cdisc

2023
EUROPE
INTERCHANGE
COPENHAGEN | 26-27 APRIL

Automate harmonization of study metadata for integrated define.xml

Presented by Nynne L. B. Storm & Matthew Phelps, Novo Nordisk A/S

Meet the Speakers

Nynne L. Berthou Storm

Title: Statistical Programming Specialist

Organization: Novo Nordisk A/S

Nynne Storm joined Novo Nordisk A/S in 2018 and has worked as a Statistical Programmer on and across various projects and trials. Currently focusing on pooling of data for submission. She has a PhD from the Technical University of Denmark within the field of Geomagnetism.

Matthew Phelps

Title: Statistical Programming Specialist

Organization: Novo Nordisk A/S

Matthew Phelps joined Novo Nordisk A/S to work on the newly formed Data Science Automation Team as a Statistical Programming Specialist in April 2022. Previously he worked as an epidemiologist at the Danish Heart foundation doing registry-based research. In addition to research, he had a lot of fun programming in R and developing public-facing Shiny apps.

Disclaimer and Disclosures

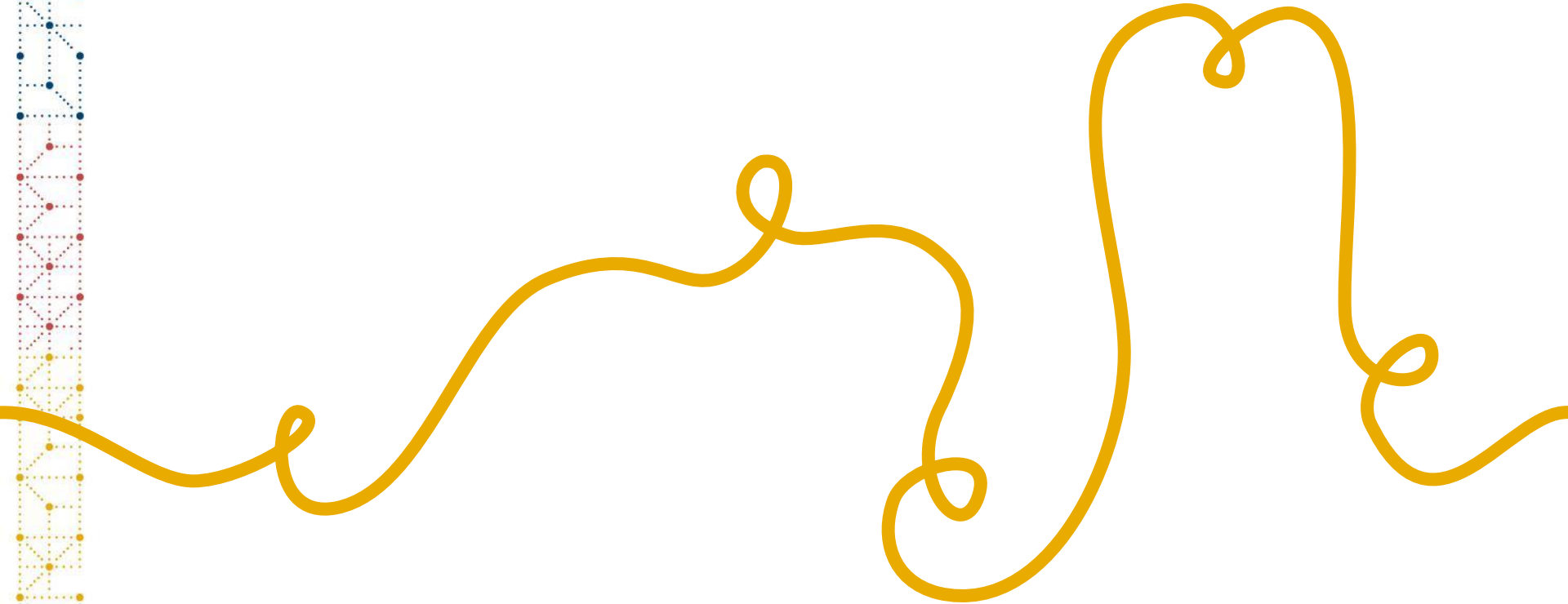
- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC or Novo Nordisk.*

- *The authors have no real or apparent conflicts of interest to report.*

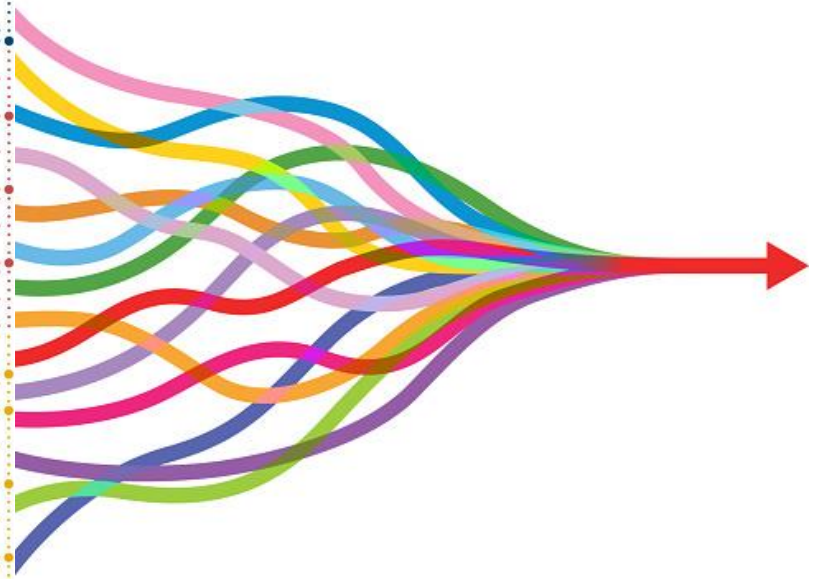
Agenda

1. Introduction
2. Challenges
3. Ambitions overview
4. Zoom in – with features
5. Future

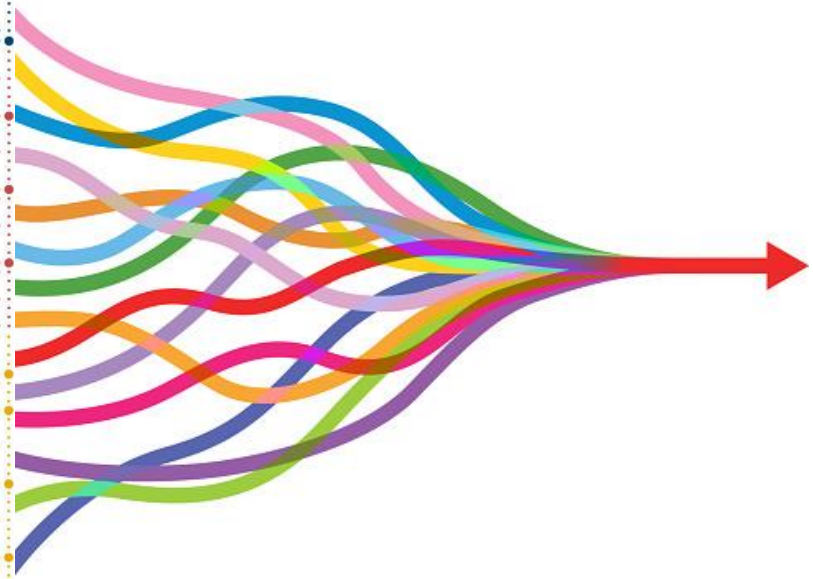
Introduction



Challenges

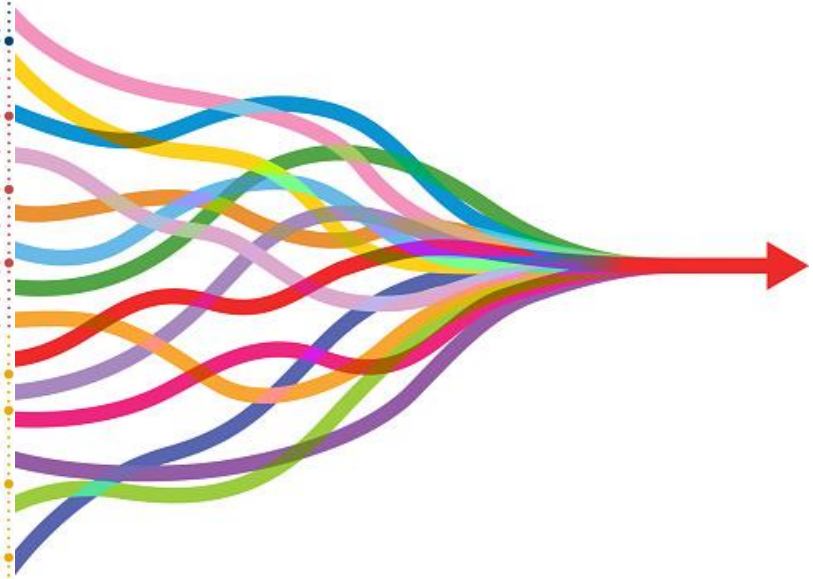


Challenges



- Integrated database for submission
- Safety update database
- Exploratory database
- Etc.

Challenges



- Integrated database for submission
- Safety update database
- Exploratory database
- Etc.

- Time consuming and tedious

Introduction

CST document - containing metadata about ADaM datasets and their variables, including CDISC ADaM IG information.

Dataset	Label	Class			
ADAE	Adverse Events Analysis	Occurence Data Structure	…	…	…
ADSL	Subject-Level Analysis Dataset	Subject Level Analysis Dataset	…	…	…

Introduction

CST document - containing metadata about ADaM datasets and their variables, including CDISC ADaM IG information.

Dataset	Label	Class			
ADAE	Adverse Events Analysis	Occurence Data Structure	…	…	…
ADSL	Subject-Level Analysis Dataset	Subject Level Analysis Dataset	…	…	…

Dataset	Variable	Label	Core	Origin	Description	
ADAE	USUBJID	Unique Subject Identifier	Req	Predecessor	AE.USUBJID	…
ADAE	APERIOD	Period	Perm	Derived	xxx	
. . .		. . .		. . .		

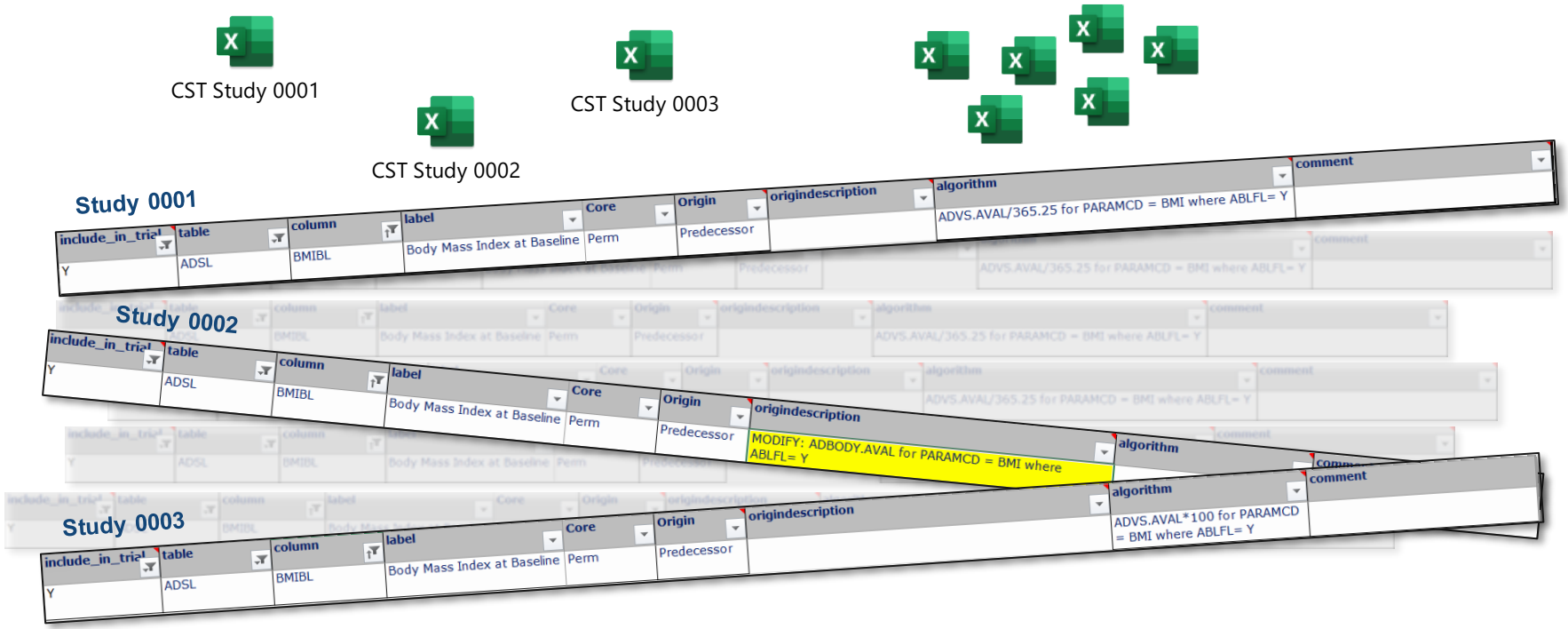
Challenges

CST Study 0001

Study 0001

include_in_trial	table	column	label	Core	Origin	origindescription	algorithm	comment
Y	ADSL	BMIBL	Body Mass Index at Baseline	Perm	Predecessor		ADVS.AVAL/365.25 for PARAMCD = BMI where ABLFL= Y	

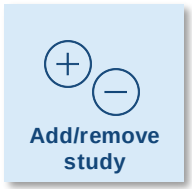
Challenges


CST Study 0001
CST Study 0002
CST Study 0003
Study 0001
include_in_trial
table
column
label
Core
Origin
origindescription
algorithm
comment
Y
ADSL
BMIBL
Body Mass Index at Baseline
Perm
Predecessor
ADVS.AVAL/365.25 for PARAMCD = BMI where ABLFL= Y
Study 0002
MODIFY: ADBODY.AVAL for PARAMCD = BMI where ABLFL= Y
Study 0003
ADVS.AVAL*100 for PARAMCD = BMI where ABLFL= Y

Challenges

- Overview
- Early start-up
- Add/remove study
- Exception

Ambitions



Ambitions

Overview

Early start-up

Add/remove study

Exception

Ambitions

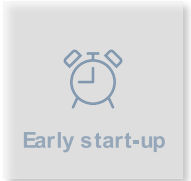
Overview

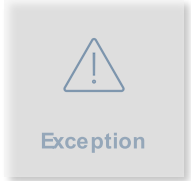
Early start-up

Add/remove study

Exception

Ambitions





Only focuses on variables that are differently derived between studies

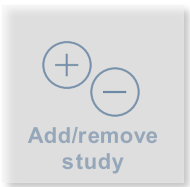
Ambitions

Only focuses on variables that are differently derived between studies

	trial	origindescription	algorithm
ADSL-BMIBL			
1	0001	__blank__	ADVS.AVAL*100 for PARAMCD = BMI where ABLFL= Y
2	0002	__blank__	ADVS.AVAL*100 for PARAMCD = BMI where ABLFL= Y
3	0003	__blank__	ADVS.AVAL / 100 for PARAMCD = BMI where ABLFL= Y

Showing 1 to 3 of 3 entries

Ambitions

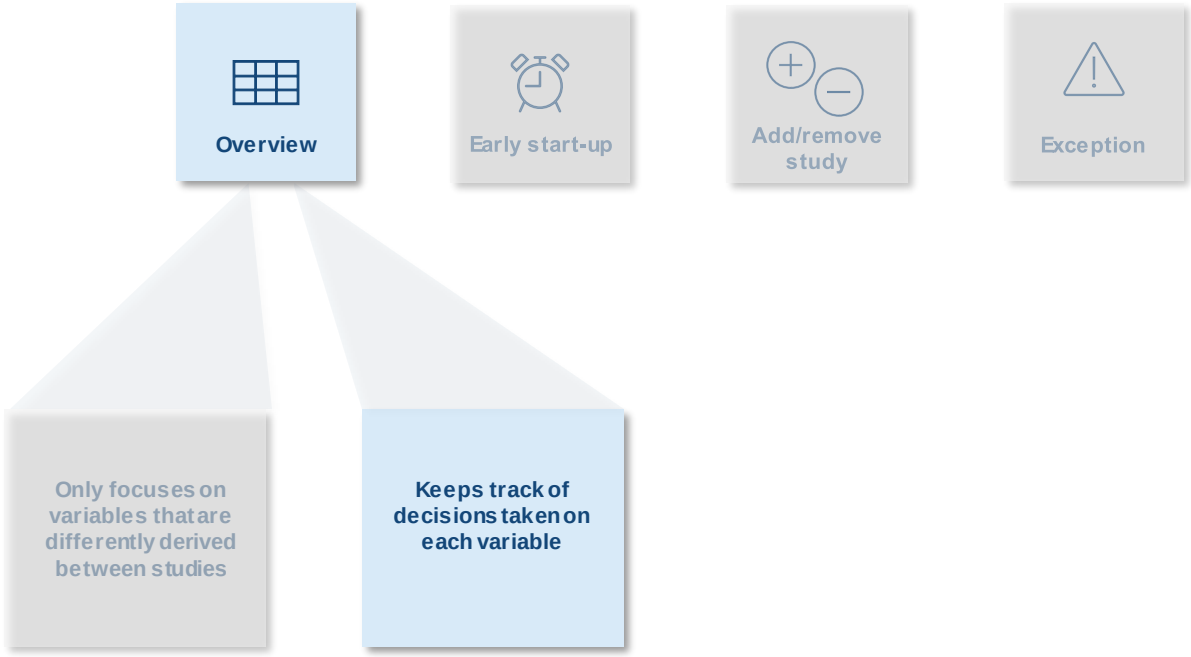


Only focuses on variables that are differently derived between studies

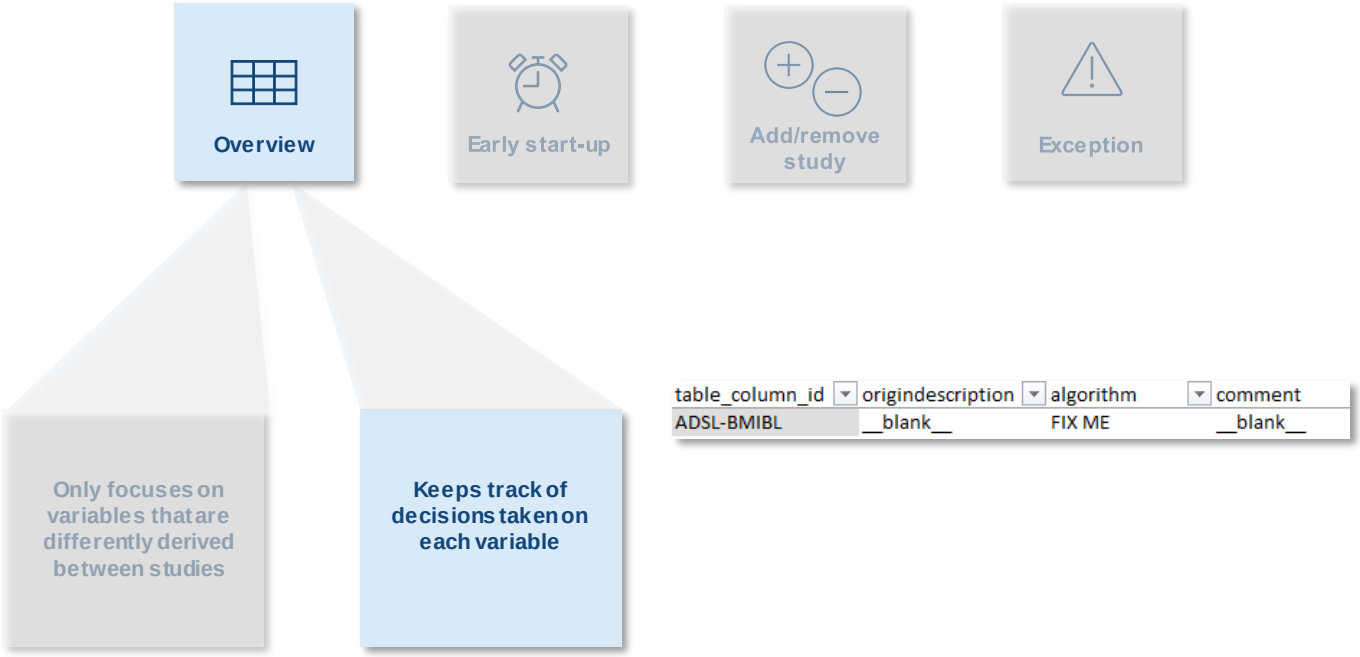
	trial	origindescription	algorithm
ADSL-BMIBL			
1	0001	__blank__	ADVS.AVAL*100 for PARAMCD = BMI where ABLFL= Y
2	0002	__blank__	ADVS.AVAL*100 for PARAMCD = BMI where ABLFL= Y
3	0003	__blank__	ADVS.AVAL / 100 for PARAMCD = BMI where ABLFL= Y

Showing 1 to 3 of 3 entries

Ambitions

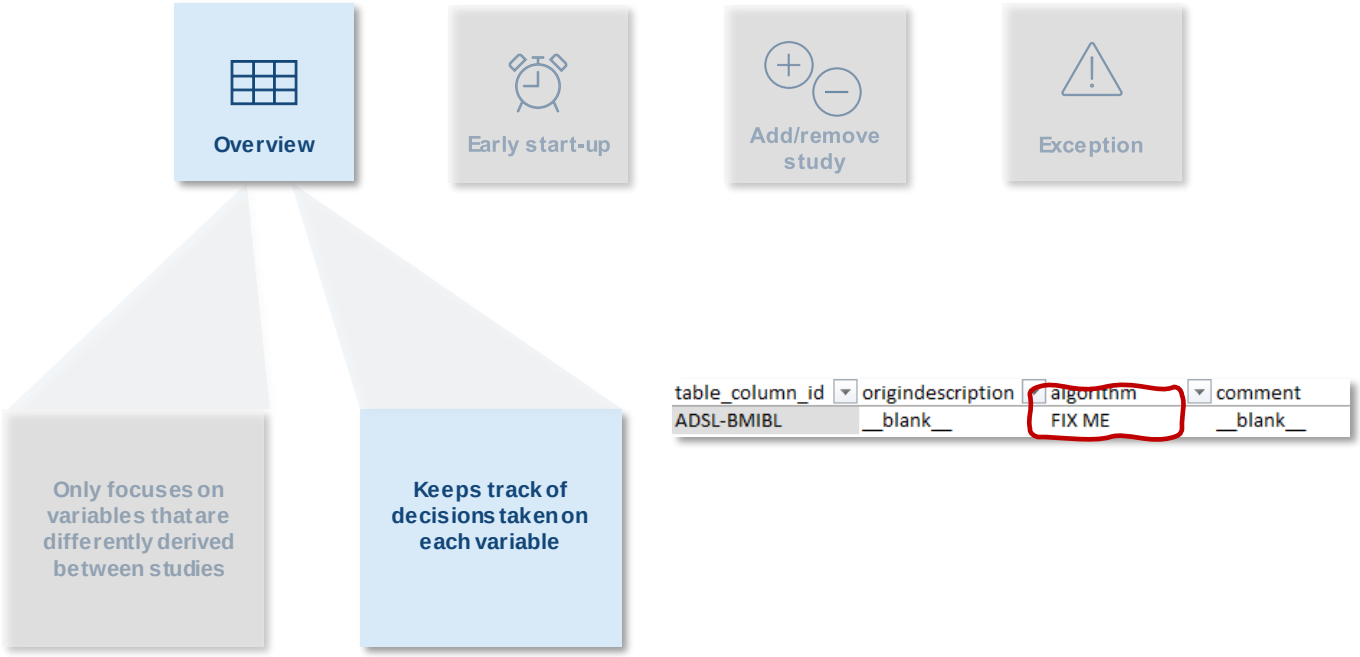

Overview
Early start-up
Add/remove study
Exception
Only focuses on variables that are differently derived between studies
Keeps track of decisions taken on each variable

Ambitions

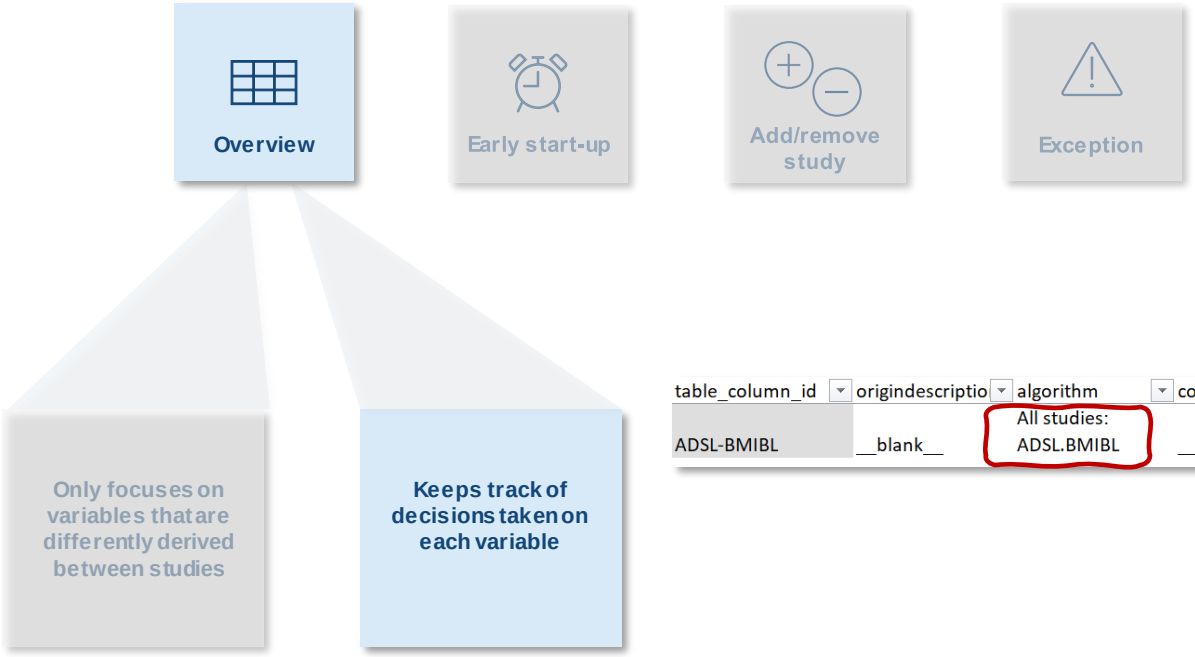


table_column_id	origindescription	algorithm	comment
ADSL-BMIBL	__blank__	FIX ME	__blank__

Ambitions

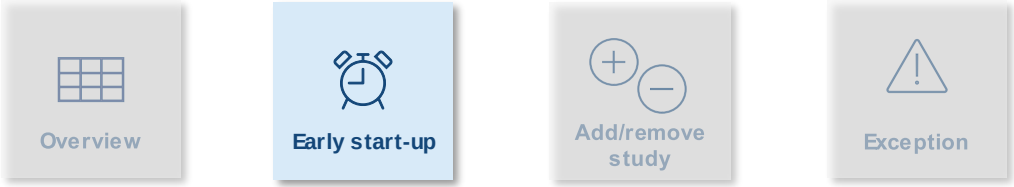

Overview
Early start-up
Add/remove study
Exception
Only focuses on variables that are differently derived between studies
Keeps track of decisions taken on each variable
table_column_id
origindescription
algorithm
comment
ADSL-BMIBL
__blank__
FIX ME
__blank__

Ambitions

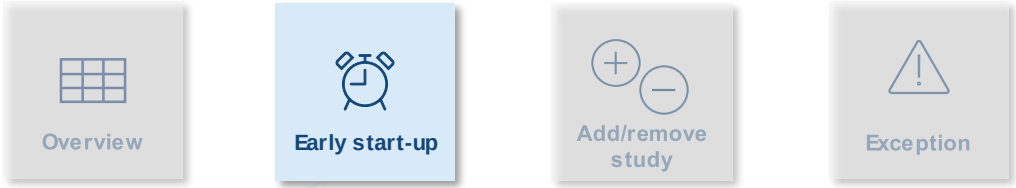


table_column_id	origindescriptio	algorithm	comment
ADSL-BMIBL	__blank__	All studies: ADSL.BMIBL	__blank__

Ambitions

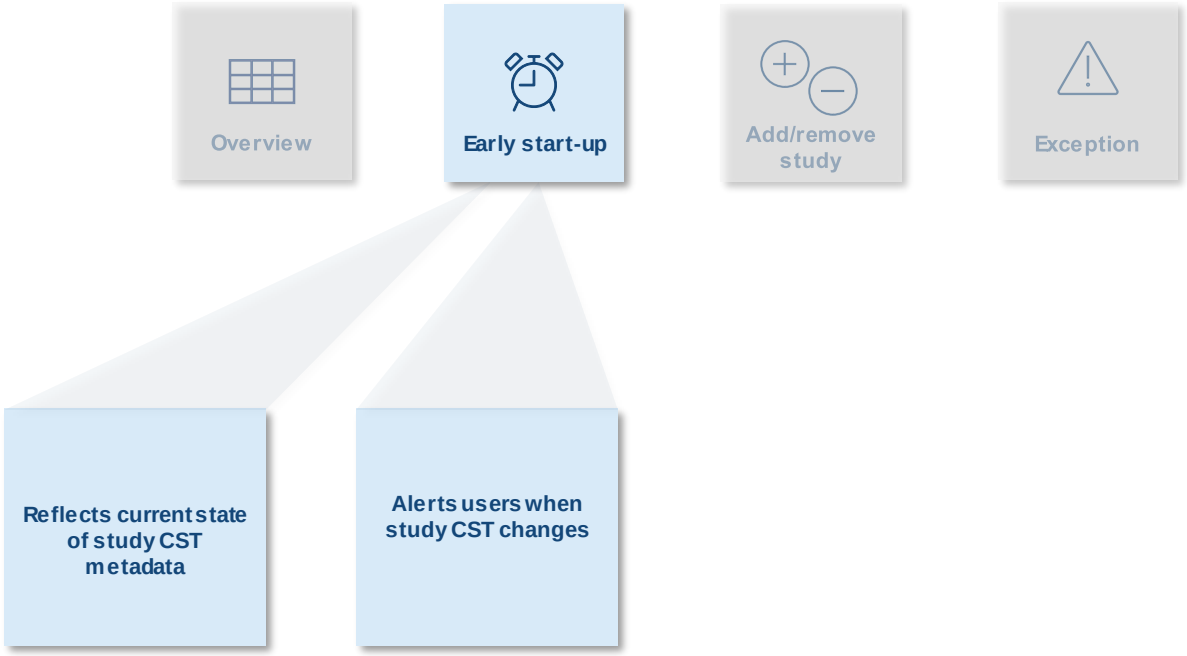


Ambitions

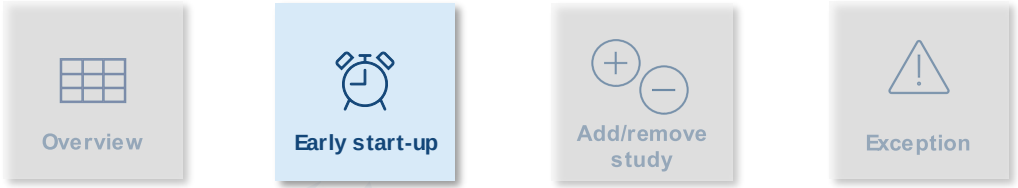


Reflects current state of study CST metadata

Ambitions


Overview
Early start-up
Add/remove study
Exception
Reflects current state of study CST metadata
Alerts users when study CST changes

Ambitions

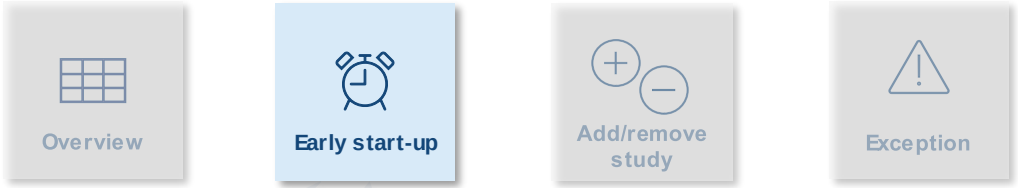


Reflects current state of study CST metadata

Alerts users when study CST changes

	trial	table	column	variable	old_value	new_value	reviewed
1	0001	ADSL	BMIBL	algorithm	ADVS.AVAL*100 for PARAMCD = BMI where ABLFL= Y	ADVS.AVAL * 100 for PARAMCD = BMI where ABLFL= Y	N
2	0001	ADSL	BMIBL	origindescription		none	N
3	0002	ADSL	BMIBL	origindescription		none	N

Ambitions

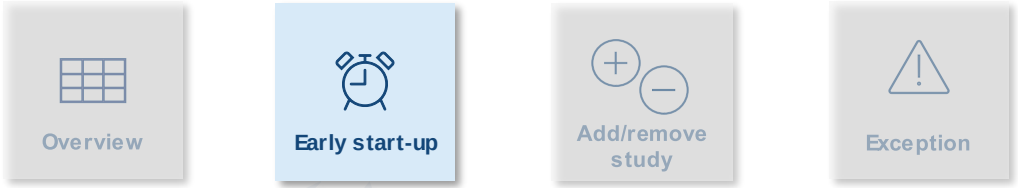


Reflects current state of study CST metadata

Alerts users when study CST changes

	trial	table	column	variable	old_value	new_value	reviewed
1	0001	ADSL	BMIBL	algorithm	ADVS.AVAL*100 for PARAMCD = BMI where ABLFL= Y	ADVS.AVAL * 100 for PARAMCD = BMI where ABLFL= Y	N
2	0001	ADSL	BMIBL	origindescription		none	N
3	0002	ADSL	BMIBL	origindescription		none	N

Ambitions

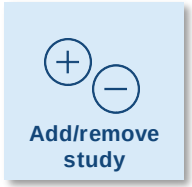


Reflects current state of study CST metadata

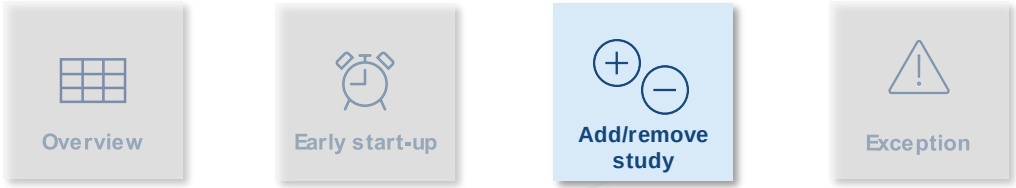
Alerts users when study CST changes

	trial	table	column	variable	old_value	new_value	reviewed
1	0001	ADSL	BMIBL	algorithm	ADVS.AVAL*100 for PARAMCD = BMI where ABLFL= Y	ADVS.AVAL * 100 for PARAMCD = BMI where ABLFL= Y	N
2	0001	ADSL	BMIBL	origindescription		none	N
3	0002	ADSL	BMIBL	origindescription		none	N

Ambitions

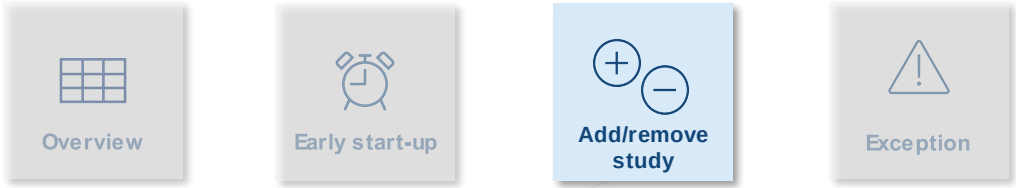


Ambitions



New studies can easily be added

Ambitions



New studies can easily be added

trial	instance	stack
0001	current	Y
0002	csr_er	Y
0003	current	Y
0004	current	N
0005	current	N

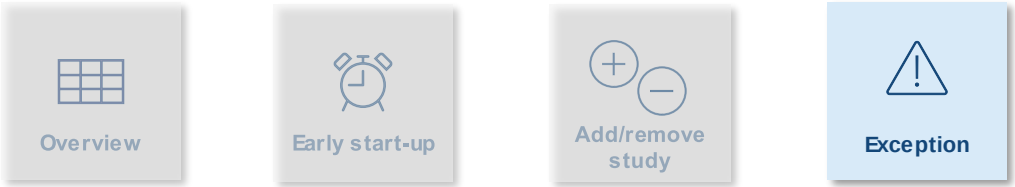
table	label	class		t_0001	t_0002	t_0003	t_0004	t_0005
ADAE	Adverse Events Analysis	OCCURRENCE DATA STRUCTURE	...	Y	Y	Y	N (auto-ge	N (auto-generated)
ADSL	Subject-Level Analysis Dataset	SUBJECT LEVEL ANALYSIS DATASET	...	Y	Y	Y	N (auto-ge	N (auto-generated)

Ambitions

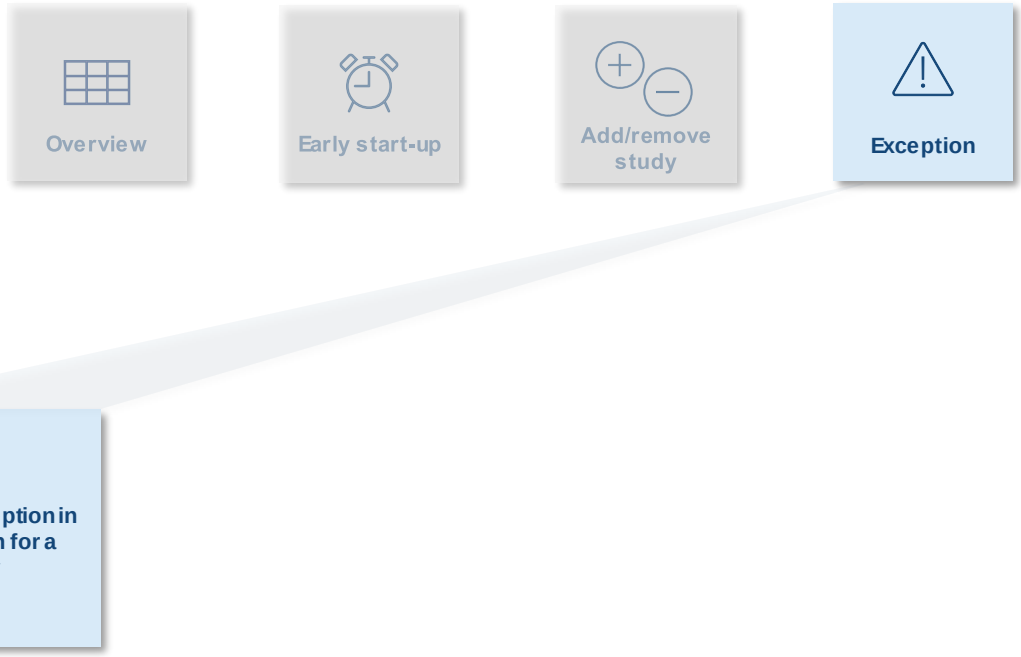
trial	instance	stack
0001	current	Y
0002	csr_er	Y
0003	current	Y
0004	current	N
0005	current	N

table	label	class		t_0001	t_0002	t_0003	t_0004	t_0005	
ADAE	Adverse Events Analysis	OCCURRENCE DATA STRUCTURE	...	Y	Y	Y	N (auto-ge	N (auto-generated)	
ADSL	Subject-Level Analysis Dataset	SUBJECT LEVEL ANALYSIS DATASET	...	Y	Y	Y	N (auto-ge	N (auto-generated)	

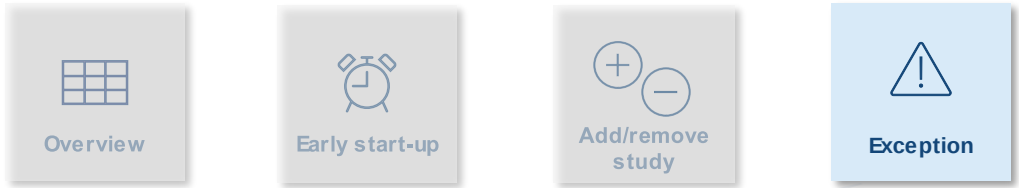
Ambitions



Ambitions


Overview
Early start-up
Add/remove study
Exception
Specific exception in description for a study

Ambitions

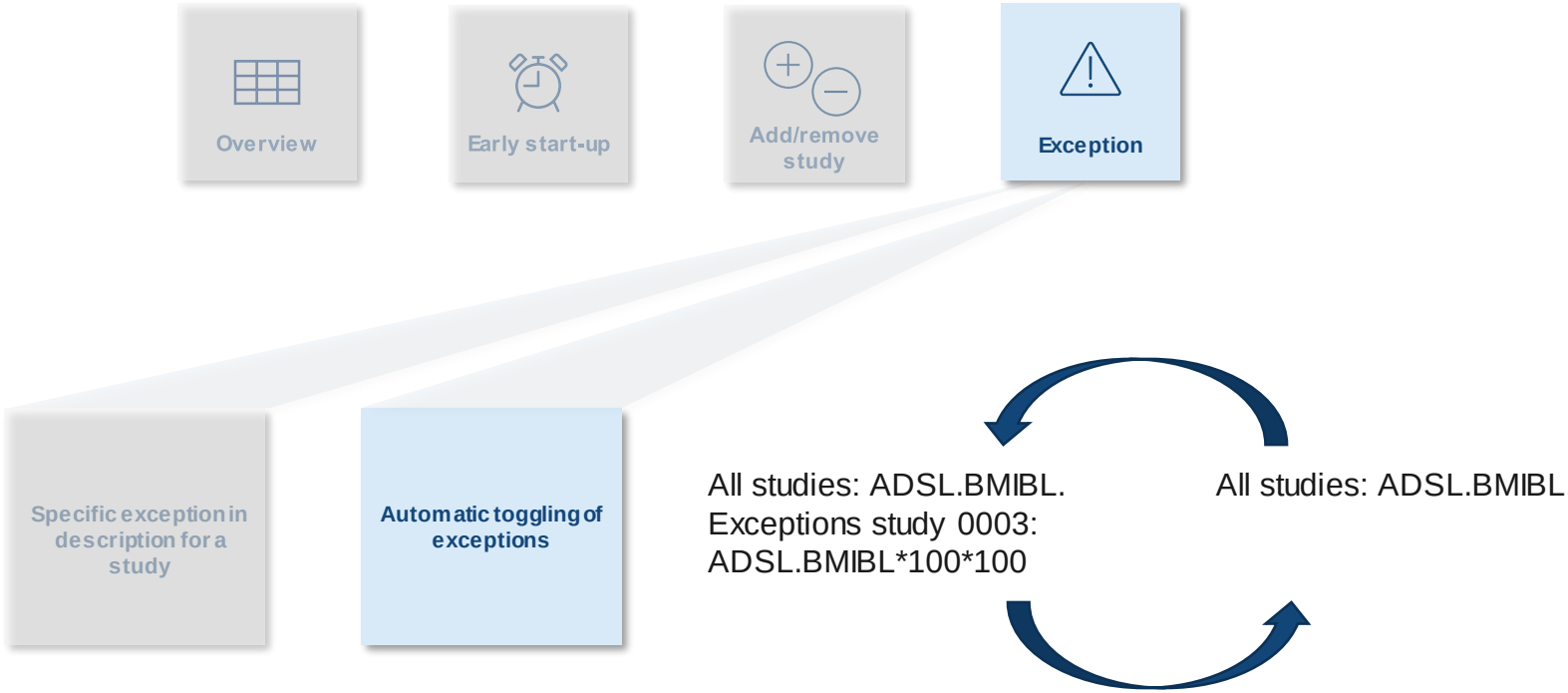


Specific exception in description for a study

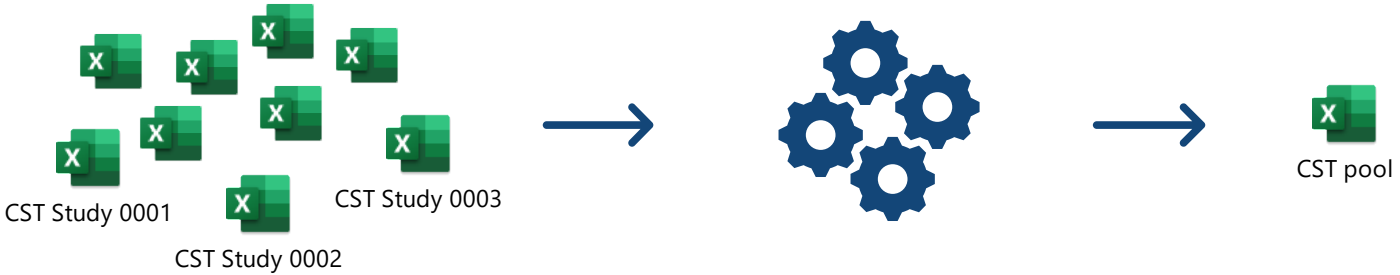
table_column_id	origindescriptio	algorithm	comment	exceptions	stacked_or_pooled
ADSL-BMIBL	__blank__	All studies: ADSL.BMIBL	__blank__	{0003}{ADSL.BMIBL*100*100}	stacked

All studies: ADSL.BMIBL.
Exceptions study 0003: ADSL.BMIBL*100*100

Ambitions

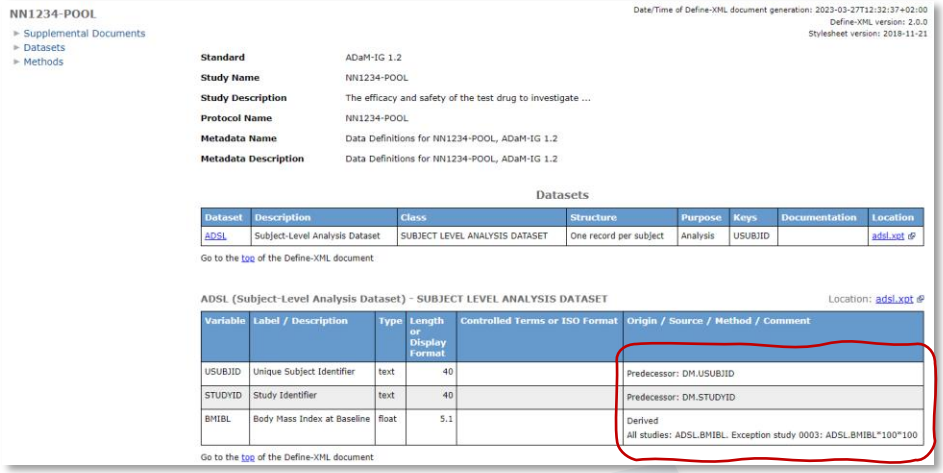

Overview
Early start-up
Add/remove study
Exception
Specific exception in description for a study
Automatic toggling of exceptions
All studies: ADSL.BMIBL.
Exceptions study 0003: ADSL.BMIBL*100*100
All studies: ADSL.BMIBL.

Output



Output

- CST file with the descriptions for the pooled database – input for define.xml



NN1234-POOL

- Supplemental Documents
- Datasets
- Methods

Date/Time of Define-XML document generation: 2023-03-27T12:32:37+02:00
Define-XML version: 2.0.0
Stylesheet version: 2018-11-21

Standard	ADaM-IG 1.2
Study Name	NN1234-POOL
Study Description	The efficacy and safety of the test drug to investigate ...
Protocol Name	NN1234-POOL
Metadata Name	Data Definitions for NN1234-POOL, ADaM-IG 1.2
Metadata Description	Data Definitions for NN1234-POOL, ADaM-IG 1.2

Datasets

Dataset	Description	Class	Structure	Purpose	Keys	Documentation	Location
ADSL	Subject-Level Analysis Dataset	SUBJECT LEVEL ANALYSIS DATASET	One record per subject	Analysis	USUBJID		adsl.xpt

Go to the top of the Define-XML document

ADSL (Subject-Level Analysis Dataset) - SUBJECT LEVEL ANALYSIS DATASET — Location: adsl.xpt

Variable	Label / Description	Type	Length or Display Format	Controlled Terms or ISO Format	Origin / Source / Method / Comment
USUBJID	Unique Subject Identifier	text	40		Predecessor: DM.USUBJID
STUDYID	Study Identifier	text	40		Predecessor: DM.STUDYID
BMIBL	Body Mass Index at Baseline	float	5.1		Derived All studies: ADSL.BMIBL. Exception study 0003: ADSL.BMIBL*100*100

Go to the top of the Define-XML document

Origin / Source / Method / Comment
Predecessor: DM.USUBJID
Predecessor: DM.STUDYID
Derived All studies: ADSL.BMIBL. Exception study 0003: ADSL.BMIBL*100*100

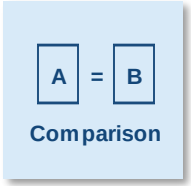
Output

- CST file with the descriptions for the pooled database – input for define.xml

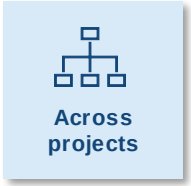
- Stacked study data – ready for data wrangling for final ADaM pool

Future

- Compare variables that are the same, but are named differently in the studies



- Compare studies across projects



- “Smart” comparison of description text

Learnings

- Only doable due to standardization (CDISC & Novo Nordisk)
- Work iteratively
- Starting early **>** starting later

Thank You!