SANOFI Clinical Data Integration Solution as Part of SDTM Generation

Hasnaa LIKAOUI, KEYRUS, Levallois-Perret, FRANCE
Melissa MARTINEZ, SAS Institute, Inc., Austin, Texas USA
Nadine MOURIN, SANOFI, Chilly Mazarin, FRANCE

ABSTRACT

SAS[®] Clinical Data Integration tool has been selected by SANOFI in order to prepare clinical data for submission using SDTM standards and adapt the CDISC linear data model to SANOFI needs. A standard structure will be created from Clinical Data Management System (CDMS) metadata and will include all applicable SDTM mapping and global derivations. The challenges of this project are industrializing these activities based on the management of multiple SDTM versions, creating a standard job template library, managing customized domains, and implementing customized variables. This paper fully describes how SANOFI has integrated, customized and used SAS[®] CDI tool to support the linear data model and how the data integration team will ensure the production of validated output SAS datasets.

INTRODUCTION

In order to enhance the SANOFI planned core model supporting the integration of all captured and derived clinical data elements through the drug development and data lifecycles, a standardization of captured clinical data into an industry-standard format to support user analysis through a linear data model flow is necessary. The goal of the Clinical Data Integration project is to govern a part of SDTM transformations using SAS[®] Clinical Data Integration (SAS[®] CDI).

BUSINESS CONTEXT

In the context of the UNIFY (Re-inventing a new common way of operating clinical development across all Sanofi R&D entities, through business process management, innovation and technology) and MAESTRO (Providing a business and technical framework to facilitate the integration of clinical data from diverse sources to allow easy access to customers) programs, new capabilities are required to support some business needs for Sanofi, Sanofi Pasteur and Genzyme entities. One of those capabilities is related to the enhancement of the clinical data flow from source systems to Statistical Computing Environments (SCE) and managed within SCALA (SCE Application for Linear dAta model) project.

Current processes for extracting data from Clinical Data Management Systems (CDMS) and importing this data into SCE are complex. Multiple tools are used and the processes are typically performed on-demand. The whole data flow needs to be refined to feed UNIFY SCE and provide Biostatisticians & Programmers (B&P) with data in SDTM structure, preformatted and validated (SDTM Hybrid), prior to running their daily activities in SCE (SDTM for submission, ADaM generation, etc.).

SCALA objectives are:
- To implement a tool providing flexibility to produce standard and study-specific data in SDTM Hybrid structure
- To create a standard library of jobs for populating standard and Therapeutic Area SDTM Hybrid domains
- To propose an efficient organization and process to enable the generation of SDTM Hybrid

The figure below describes the new clinical data integration process with metadata driven approach. This process is based on the integration of clinical data from the new CDMS and clinical data from external providers within an Operating Data System (ODS). SCALA project consumes this integrated data after extraction, transformation, and storage for study milestones as source data. SAS[®] CDI has been selected in order to produce standard library and study data based on SDTM Hybrid structure and SANOFI standards by using a combination of metadata and jobs. The output is generated by study/domain and called SDTM Hybrid as it includes all study SDTM variables, but SDTM variables requiring advanced statistical derivations will not yet be populated. There are no supplemental qualifier datasets in the SDTM Hybrid mart. Supplemental qualifier variables will be retained within the parent domain. Study/domain SAS datasets are delivered to the B&P team in order to finalize submission SDTM domains and create ADaM domains for analysis purpose.
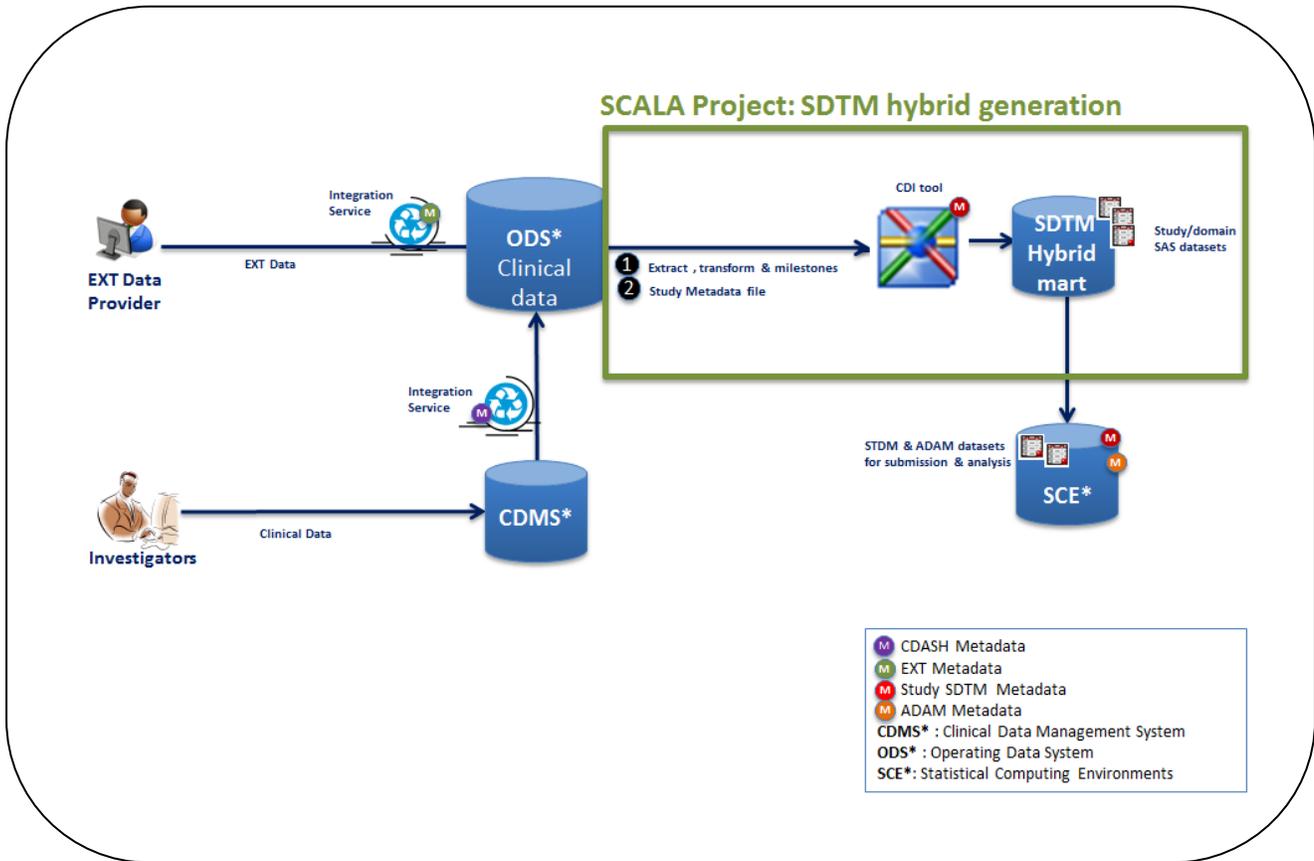
Figure 1 : Clinical Data Integration with metadata driven approach

HOW TO IMPLEMENT SAS® CDI SOLUTION AS PART OF SDTM GENERATION?

SAS® CDI is a metadata driven tool which meets the metadata driven approach described above. The global approach is to run jobs that map and derive source metadata to target domain metadata using transformations in SAS® CDI. In addition, compliance checks allow the validation of the target structure output.

Note that:

- Source metadata is obtained by extracting study clinical data from ODS in SAS dataset format

- Target metadata is created by creating study-specific custom data standards based on the SDTM metadata model, registering these custom data standards to SAS Clinical Standards Toolkit, and then importing these custom data standards into SAS® CDI.

In term of infrastructure, the installation of SAS® CDI requires a dedicated physical platform for compute and metadata servers and the Network Area Storage (NAS)

To satisfy quality & business process requirements, it has been decided to implement within the NAS production environment three sub-environments with appropriate user access rights:

- DEVOPS: is dedicated to study SDTM transformation job development and validation. In this sub-environment, job development can be performed by using test or production study clinical data from ODS according to the volume of the study clinical data collection within the CDMS.

- PRODOPS: is dedicated to SAS® CDI job deployment, deployed job execution, and production SDTM Hybrid SAS datasets output storage.

- COMMUN SOURCES: is dedicated to study source data (test and production) shared with DEVOPS and PRODOPS and any study related documentation.
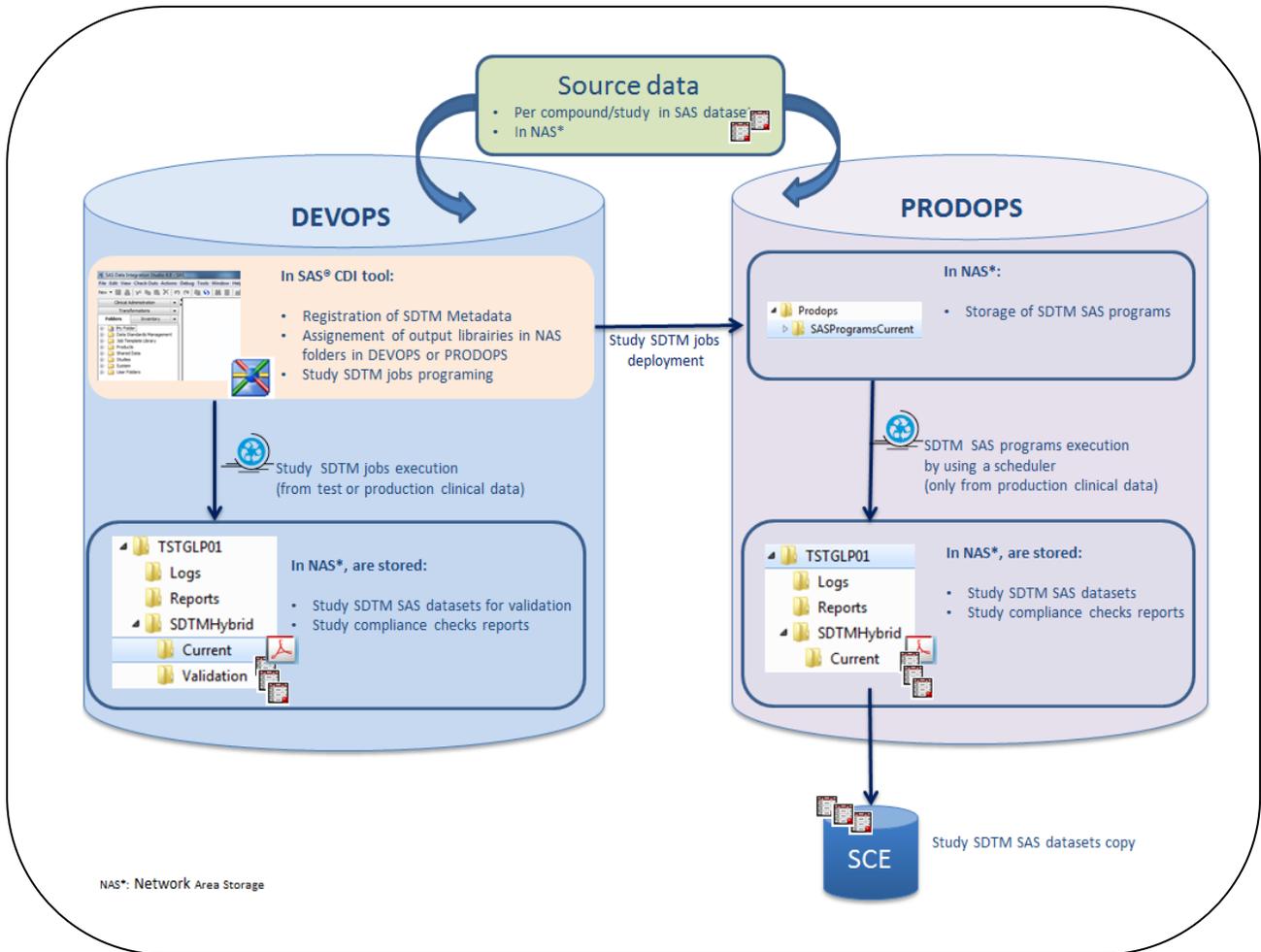
Figure 2 : SCALA architecture plateform

Physical Folder Setup

SCALA project has been focused on the setup of the physical platform with a dedicated folder structure that will be matched with SAS® CDI folders. This step takes into account the business process and the workflows defined within the project. A template folder structure for studies, including data and deployed job execution folders, are set up in Development/Validation (DEVOPS), Production (PRODOPS), and source data environments. Within each study, study-specific SDTM metadata, SAS programs (deployed jobs), output SAS datasets, any study-related documents, and reports such as reports on compliance check results are stored.

When a new study has to be implemented in the SDTM Hybrid Mart, a dedicated script creates all study folders with the appropriate user access rights based on the study folder template. This automated process is one of the key steps for the process industrialization purpose.

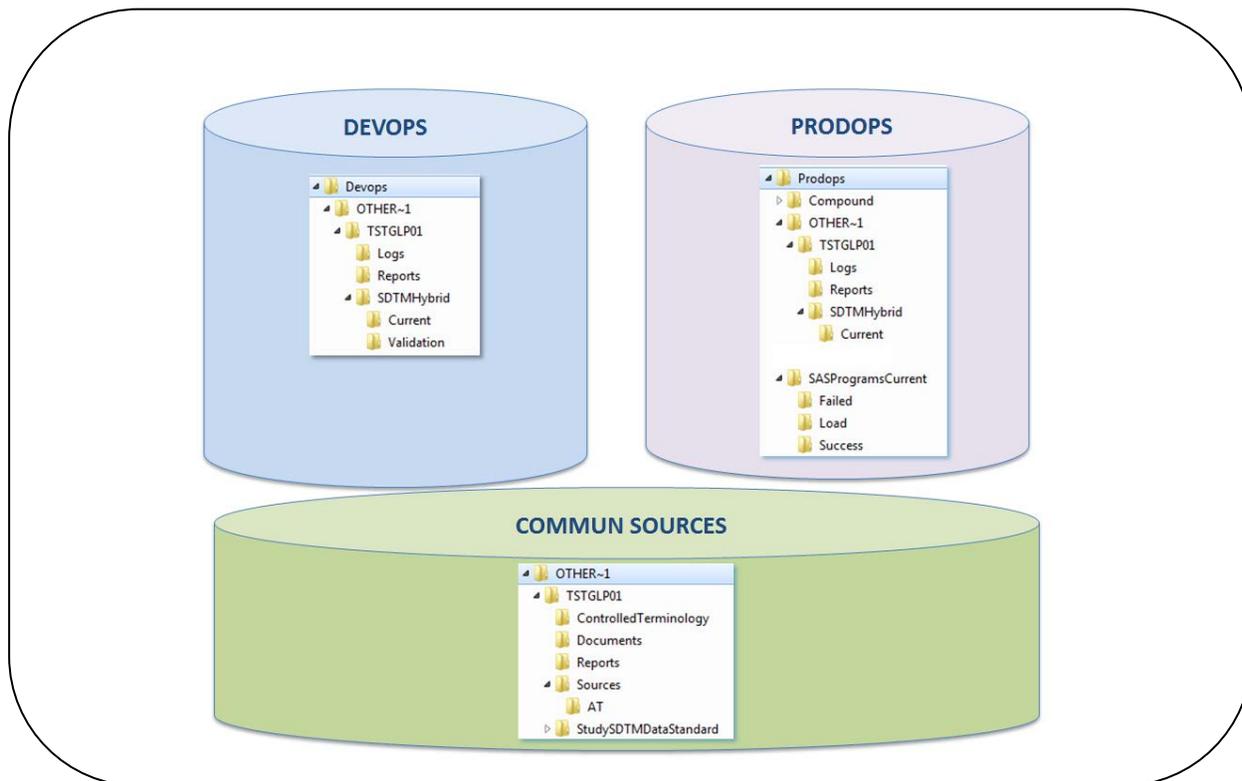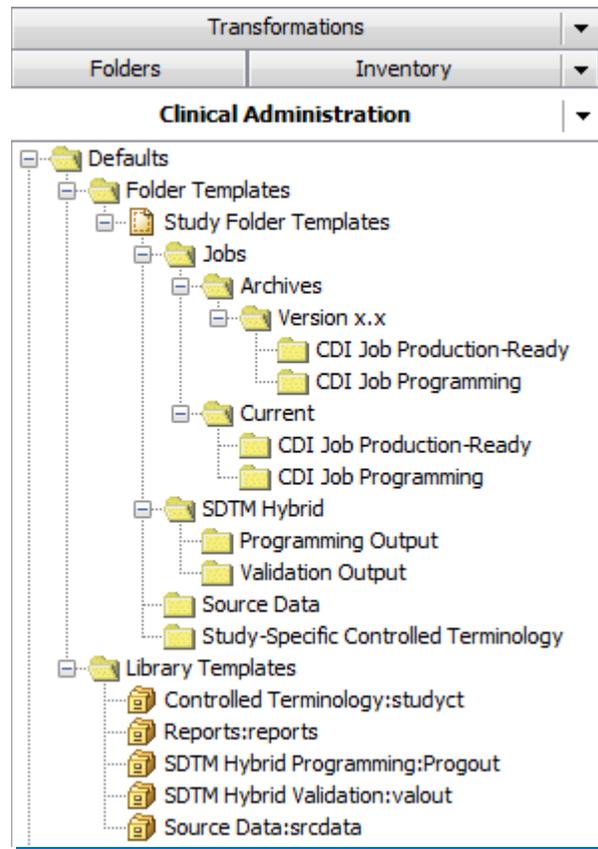The figure below describes the folder structure hierarchy for a given study:

Figure 3: Physical folder hierarchy for a study named TSTGLP01

SAS® CDI Global Configuration

This step has been related to the global configuration of SAS® CDI to fulfill the SANOFI needs described below:

Definition of Default Content: Study Folder Template and Library Templates

Advantages: Allows to setup quickly generic folders hierarchy required for each new study by avoiding any typing or structure error.

Limits: When Library Templates are used for a new study setup, library metadata objects are created using a consistent naming convention and librefs; nevertheless, the physical path of each library has to be manually defined at study level.

Registration of Custom SANOFI SDTM-Based Data Standards Using SAS Clinical Standards Toolkit and SAS® CDI
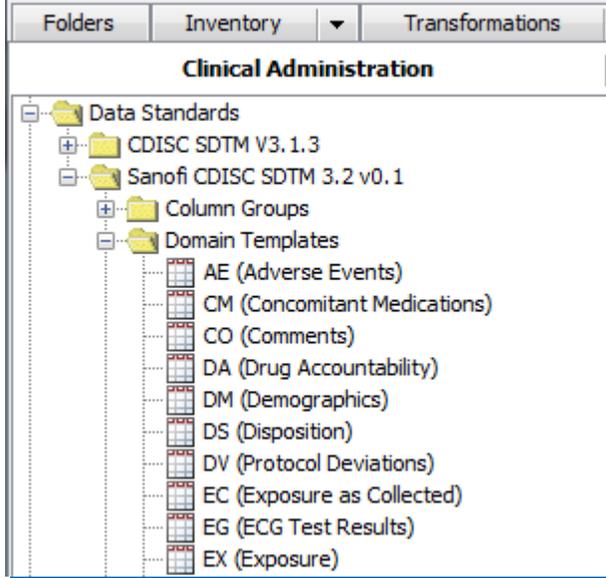
Advantages: Allows the development of a library of SANOFI standard SDTM Hybrid job templates, which are easily reusable at study level.

Limits: Each new customized SANOFI data standard version must be created and registered to SAS Clinical Standards Toolkit, which is an advanced topic. This includes creating approximately 15 datasets to define the custom data standard metadata, quality checking the data sets, registering the data standard to SAS Clinical Standards Toolkit, and then importing the data standard into SAS® CDI.
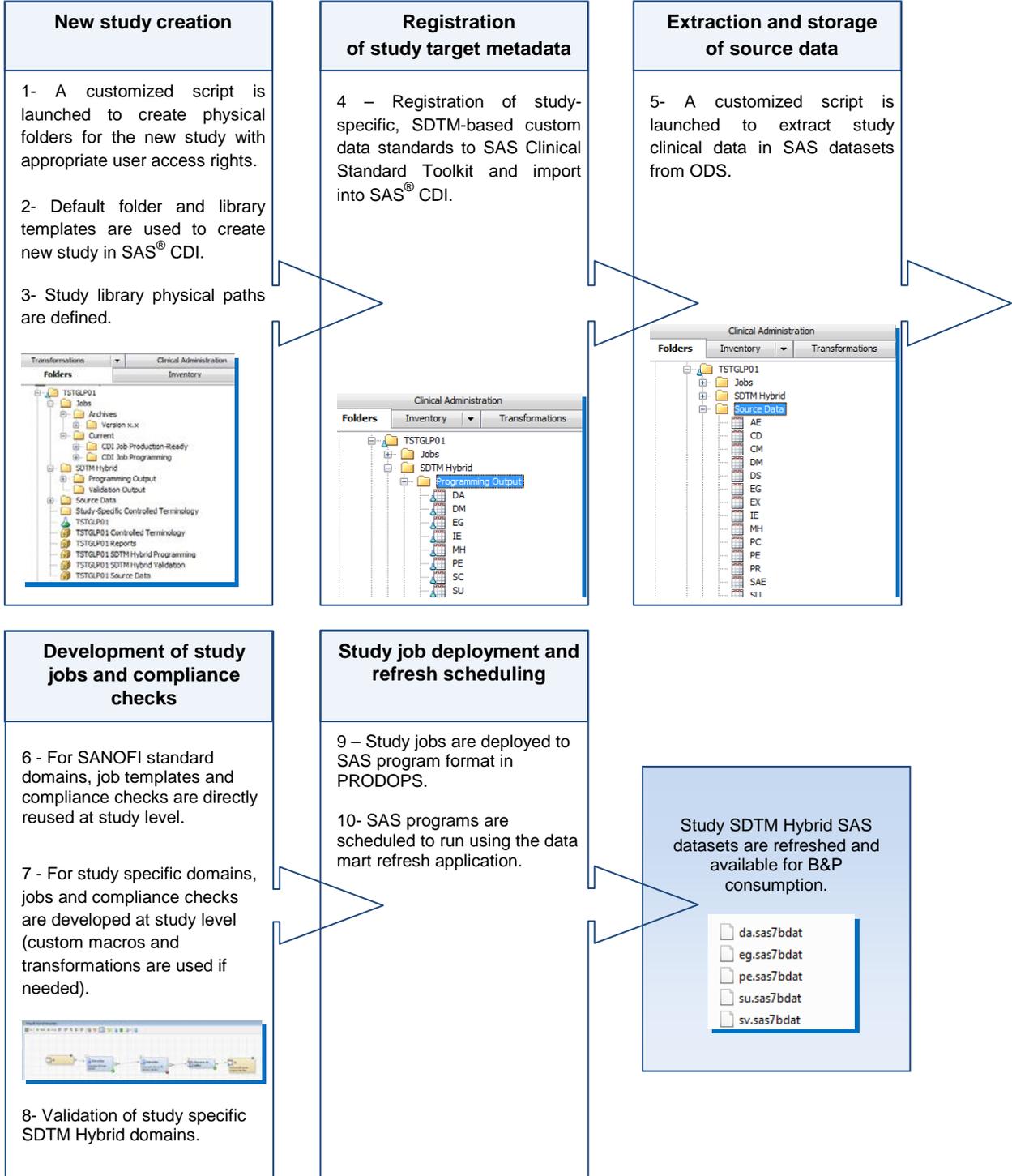
Library Management for Industrialization Approach

To allow the industrialization of study-specific SDTM Hybrid programing, SCALA project has developed a library of SANOFI SDTM job templates, easily reusable at study level by adding the study's source tables and target domains into the job templates. The more the collection of study clinical data is aligned with SANOFI standardization in CDMS, the faster the study-specific SDTM Hybrid programing is. The naming convention of job templates has been defined according to the standard CDMS forms and/or therapeutic areas to identify easily which template has to be copied at study level.

SCALA project has also developed some customized macros and transformations in order to reduce manual programming efforts and ensure the same behavior for the same type of transformation (e.g.: transformation of DATE/TIME source variables into ISO8601 format target variables, or management of missing values such as UNK,MD,NA, etc.).

Generic Process for a New Study Setup and SDTM Hybrid Generation

The physical folder setup, SAS® CDI global configuration, and the library of SANOFI SDTM job templates enable SANOFI to put in place an industrialized global process for the implementation of a new study.

New study creation

1- A customized script is launched to create physical folders for the new study with appropriate user access rights.

2- Default folder and library templates are used to create new study in SAS® CDI.

3- Study library physical paths are defined.

Registration of study target metadata

4 – Registration of study-specific, SDTM-based custom data standards to SAS Clinical Standard Toolkit and import into SAS® CDI.

Extraction and storage of source data

5- A customized script is launched to extract study clinical data in SAS datasets from ODS.

Development of study jobs and compliance checks

6 - For SANOFI standard domains, job templates and compliance checks are directly reused at study level.

7 - For study specific domains, jobs and compliance checks are developed at study level (custom macros and transformations are used if needed).

8- Validation of study specific SDTM Hybrid domains.

Study job deployment and refresh scheduling

9 – Study jobs are deployed to SAS program format in PRODOPS.

10- SAS programs are scheduled to run using the data mart refresh application.

Study SDTM Hybrid SAS datasets are refreshed and available for B&P consumption.

Specific Processes

Specific procedures are set up in order to fulfill some specific business processes and/or regulatory requirements such as:

- *Management of SANOFI custom data standards:* the STDM metadata used for each study has been defined by SANOFI based on CDISC standards and SANOFI specificities. The procedure describes how to import SANOFI standards within SAS$^®$ CDI and how to maintain them using the SAS$^®$ Clinical Standard Toolkit. The challenge here is how to be able to industrialize this process and optimize the technical steps. Customized development has been integrated for the management (storage and archiving) of the physical files that define the custom data standards. A comparison report will be generated in order to identify discrepancies between two data standard versions and better analyze the impact of the data standard version upgrade on studies.

- *Management of source data for study milestones:* source data has to be archived for study milestones (interim analysis, DMC, partial and final lock, etc.). After notification of study milestones and refreshing of the ODS, a SCALA developer is ready to launch the refresh process with the milestone in order to store the source data in a specific folder. This data could be submitted to regulatory authorities.

- *Job versioning:* After each update of SAS$^®$ CDI jobs, an archived version with the applicable metadata is created in order to ensure traceability of job modifications and to be able to re-run previous job versions if needed.

- *Job deployment:* only validated jobs will be deployed in PRODOPS. SCALA developers will deploy jobs twice: the first one to fulfill the backup & restore process and the second one to be executed by the scheduler.

- *Scheduling refresh of output SAS datasets:* there are 4 types of refreshing of the SDTM Hybrid Mart:

    1. Refresh during study conduct: This refresh is planned on a weekly basis for routine mode.

    2. Refresh during milestone period: The frequency of the scheduled refresh is increased before the study milestone date (every 2 days) as B&P workload is important.

    3. Refresh on demand: This refresh is requested occasionally by B&P for specific needs.

    4. Refresh from previous study milestones data: This refresh is requested occasionally by B&P in case of errors in produced jobs or if there is a need for further investigation.

Security Management

The management of the user access by defining specific profiles and user access rights in both SAS® CDI and physical folders is described below:

CDI Administrator is responsible for the management of the standard job library and custom SDTM-based data standard metadata. He/she also responsible for the maintenance of the SAS® CDI configuration.

He/She has access to DEVOPS and some PRODOPS folders.

He/She is not allowed to have Write access to the SDTM Hybrid Mart or source data physical folders.

CDI Developer is responsible for the setup of new studies using study folder and library templates. He/she is responsible for importing data standards (including study-specific custom data standards) into SAS® CDI, as well as the creation and/or update, validation, and deployment of CDI jobs. He/she will be in charge of the impact analysis following data standard metadata updates.

He/She has access to DEVOPS and some PRODOPS folders.

He/She is not allowed to have Write access to SDTM Hybrid Mart and source data physical folders.

Scheduler role is an automated user account responsible of the creation of the physical structure folders, extraction of source data and the execution of deployed CDI jobs based on the data refresh requests.

Support Service

The new SCALA clinical data integration service will be composed of 3-5 resources during the first year based on the expected volume of the new enrolled studies in the UNIFY program. In mid-term, around 100 studies per year are expected; therefore, resources will be increased accordingly. The first expected studies will be related to Diabetes and Oncology. Further therapeutic areas will be integrated, and the standard library will be enriched. SCALA team will work closely with the study team including data managers, study biostatisticians, and programmers but also with technical teams responsible for the integration of clinical data and the storage of metadata in ODS. The challenge is to deliver validated SDTM Hybrid domains respecting service level agreements by determining the optimal configuration, processes and performance. 3 to 6 weeks have been estimated for the study setup and delivery of SDTM Hybrid domains based on the standardization of the study source metadata. The quality and efficiency of the service put in place will be crucial for the integration of new studies.

CONCLUSION

SAS® CDI solution allows the industrialization of the SDTM Hybrid generation needs. Some additional customized developments and enhanced features have been implemented in order to optimize the business processes and satisfy end-user needs. A transition phase is necessary to anticipate and accompany this change within the new service organization.

CONTACT INFORMATION

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

> Miss MOURIN Nadine
> SANOFI
> 1 Avenue Pierre Brossolette
> 91385 Chilly-Mazarin Cedex - France
> Work Phone: +33 (0)1 60 49 45 26
> Email: Nadine.mourin@sanofi.com
> Web: www.sanofi.com

Brand and product names are trademarks of their respective companies.