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

CDISC UK Network, 3 April 2019

Lieke Gijsbers, OCS Life Sciences, the Netherlands
Nele Claes, Janssen R&D, Belgium

OCS Life Sciences

Janssen | PHARMACEUTICAL COMPANIES OF Johnson & Johnson

Framing the picture

- At Janssen, Data Management (DM) activities are outsourced to DM CROs
  - DM CROs are contracted to deliver SDTM datasets to Janssen during trial conduct
  - DM CROs prepare the SDTM Submission Package after Database Lock
  - Janssen DM performs ongoing Quality Control on these SDTM deliverables

- In 2016, Janssen identified the need to expedite clinical data review
  - Early access to data for Real Time Learning and Decision Making
  - The idea of a new data model was introduced

Proof of Concept

- PoC of a controlled & proprietary data model: Data Review Model (DRM)

- Use cases:
  - How can information in DRM be most logically clustered?
    - Avoiding the use of SUPPQUAL and Findings About data types
  - How to represent relationships in DRM?
    - Avoiding the need for RELREC as known in SDTM
  - Do new datasets and variables, not possible to include in SDTM, add value in DRM?
  - Will DRM help when mapping new exploratory data streams?
  - Will DRM allow an easy transformation to SDTM?

Proof of Concept

- Findings of proof of concept:
  - **Less complex** data model compared to CDISC SDTM
  - Janssen-controlled DRM: **less vulnerable to changing** CDISC SDTM and controlled terminology versions
  - Stores additional '**value added' content**
  - Not a data submission model: **less strict on implementation** on trial level

- Other learnings:
  - Enabling early access to the data in the DRM model requires a high-degree of re-use, from standard or previous trials
  - Data harmonization will require a controlled environment
  - DRM to SDTM conversion was relatively easy

→ Initiation of a pilot project in 2017

Next slides will cover…

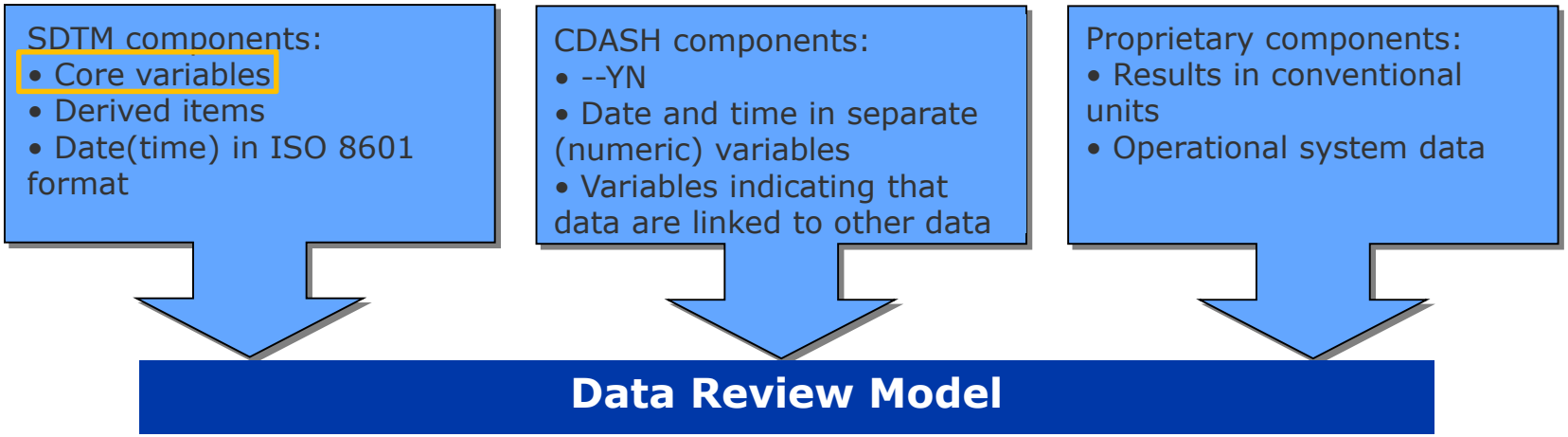
- Key principles of the Janssen Data Review Model
- Insights into the conversion framework used during the DRM pilot phase
- Learnings, next steps and future perspectives

Data Review Model (DRM)

- Strongly based on CDISC CDASH and SDTM

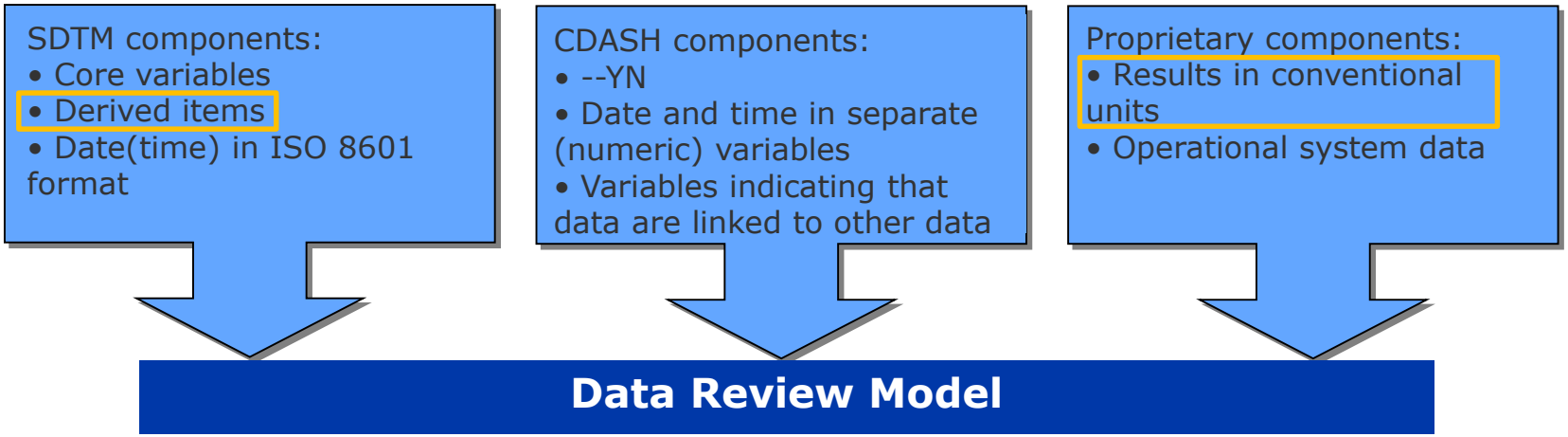
- DRM adheres to the fundamentals of SDTM:
  - Vertical structure
  - Observations are reported in a series of domains
  - Dataset and variable names are standardized

Key Principles of DRM



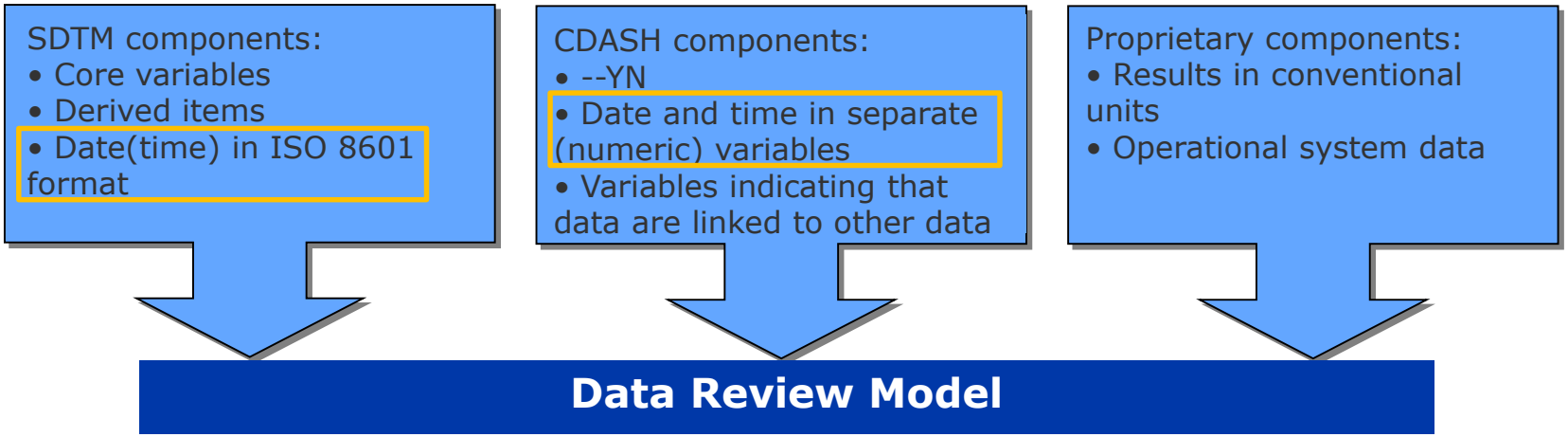
- STUDYID, SITEID, SUBJID, VISITNUM, VISIT
- --TERM, --TRT, --TESTCD
- --OCCUR, --STAT, --ORRES, --ORRESU

Key Principles of DRM



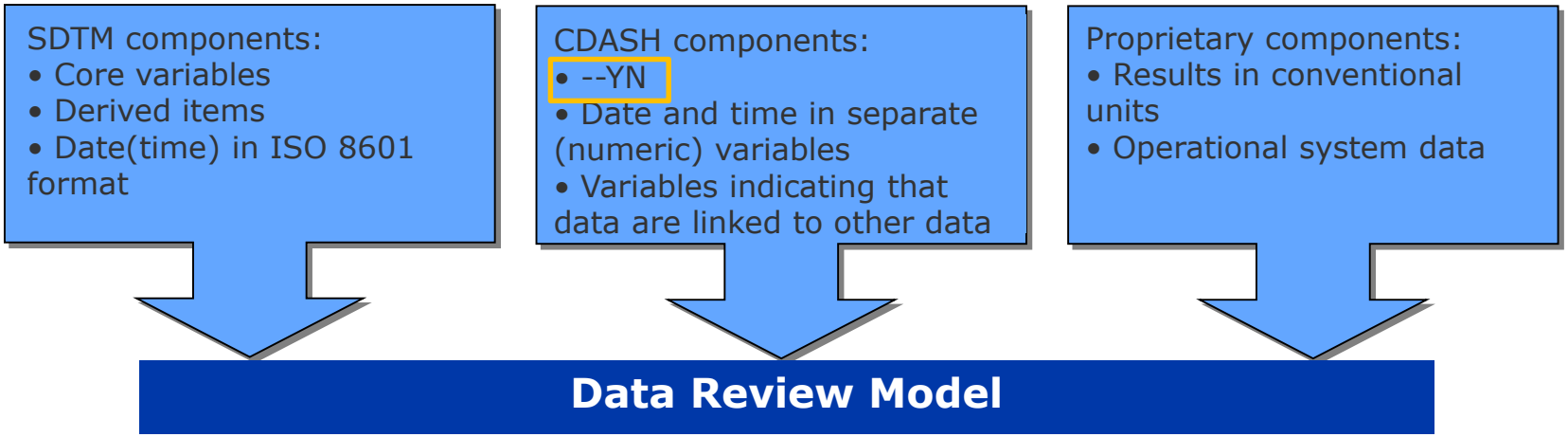
- --DY
- --STRESC, --STRESN, --STRESU

Key Principles of DRM



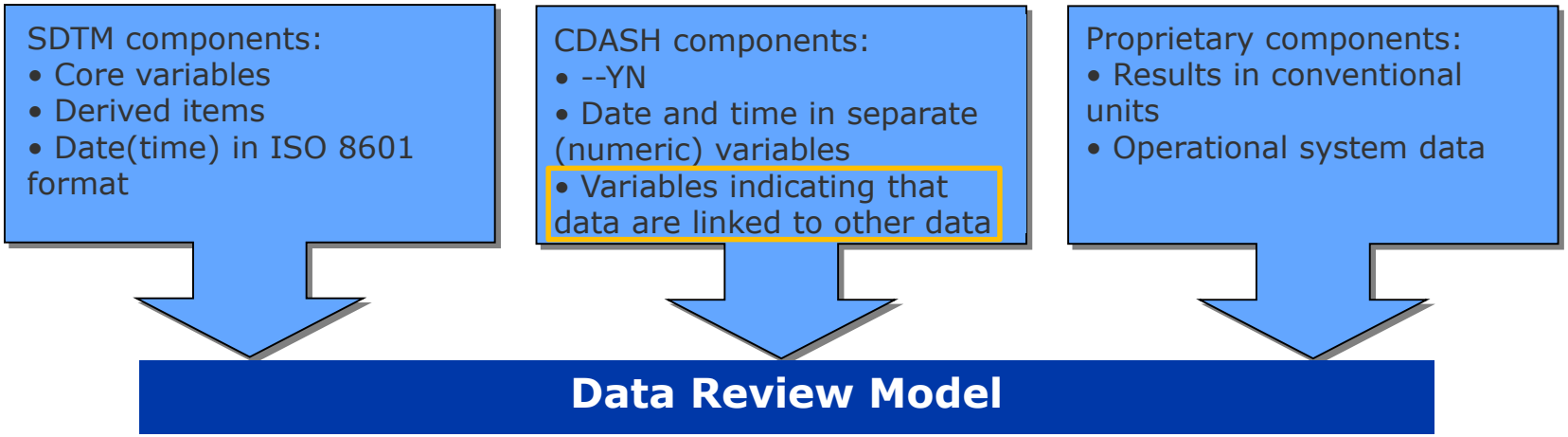
- --DTC
- --DAT
- --TIM

Key Principles of DRM



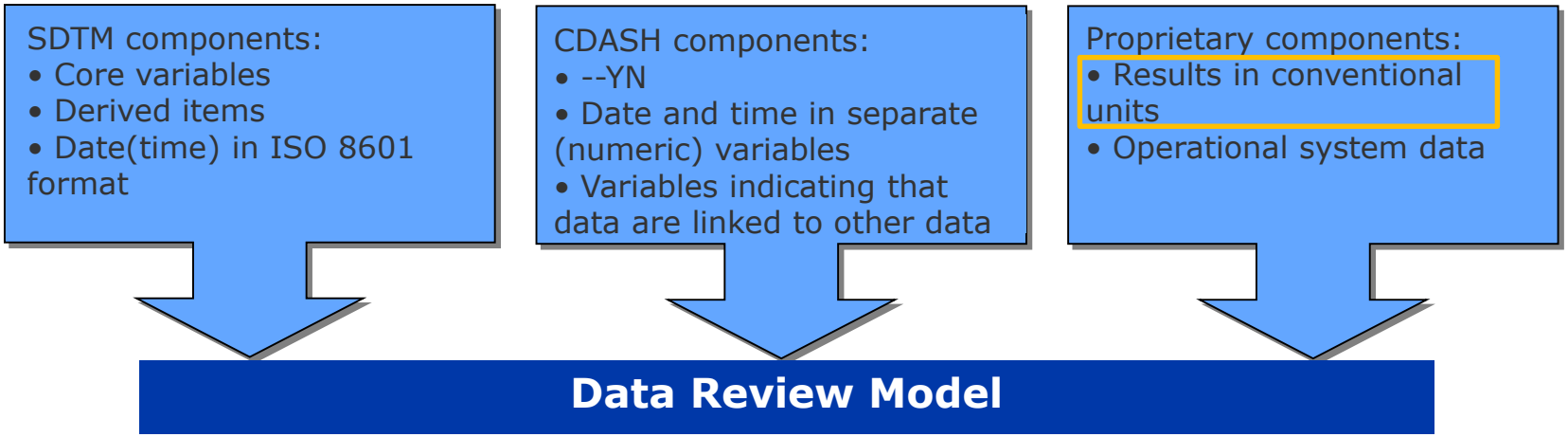
- "Were any adverse events experienced?"
- "Were any medications taken?"
- "Were examination performed?"

Key Principles of DRM



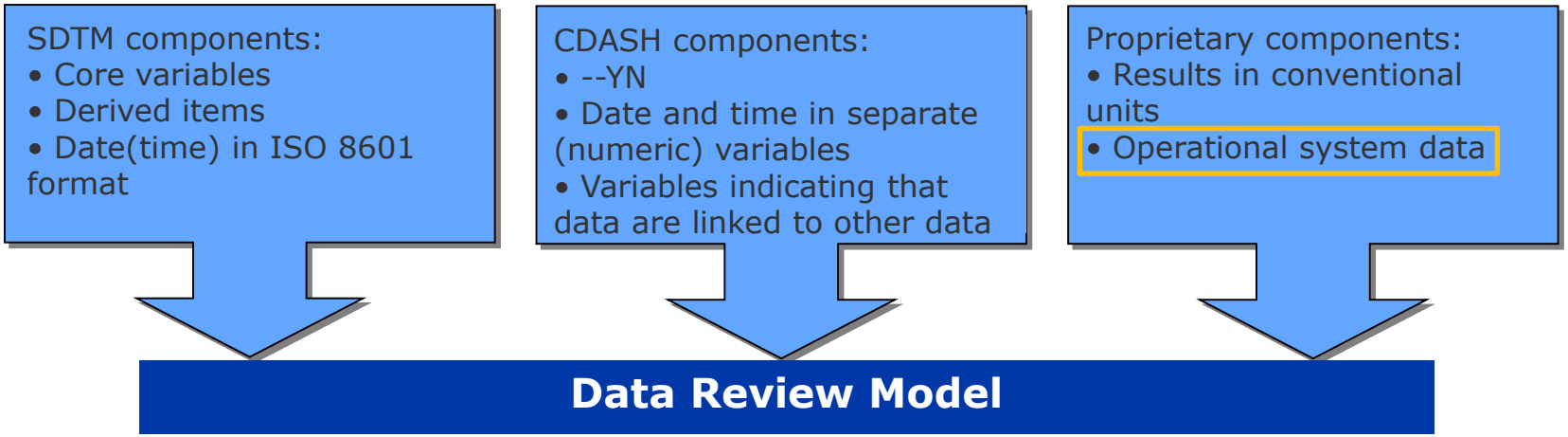
- AESEQ1
- AESEQ2
- AESEQ3

Key Principles of DRM



- --CNRESC, --CNRESN, --CNRESU

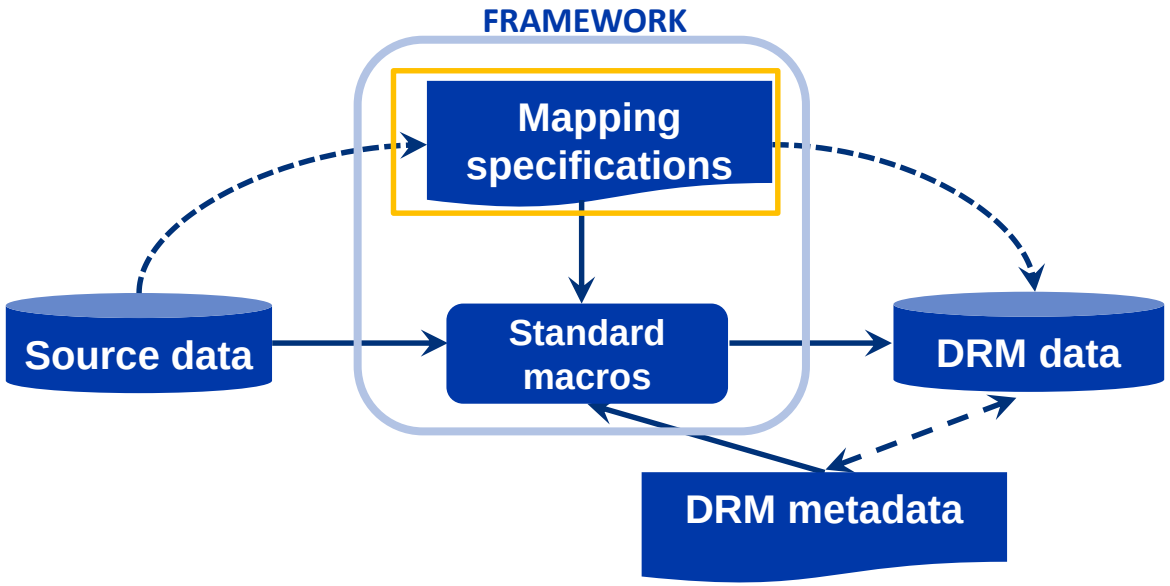
Key Principles of DRM



- Source of the raw data
- CRF page name and number
- Date indicating when the record was initially created and last updated

Mapping Framework

- Data converted using the mapping framework[1]
- Implemented in Janssen's SAS® Life Science Analytics Framework



[1]*Bas van Bakel, OCS Consulting, DIY: Create your own SDTM mapping framework, PhUSE 2016, Paper CD03*

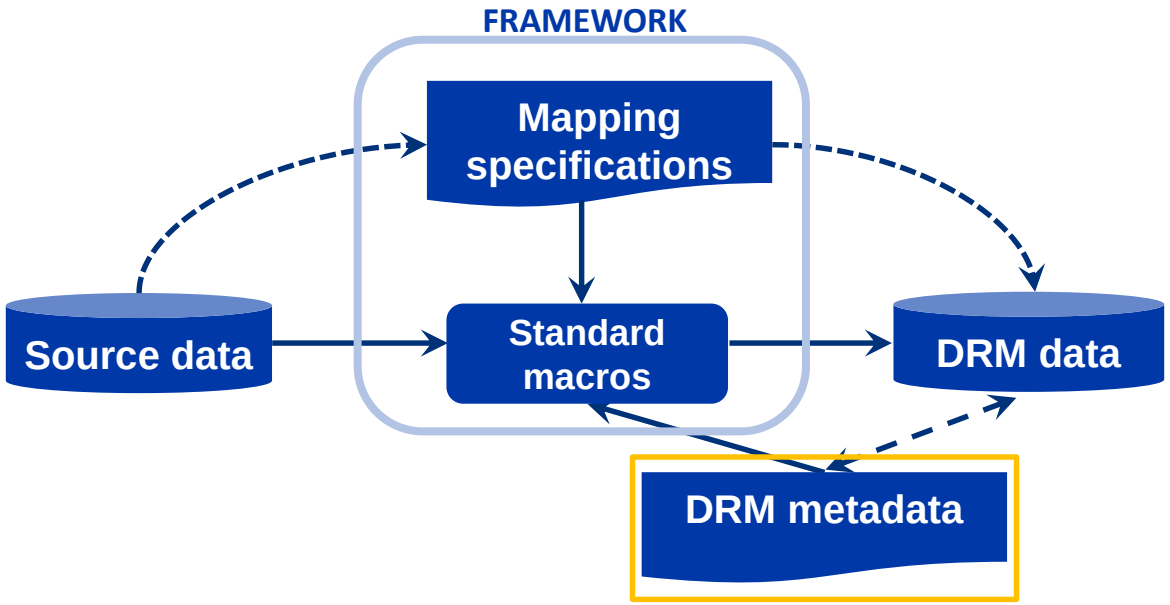
Mapping Specifications

- Excel spreadsheet containing all:
  - Source variables
  - Target DRM variables
  - Mapping specifications
  - Translation of the specifications into SAS (pseudo-)code needed to generate the DRM variables

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;]

Mapping Framework

- Data converted using the mapping framework[1]
- Implemented in Janssen's SAS® Life Science Analytics Framework



[1]*Bas van Bakel, OCS Consulting, DIY: Create your own SDTM mapping framework, PhUSE 2016, Paper CD03*

Target metadata

- Excel spreadsheet defining:
  - 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		

DRM Implementation Process

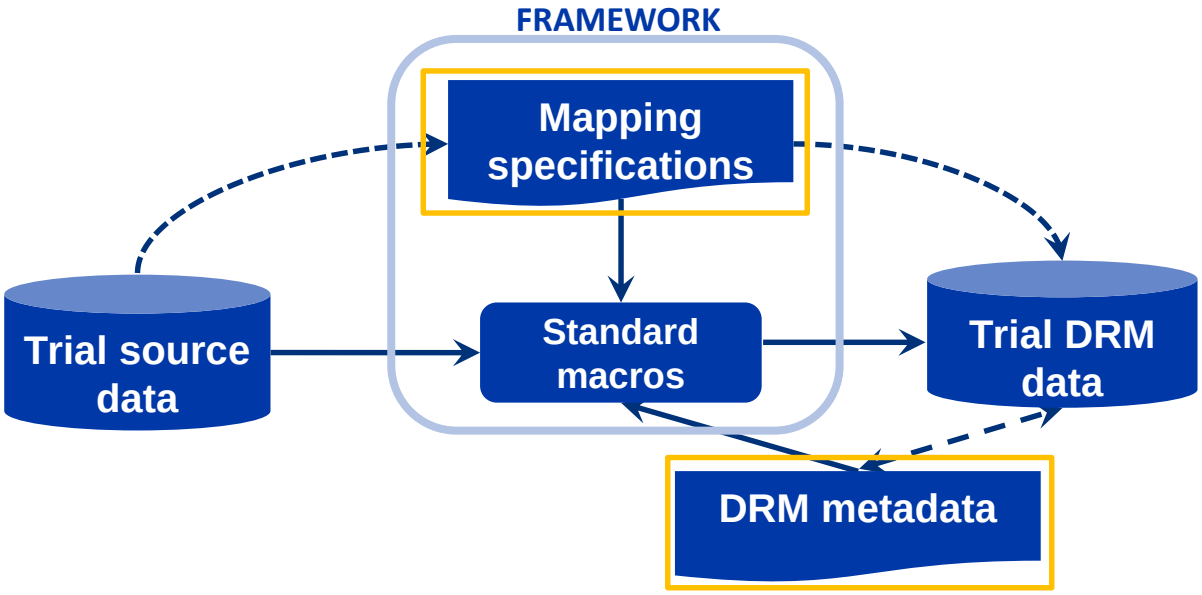
- Create standard/global (re-usable) mappings for 17 DRM domains

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;]

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		

DRM Implementation Process

- Create DRM datasets for 13 trials
  1. According to trial specific source data:
     - Adjust standard mapping specifications
     - Adjust DRM metadata
  2. Create DRM datasets on a daily basis using the mapping framework



DRM Implementation Process

- Create DRM datasets for 13 trials
  1. According to trial specific source data:
     - Adjust standard mapping specifications
     - Adjust DRM metadata
  2. Create DRM datasets on a daily basis using the mapping framework
  3. Monitor daily DRM creations:
     - Data conversion failures could occur:
       - No source data available due to failure of automatic upload of source data
       - Source variables were added or removed
       - Values exceeded the length in the 'target' metadata and were truncated
       - Et cetera

Learnings & Next Steps

- Pilot phase demonstrated business value of the Data Review Model
  - Early access to data for medical review and central monitoring
  - Full data traceability and high data availability

- Planning for a staged roll-out of DRM
  - DM CROs will continue to deliver SDTM data packages in parallel

Learnings & Next Steps

- Based on feedback from the pilot teams:
  - Refining the DRM domain models and business rules
  - Moving enhanced mapping framework to production setting
  - Preparing for a library of mapping rules for Janssen's Data Capture standards to DRM

- Introduction of new utilities to enhance the DRM conversion process
  - Metadata driven setup of trial level mapping sheet
  - Fail-safe mechanism to check incoming source data

Thank you! Questions?

Lieke Gijsbers
OCS Life Sciences
Netherlands

Nele Claes
Janssen R&D
Belgium