Automating ADaM – How to Efficiently Create CDISC ADaM Specifications and Automate their Transformation into Datasets

Christof Binder & Helen Hart, F. Hoffmann-La Roche
christof.binder@roche.com, helen_claire.hart@roche.com

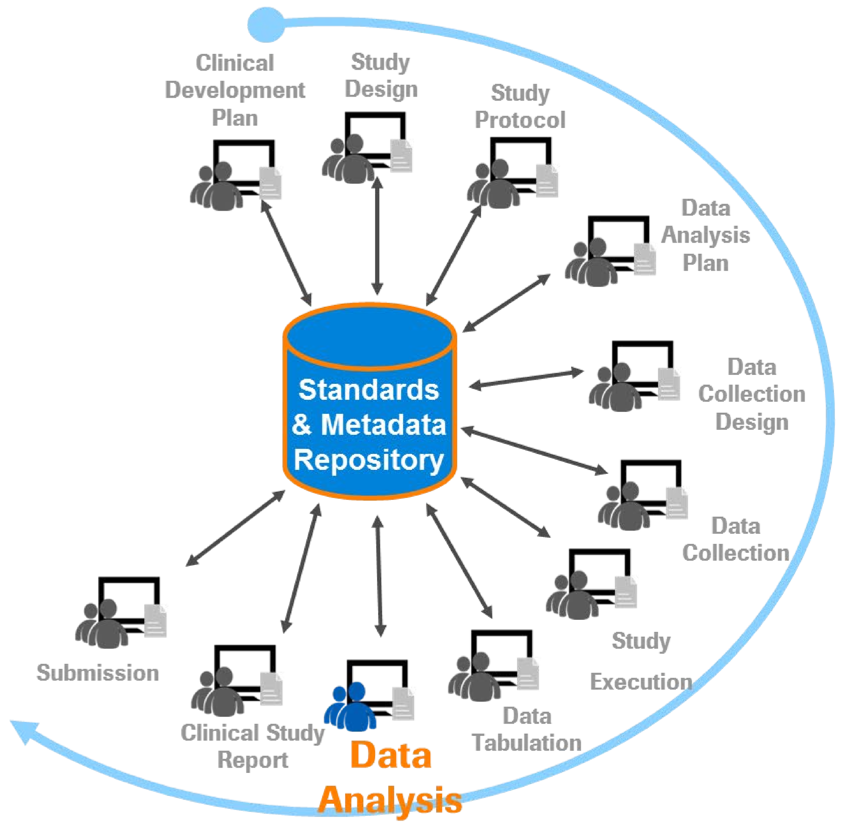
Abstract

In December 2014, the FDA published binding guidance that will require study data to be submitted in electronic format in conformance with CDISC standards. An important component is the production of the analysis datasets following the Analysis Data Model (ADaM).

Challenges

- ADaM datasets are large and repetitive in structure, therefore tedious to program manually
- The task of managing consistency between common variables across datasets can be onerous
- The ADaM standard requires metadata showing clear and full traceability, which can be complex to handle. Questions arise over how to store the metadata, how to present it, how to make it clear to a reviewer and how to make it accessible for tools and people

Where is ADaM in the Study Workflow?



AUTOMATED PROGRAMMING

- In order to create the ADaM datasets efficiently Roche has developed a series of standard programming modules in the form of standard SAS macros
- For the Basic Data Structure (BDS) datasets, the analysis parameters are stored in files in machine readable format and are processed automatically by the macros
- The programming modules can be used independently of each other and individual modules can be replaced with custom code, allowing users the flexibility to deviate from the standard if required
  - o For example, if a user wanted to employ a new method to calculate patient age, then the user would be able to provide a new module within the defined structure without affecting the overall program

DEVELOPMENT OF ANALYSIS CONCEPTS

Analysis concepts drive and underpin the variable derivations and are applied consistently across parameters or to variables across datasets where appropriate

Example

The **analysis concept** "PCHG" (Percentage change from baseline) is as follows:

> *"(Change from baseline value/Corresponding baseline value) * 100*
> *Baseline value must be non-zero."*

A corresponding associated derivation for the corresponding variable "PCHG" for a specific BDS dataset could be:

> *"Set to (Change from Baseline [ADEG.CHG] divided by Baseline Value [ADEG.BASE]) multiplied by 100*
> *Do not compute if Baseline Value [ADEG.BASE] is 0."*

Here, the analysis concept would be applied in all ADaM datasets where percentage change from baseline is calculated. This leads to a generic programming module (e.g. SAS macro) being called within a standard program

GDSR

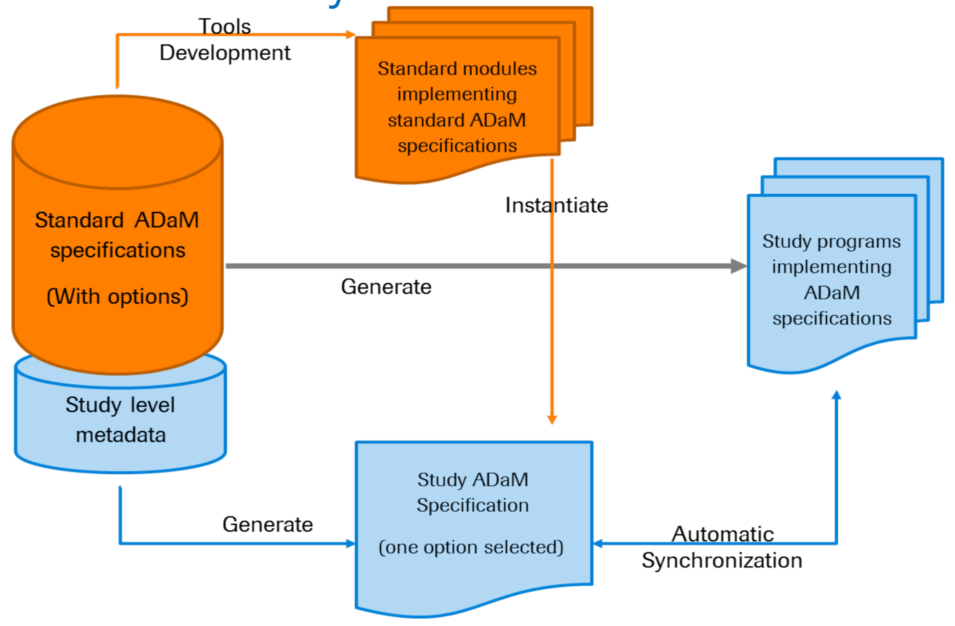
What is it?

- Global Data Standards Repository (GDSR), the metadata repository at Roche, is the central source for standard study metadata
- It is a triple store based on semantic web technology, describing and linking data using Resource Description Framework (RDF)

Functions

- Our ADaM specifications are generated directly from the GDSR
- All the dataset, variable and value level metadata for our ADaM datasets are populated in the GDSR from the point of initial creation and defined only once, offering a "single source of truth"

Analysis Data Generation



Advantages of approach

- The use of the GDSR facilitates the traceability of analysis data
- Wherever the same analysis concepts and derivations are used, the same generic programming modules (eg. SAS macros) are applied
- The centralized, generic and automated approach minimizes the risk of discrepancy across datasets and facilitates their production