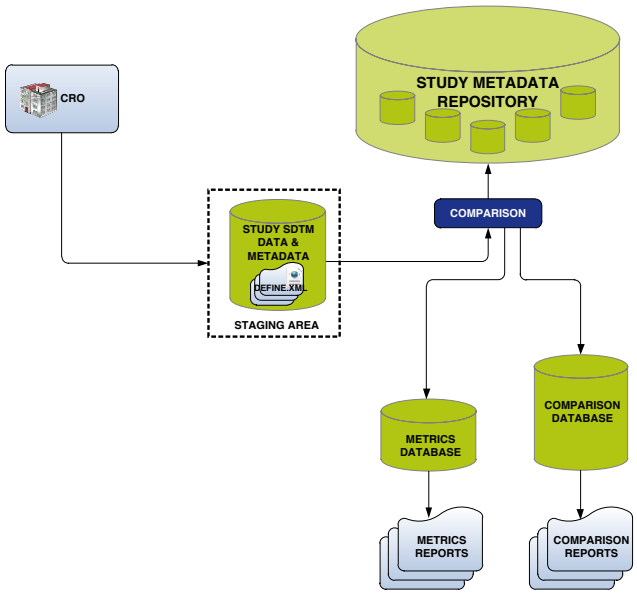

Figure 6 : Comparison of uploaded study data and metadata with Study Metadata Repository

The external partner can then upload the datasets and a define.xml in the staging area. The staging area is the location in SAS Drug Development where the sponsor data manager user can decide whether they want to further process the study and copy the study in the Production area or whether they want to return the study to the external partner to fix the possible issues.

Issues can be caused by inconsistency of the received study from the external partner with the specifications sent by the data manager. These inconsistencies are captured with the comparison program. This program compares the metadata of the received study with the predefined metadata in the Study Metadata Repository. The program will also compare the Trial Design datasets received from the external partner with the Trial Design datasets defined by the data manager at study build.

Validation is performed by checking for consistency with the Data Standards Library. This consistency is important for the submission of the study to the FDA. Studies that do not pass through the validation process will be returned to the external partner for further processing. Reports and listings are provided through the system. A SAS database is kept that allows to produce metrics reports. The validation program will run a series of checks against study metadata and create output reports. The checks are in the form of SAS programs and can be divided into categories for ease of selection. The job selection can be implemented via the user interface. The following categorizations are implemented:

- **General Classification**
  - Data Checks,
  - Metadata checks.

- **Classification based on SDTM class of CDISC Domains**
  - Events,
  - Interventions,
  - Findings,
  - Findings About,

- Trial Design,
- Special Purpose,
- All Observation Classes.

- **Classification based on CDISC Model Version and Therapeutic Area Model**
- SDTMIG V3.1.1,
- SDTMIG V3.1.2,
- ADaMIG V1.0,
- Therapeutic Area.

The output of the Structural Validation and the Data Validation checks are added to a SAS table which will provide an error report in Excel or PDF format.

The execution of the above described validation programs are governed by the second workflow in SAS Drug Development. When the SDTM validation program and the comparison of the incoming study metadata and trial design datasets with the Study Metadata Repository have run, the data manager of the trial sponsor organization is alerted by an e-mail or by the automatic messaging system in SAS Drug Development. This message contains a hyperlink to a summary report securely stored in SAS Drug Development. It is then up to the data manager to accept the data, whereby it gets copied to an internal staging area visible only to internal users, or reject the data, whereby the CRO will automatically be notified of the rejection.

CONCLUSION

We have proven it is possible to set up the data standards specification and dissemination and subsequent validation in SAS Drug Development with the objective to produce high-quality clinical data. CDISC standards are powerful logical data models that allow life sciences organizations to collaborate across different parts of the clinical data management process. Not only can these standards be defined and managed in a central repository, but more importantly the time needed by a data manager to harmonize data and ensure high clinical data quality during and after a trial ends, is significantly reduced. Incoming data can be explored and validated every time it is uploaded by the external partner. SAS Drug Development automates and supports the business process by providing workflow mechanisms and controlled access to the global and project-specific data and metadata assets.

The described system and process brings operational advantages to the data management group of a pharmaceutical sponsor by working more efficiently with external partners on the basis of the agreed specifications, and it also allows biostatistics and medical investigators to get access to clinical standards data faster. This should lead not just to better collaboration between CRO's and pharma companies, but also between pharmaceutical companies when co-developing therapies and during submission to authorities in an approval process.

REFERENCES

Rozwell *et al.*, 2009, online at http://www.cdisc.org/business-case. A business case for standards by CDISC and Gartner.

ACKNOWLEDGMENTS

The authors thank Tina Apers (B&D Life Sciences) and Matt Gross (SAS) for reviewing this paper and the involved B&D and SAS teams for their work during the set-up of the described system.

CONTACT INFORMATION

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

Mark Lambrecht
SAS Institute
Hertenbergstraat 6
B-3080 Tervuren
Belgium
Office Phone: +32 2 766 07 00
Web: http://www.sas.com/

Peter Van Reusel
Business & Decision Life Sciences
Sint Lambertusstraat 141 rue Saint Lambert
B-1200 Brussels
Belgium
Office Phone: +32 2 510 05 40
Web: http://www.businessdecision.be