

Paper number: AD03

Data Standardisation, Clinical Data Warehouse and SAS® Standard Programs

Jean-Marc Ferran, Standardisation Manager, MSc¹
Mikkel Traun, Functional Architect, MSc¹
Pia Hjulskov Kristensen, Business Implementation Specialist, MSc¹

¹ Clinical Data Warehouse Operations, Global Development, Novo Nordisk A/S, Bagsværd, Denmark

Keywords: Data standards, Clinical Data Warehouse, Metadata, SAS® Standard programs, CDISC Data model, Module based development, Standard implementation methodology.

Abstract

At Novo Nordisk A/S, a Clinical Data Warehouse project was initiated in order to optimize data processing and reporting within the conduct of our clinical trials and cross-trials for submission to the authorities. The Clinical Data Warehouse is a 21 CFR part 11 compliant global system and resides on a common data model based on CDISC SDTM and extensive use of protocol and statistical metadata.

For analysis and reporting in the Clinical Data Warehouse system, SAS® standard programs are used to report trial results. Development of such components follows a standard implementation methodology and a module based approach in order to ensure code quality, consistency, portability, reusability and easy maintenance.

A library of more than one hundred SAS® standard program components was initially released and is used in the first trials reported in the Clinical Data Warehouse enabling faster delivery of output and easy development of new SAS® programs.

Introduction

At Novo Nordisk A/S, a Clinical Data Warehouse project was initiated in order to optimize data processing and reporting within the conduct of our clinical trials and cross-trials for submission to the authorities. The Clinical Data Warehouse is a 21 CFR part 11 compliant global system and resides on a common data model based on CDISC SDTM and extensive use of protocol and statistical metadata. At first, the clinical data is sourced from OC to the Clinical Data Warehouse, mapped to a common data model within a clinical data repository and finally exported to the statistical computing environment (SAS Drug Development®) where SAS standard components are available for generating business rules and reports. Protocol and statistical metadata are captured via a Java application, stored in the Clinical Data Warehouse and are then exported into SAS datasets within the statistical computing environment of the Clinical Data Warehouse for use in the reporting of our trials.

At an early stage in the project, a crucial need for standardisation and alignment of all layers of the data chain was established from protocol, CRF to statistical standards (Reports, business rules and statistical metadata) in order to get the expected benefits. Different working groups specialised in each data area (protocol, CRF and statistical standards) were set-up outside the project to design processes and tools to collect and maintain the different data standards to be used in the Clinical Data Warehouse.

For analysis and reporting in the Clinical Data Warehouse system, SAS standard components are developed and validated to be used in Novo Nordisk trials across R&D centres, indications and therapeutic areas. Development of such components follows a standard implementation methodology and a module based approach in order to ensure code quality, consistency, portability, reusability and easy maintenance. Each standard program component follows the same development cycle: requirement specification, code writing, program documentation, code review, user guide writing, test case writing, test case execution and validation report writing. Each standard program is broken down in different standard "building-block" SAS macros (BBM) dealing with specific parts of the code. The idea behind this framework is to design sub-components that can be reused in trial-specific programs and across SAS standard programs in order to minimize overall effort in programming, documenting, validating and maintaining the standard program library.

In this paper, we detail implementation methodology and module based approach we used for the development of our library of more than one hundred SAS standard macros, programs or utilities and discuss benefits and challenges associated with such a new way of working in our organisation.

Clinical Data Warehouse data flow and standards

At an early stage in the CDW project, a crucial need for standardisation and alignment of all layers of the data chain was established from protocol, CRF to statistical standards in order to get the expected benefits. See Figure 1.

Different working groups specialised in each data area (protocol, CRF and statistical standards) were set-up outside the project to design processes and tools to collect and maintain the different data standards to be used in the Clinical Data Warehouse:

- Protocol metadata standards group -> Protocol metadata
- CRF standards group -> CRF standards
- Statistical standards group -> Reports, business rules and statistical metadata

Reports templates and business rules are agreed upon across our R&D centres and across functions. Departments in Denmark, US, Singapore and Japan were involved with contributions from trial managers, data managers, statisticians, statistical programmers, medical writers and international medical officers.

Report templates are general layouts that are specific to a type of Report (Summary table by visit, shift table, mean plot, etc.). Reports are of 3 types: Tables, Figures or Listings. Figure 2 shows a report template specification with layout and associated requirements for shift tables. Labels in the report such as treatment labels, visit description labels, topic code and associated category labels, formats or code lists are controlled and sourced from metadata.

Business rules refer to data decisions and derived variables (Definition of analysis sets such as ITT or PP, missing value rule, BMI calculation rule, etc.).

Trial Managers define Protocol metadata according to controlled terminology when writing protocols and Trial statisticians define then statistical metadata when writing Statistical Analysis Plans. Protocol and statistical metadata are then captured in the system via a Java application and then exported into SAS datasets within our statistical computing environment for use in the reporting of our trials.

Illustration of CDW Data Flow & Standards

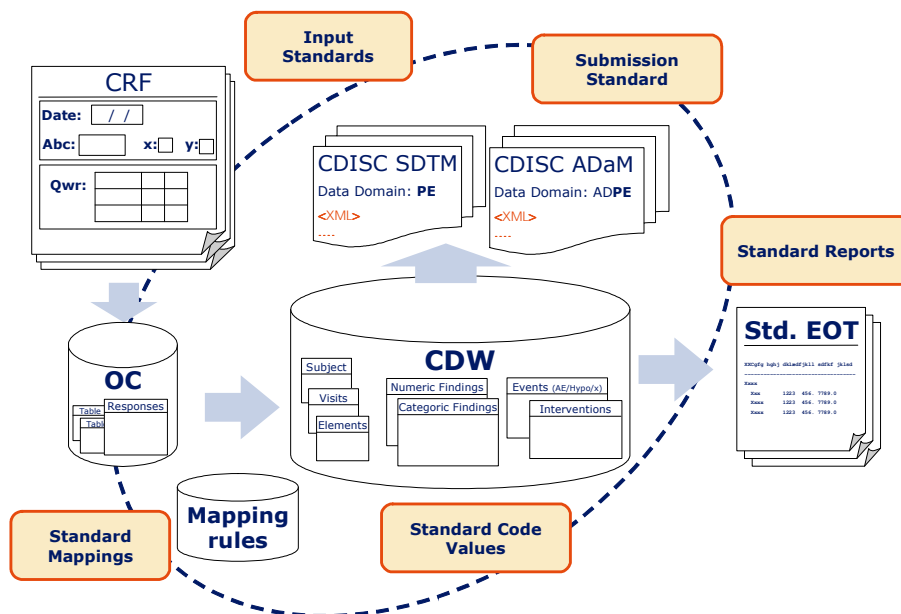


Figure 1: Clinical Data Warehouse (CDW) Data Flows and Standards

Shift Table for [parameter] - [Analysis Set]<, [subpopulation selection criteria]>								
Treatment1					Comparator			
Number of Subjects	xxxx				xxxx			
[Visit X description]								
[Baseline Visit description]	Cat1 N	Cat2 N	CatN N	Missing N	Cat1 N	Cat2 N	CatN N	Missing N
[BBB]								
category1	xxxx	xxxx	xxxx	xxxx	xxxx	xxxx	xxxx	xxxx
category2	xxxx	xxxx	xxxx	xxxx	xxxx	xxxx	xxxx	xxxx
categoryN	xxxx	xxxx	xxxx	xxxx	xxxx	xxxx	xxxx	xxxx
Missing	xxxx	xxxx	xxxx	xxxx	xxxx	xxxx	xxxx	xxxx
[CCC]								
Repeat...								
<Footnotel-n>								
Footer text								

Specific Layout decisions

- Selection (where clause): Characteristics (parameter, e.g. Physical examination, ECG, Biochemistry), Visit X, population, subpopulation.
- As default all individual categorical findings within the group (BBB, CCC, ...) will be presented but it is possible to select subsets.
- BBB, CCC, etc. will be depend on selected characteristics group (e.g. Skin disorders, Abdomen, Haemoglobin, WBC). Default order of BBB, CCC, etc. will be defined for each report table, but the order will be changeable.
- Categories 1-N will depend on the selected characteristics (ex. ECG: Normal, Abnormal NCS, Abnormal CS; Laboratory Measurements: High, Normal, Low). Default order of categories (category1-N) will be defined for each report table, but the order will be changeable.
- Categories with no subjects will also be presented with 0 as N.

The total column has in this template been omitted.

Figure 2: Example of report template specification: Report template for shift table

Implementation of SAS standard programs

SAS standard programs are developed and validated in order to support production of standard reports and business rules. Each standard program is broken down in different standard “building-block” SAS macros (BBM) dealing with specific parts of the code (extraction, processing, reporting, etc., See Figure 3) or specific functions (computing geometric mean, compute column width in tables, etc.). Those standard building-block macros are developed, documented and validated separately and then integrated into main standard programs (See Figure 4 where a full design is illustrated) producing standard reports or business rules. The aim is to develop standard “building-block” macros that are portable across SAS servers and can be reused across several main standard programs in order to minimize the overall effort in programming, documenting, validating and maintaining the standard program library. Each “building-block” macro can also be reused within trial-specific programs to deal with one specific part of the code.

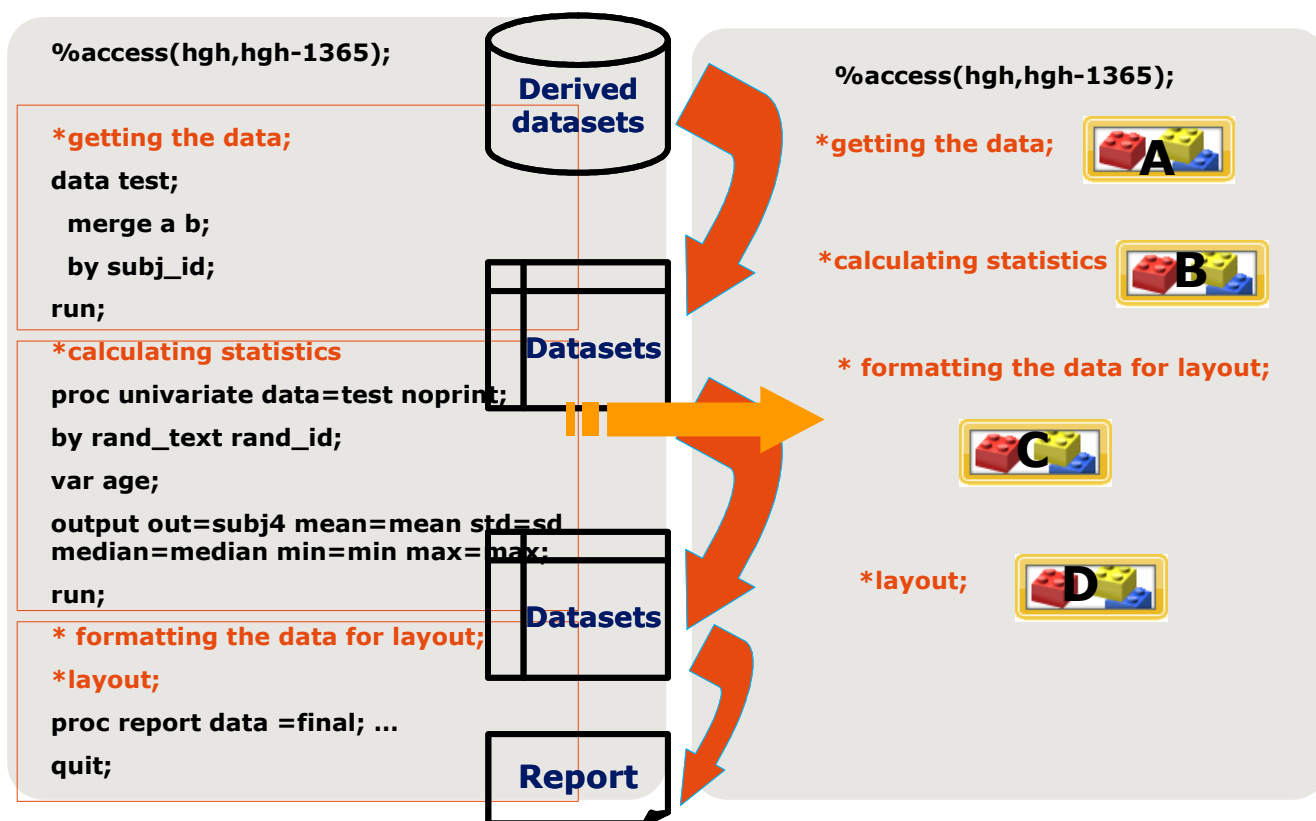


Figure 3: Building block macro principle or the “Lego®” brick principle.

Rather strict coding standards and practices were defined in order to ensure consistency and quality across the standard components. Those coding standards deal with syntax (indentation, comments, etc.), naming convention or SAS code optimization and practices (Defensive programming, work library management, global/local macro variable use, etc.). Each piece of code is peer-reviewed and this process is documented using a code review sheet listing the most important coding areas to be verified.

For each standard component, a test plan is written in order to validate expected functionalities and error handling against requirement specifications.

User guides are written and provide necessary information to the users in order for them to understand and use each component. Bugs and restrictions found along the way are also listed in user guides until they can be fixed in connection with later releases.

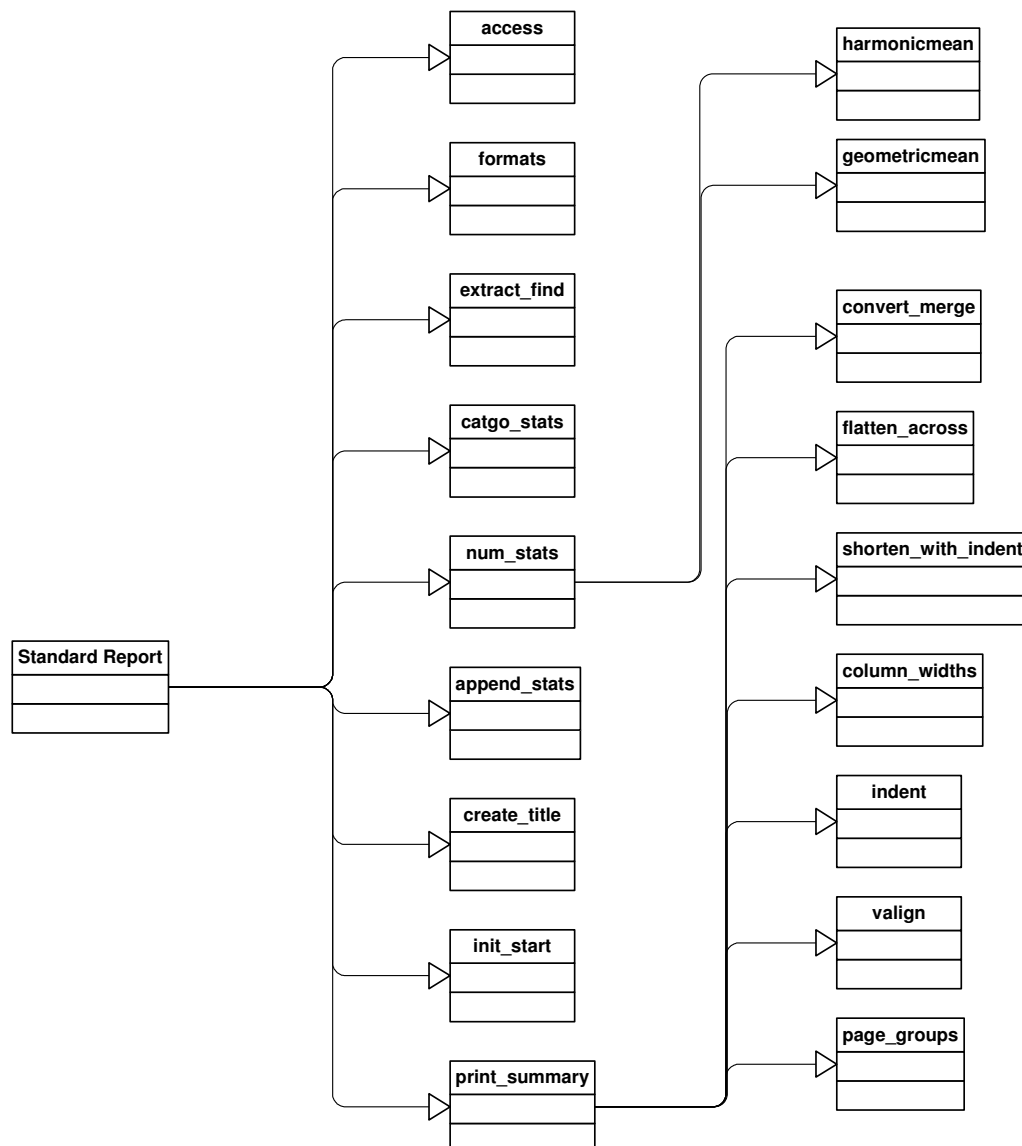


Figure 4: Example of standard program design using “building block” principle

CDISC SDTM implementation and use of metadata

Standard programs run on a denormalised data layer residing on a common data model based on CDISC SDTM+ called Derived Data Marts (DDMs). The DDMs have all data decisions and derivations implemented according to business rules and are ready for reporting. The physical implementation of the DDMs layer is organised by data areas (trial, flowchart, subject), data types (Numeric, categorical, textual) and general domain classes (Finding, Intervention, Events) in order to facilitate development of general and reusable SAS standard components.

In the DDMs layer, metadata datasets holding protocol and statistical metadata where trial flowchart, labels, units, SAS formats and categories for categorical findings are controlled and used in the standard programs in order to ensure consistency across our trial reports.

Standard Program and Statistical Standards Management

Standard reports and business rules and their implementation into SAS standard programs are subject to changes due to bugs or need for extra features. Bugs found in SAS standard programs from the users are reported in an IT service management system (Remedy®) and then reported in the corresponding user guides together with possible work-arounds until they can be fixed in later versions of the relevant components. Change requests to report standards or business rules are sent to the Statistical Standards group that assess them and decide upon implementation. On each study, a SAS programmer has the responsibility to ensure that standards are used as much as possible. Use of statistical standards and standard programs are monitored using KPIs and programming plans for each study are reviewed by statistical standards specialists. This process enables capturing needs for extra statistical standards and ensures extensive use of the standard library.

Use of Standard Program Library

Statistical standards used for each trial must be specified in the Statistical Analysis Plan. A Catalogue of reports and business rules is available. Standard programs supporting implementation of statistical standards are available in standard program library. Building block macros can also be used in trial-specific programs. Figure 5 below describes the process flow for implementation of a new report or business rule. In the best case, a standard program is available to support the given report or business rules. Otherwise, a building block macro (BBM) may support one step of the code. In the worst case, no standard component can be used and the trial programmer must code from scratch or copy and modify an existing program and therefore a third person must thoroughly revalidate it. It is also possible for the trial programmer if a need is identified and resources are available to request a new standard program and have it available in due time.

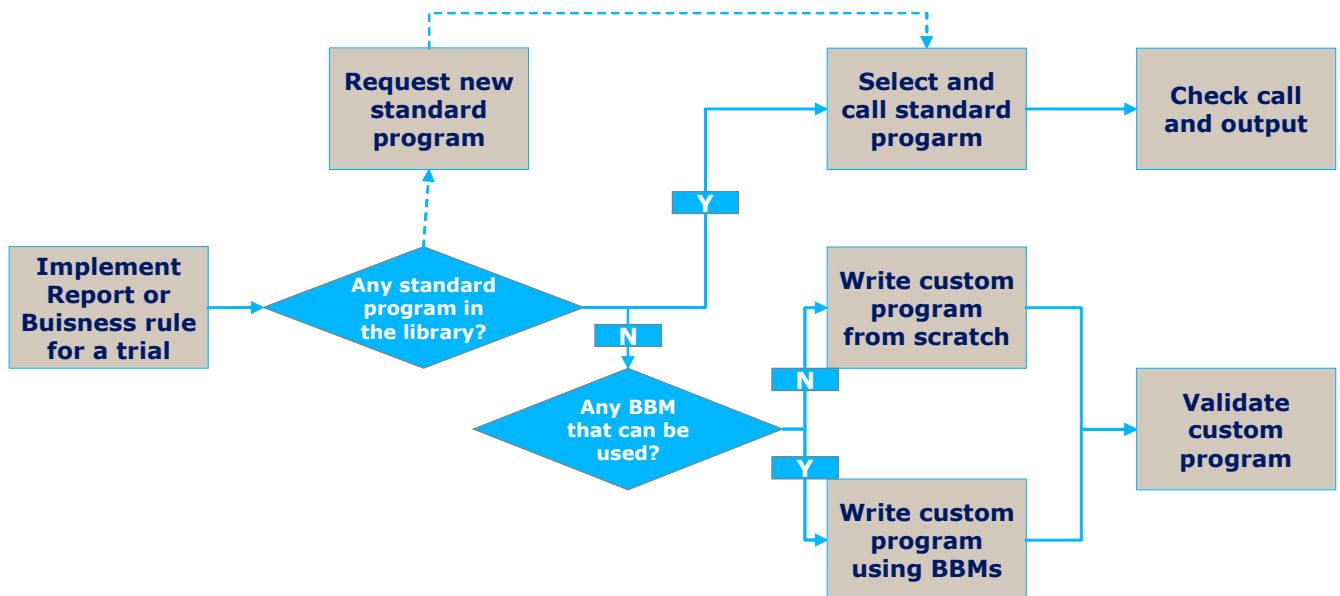


Figure 5: Process flow for use of standard library for a trial

Conclusions

Benefits

- The extended use of standards and metadata in the reporting of our clinical trials and a common adequate data model enabled to develop a smarter and more structured approach to programming at Novo Nordisk.
- A standard library of more than one hundred SAS standard components is implemented and is available for use in our clinical trials across R&D centres, therapeutic areas and projects enabling faster delivery of output and easy development of new SAS programs.
- The communication with other stakeholders such as medical writers or international medical officers has been simplified thanks to a broad and accurate standard reports catalogue.
- Statistical Analysis Plan writing is accelerated too as references to statistical standards can be made.
- And finally, time is saved from routine programming work for statisticians and statistical programmers for more challenging coding and analysis tasks.

Challenges

- Users must learn to work with metadata, controlled terminology and a new data model.
- Users must learn to use new standard program components instead of copying and modifying code they already know, which may give the feeling to have less control on the code and output.
- There is more paper work involved in standards management, e.g. bugs reported in IT system management (Remedy) and change request raised for statistical standards.
- Development of standard programs is a difficult and time-consuming exercise.
- It can also be difficult to have the necessary resources as clinical projects often have the priority.

Contact Information

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

Jean-Marc Ferran (jmferran@gmail.com - +4526214077)

Mikkel Traun (MT@novonordisk.com - +4530758200)

Pia Hjulskov Kristensen (PHJK@novonordisk.com - +4530799939)

Krogshøjvej 51

2880 Bagsværd, Denmark

Web: www.novonordisk.com

Brand and product names are trademarks of their respective companies.