A proprietary, CDASH/SDTM-hybrid data model to expedite clinical data review

M.A. Prodhomme[1], L. Gijsbers[2], T. Van der Spiegel[1]

[1]Janssen R&D, Beerse, Belgium; [2]OCS Life Sciences, 's-Hertogenbosch, the Netherlands

Abstract

In 2016 Janssen identified the need to expedite clinical data review. A proof of concept demonstrated the value of pursuing a new and proprietary data model for data review, serving as single source of truth. The Data Review Model (DRM) that was introduced is strongly based on CDISC SDTM and CDASH. DRM provides full traceability and describes both clinical and operational (system) data consistently across studies. On the longer term, Janssen plans to implement a metadata-driven environment, including data conversion from source data into DRM.
In 2017, OCS Life Sciences and Janssen piloted DRM by implementing a mapping framework that supports both documentation and execution of source to target data mapping. This poster will describe how multiple trials were mapped to support the pilot phase of DRM, to learn, refine and document the value of DRM prior to moving to production implementation [1].

Introduction

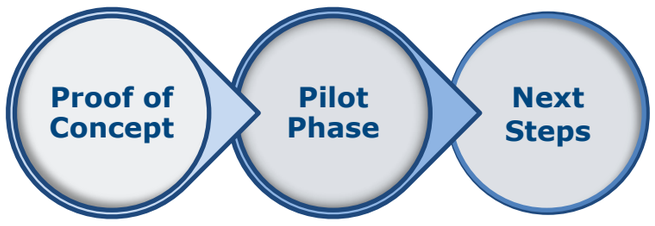
At Janssen, Data Management (DM) activities are outsourced to DM CROs:
- Delivering SDTM datasets to Janssen during trial conduct
- Preparing the SDTM Submission Package after Database Lock.

Janssen DM performs ongoing Quality Control on these SDTM deliverables.

In 2016, spending time evaluating the current data flow, Janssen identified the need to expedite clinical data review. The idea of a new data model was introduced, serving as single source of the truth to all consumers of clinical data. What followed was a proof of concept (POC), showcasing 5 newly designed domains. The Data Review Model, in its very early stages, was born. As part of the POC we tried out several use cases:

- How can we most logically cluster/group information in DRM?
- Avoiding the use of SUPPQUAL and Findings About datatypes.
- How to represent relationships in DRM, without the need for RELREC?
- Can we add value by adding new data (variables or datasets) in DRM?
- How will DRM help when mapping a new exploratory data stream?
- Will DRM allow an easy transformation to SDTM?

Project Life Cycle



Proof of concept expectations:
- Less complex data model
- Focus on data review not on data submission
- Control both structure and content
- Independency from industry guidance
- Additional 'value added'
- Facilitate early access to data

Pilot Phase results:
- Validation of model
- Validation of Mapping Framework
- Confirmation of early access to data for medical review and central monitoring
- Confirmation of full data traceability and high data availability

Next steps:
- Creation of new utilities in the conversion process
- Metadata driven way to populated the DRM Mapping and Metadata
- A fail-safe mechanism to check the incoming source data (before conversion)
- First real live use cases started beginning 2019

Conclusion

To improve the data flow, following a successful proof of concept and pilot phase, Janssen is introducing a new data model to help expedite access to data and facilitate data review operations.

This Data Review Model provides a general framework for describing clinical trial and operational data in a rather simple, well-structured and consistent/uniform way.

It provides clear traceability to the collected source data, it positively impacts the review of data and it allows an easy and controlled transformation to SDTM.

References

1. Lieke Gijsbers, OCS Life Sciences, 's-Hertogenbosch, the Netherlands, Tom Van der Spiegel, Janssen R&D, Beerse, Belgium, A proprietary, CDASH/SDTM-hybrid data model to expedite clinical data review, PhUSE 2018, Paper SI25

2. Bas van Bakel, OCS Consulting, 's-Hertogenbosch, the Netherlands, DIY: Create your own SDTM mapping framework, PhUSE 2016, Paper CD03

DRM Mapping Specification

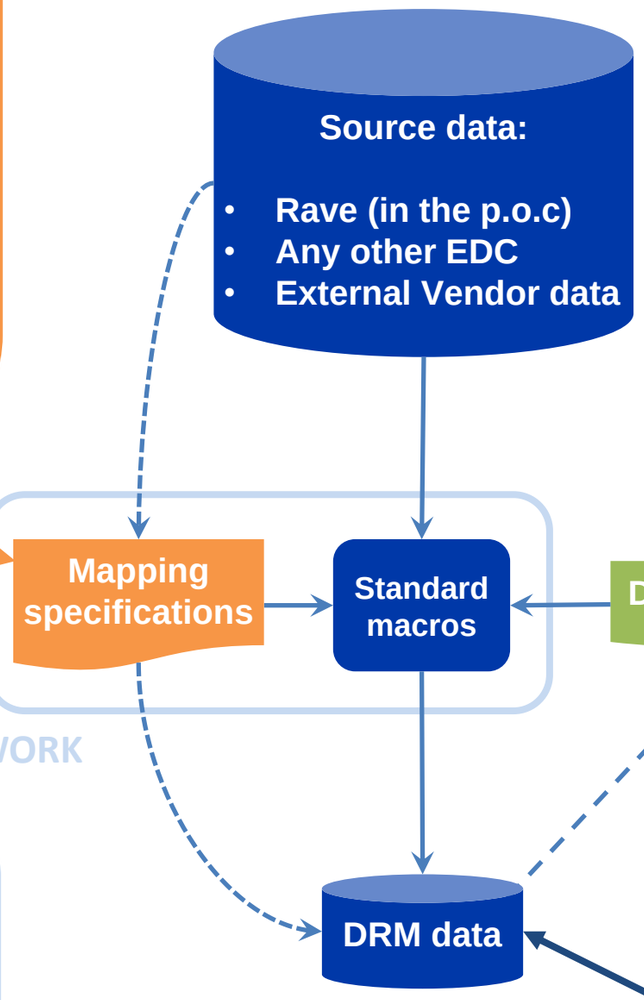
- All source-to-target specifications
- Translation of specifications into code or pseudo-code.
- Recode table
- Data sets order

DATASET	VARIABLE	DRM_DS	DRM_VAR	SPECIFICATION	FUNCTION
source.ae_gl_900yn	PROJECT	AE	STUDYID	Recode according to STUDYID recoding list	RECODE [STUDYID]
source.ae_gl_900yn	STUDYID			Not mapped	NOT MAPPED
source.ae_gl_900yn	SUBJECT	AE	SUBJID	Left justify and uppercase source variable	FUNCTION [SUBJID = STRIP(UPCASE(SUBJECT));]
source.ae_gl_900yn	SITENUMBER	AE	SITEID	Left justify and uppercase source variable	FUNCTION [SITEID = STRIP(UPCASE(SITENUMBER));]
source.ae_gl_900yn	INSTANCENAME	AE	VISIT	Recode according to VISIT recoding list	RECODE [VISIT]
source.ae_gl_900yn		AE	AECAT	Assign value 'ADVERSE EVENTS/SERIOUS AES'	FUNCTION [AECAT = 'ADVERSE EVENTS/SERIOUS AES';]
source.ae_gl_900yn	AEYN	AE	AEYN	Copy from source variable	COPY
source.ae_gl_900yn	AEYN_STD			Not mapped	NOT MAPPED
source.ae_gl_900	PROJECT	AE	STUDYID	Recode according to STUDYID recoding list	RECODE [STUDYID]
source.ae_gl_900	STUDYID			Not mapped	NOT MAPPED
source.ae_gl_900	SUBJECT	AE	SUBJID	Left justify and uppercase source variable	FUNCTION [SUBJID = STRIP(UPCASE(SUBJECT));]
source.ae_gl_900	SITENUMBER	AE	SITEID	Left justify and uppercase source variable	FUNCTION [SITEID = STRIP(UPCASE(SITENUMBER));]
source.ae_gl_900	INSTANCENAME	AE	VISIT	Recode according to VISIT recoding list	RECODE [VISIT]
source.ae_gl_900	AETERM	AE	AETERM	Copy from source variable	COPY
source.ae_gl_900	AECAT	AE	AESPINT	Copy from source variable	COPY
source.ae_gl_900	AECAT_STD			Not mapped	NOT MAPPED
source.ae_gl_900	AESEV	AE	AESEV	Copy from source variable	COPY
source.ae_gl_900	AEREL	AE	AEREL	Copy from source variable	COPY
		AE		**POSTSTEP1: COMBINE SOURCE DATASETS.** Combine the mapped source datasets ae_gl_900yn and ae_gl_900 by merging on calculated values of STUDYID, SITEID and SUBJID.	POSTSTEP1 [PROC sort DATA=work.mapped_source_ae_gl_900; BY studyid siteid subjid; RUN; PROC sort DATA=work.mapped_source_ae_gl_900yn; BY studyid siteid subjid; RUN; DATA work.mapped_combined_ae1; MERGE work.mapped_source_ae_gl_900 work.mapped_source_ae_gl_900yn; BY studyid siteid subjid; RUN;]

DRM Metadata

- Order of DRM variables
- The attributes of the DRM variables
- Key variables for the sorting of records

DOMAIN	NAME	LABEL	TYPE	LENGTH	FORMAT	SORTVAR
AE	STUDYID	Study Identifier	C	40		1
AE	SITEID	Site Number	C	20		
AE	SUBJID	Subject Identifier	C	10		2
AE	VISIT	Visit	C	60		
AE	AECAT	Category	C	200		3
AE	AEYN	Were any Adverse Events Experienced?	C	9		
AE	AETERM	What is the Adverse Event Term?	C	200		4
AE	AESPINT	Is this an AE of Special Interest?	C	120		
AE	AESEV	Severity	C	24		
AE	AEREL	Relationship to Study Treatment	C	33		

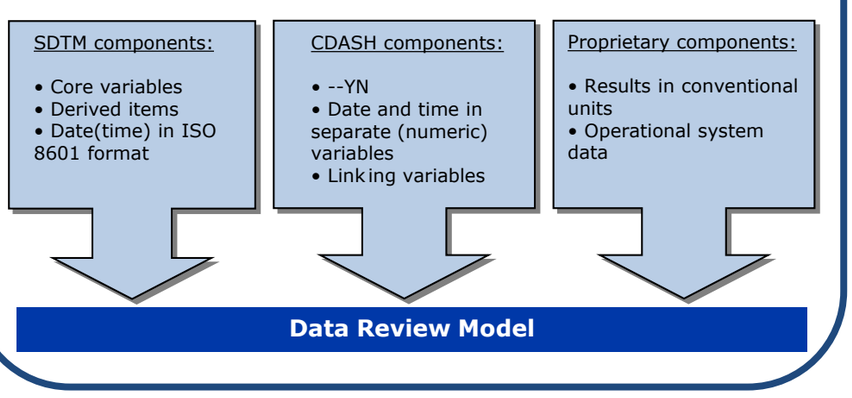


Mapping Framework

- Developed by OCS Life Sciences
- Implemented in Janssen's SAS® Life Science Analytics Framework (LSAF)
- Based on modular structure -> implementation with minimum effort [2]
- Describe and execute source-to-target mappings in a structured way

Data Review Model

The Data Review Model describes both clinical and operational (system) data and is strongly based on CDISC CDASH and SDTM



janssen | PHARMACEUTICAL COMPANIES OF Johnson&Johnson

OCS Life Sciences