Unraveling the Complex Web of Data Relationships

Presented by Emma Törner, Statistical Programming Director, CVRM Biometrics, AstraZeneca
& Parag Wani, Senior Director Statistical Programming, CVRM Biometrics, AstraZeneca

Meet the Speakers

Emma Törner

Title: Statistical Programming Director

Organization: AstraZeneca

Emma Törner is currently a Statistical Programming Director at AstraZeneca, leading drug project development in a key therapeutic area as a portfolio and strategy lead. Emma has a background in Engineering Mathematics, Medical modelling and joined the industry and AstraZeneca in 2017. She has led multiple clinical studies and drug projects as a statistical programmer and has been involved in regulatory submissions. Emma has previously held the role of a lead for therapy area standards governance, development, and training. She has a key interest and expertise in Cardiovascular CDISC standards and foundational SDTM/ADaM standards.

Parag Wani

Title: Senior Director Statistical Programming

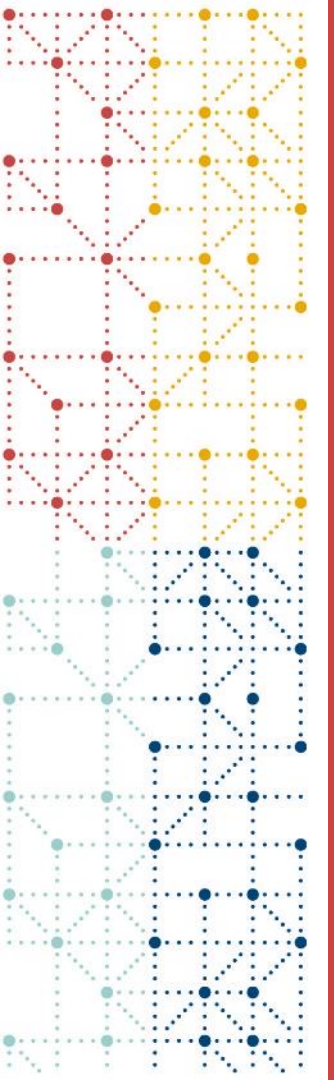
Organization: AstraZeneca

Parag Wani has more than 21 years of experience in statistical programming in the drug development domain with diverse roles as a contributor, SME and leader across major pharmaceuticals and CROs. Parag is currently a Senior Director for Statistical Programming in AstraZeneca within a key therapeutic area and heads up a global team of managers and programmers. With a background in Computer Science, he has always had a strong technical and analytical aptitude and has a keen interest in driving and contributing to strategies across clinical drug development value chain.

Disclaimer and Disclosures

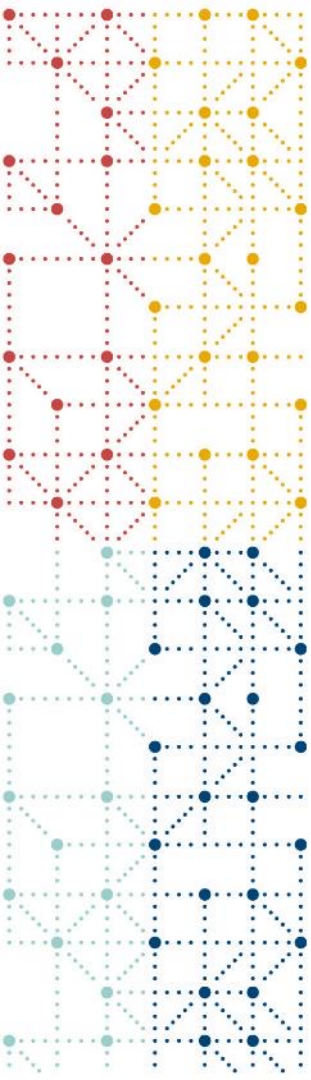
- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC, AstraZeneca or CVRM Biometrics.*

- *The author(s) have no real or apparent conflicts of interest to report.*



Agenda

1. CDISC RELREC Introduction
2. Sponsor (AstraZeneca) Data Standards
3. RELREC Linking Variables Approach
    1. Standardizing Linking Variables
    2. Standardizing Linking Variable Values
4. Automation and Testing Strategy



CDISC RELREC Introduction

RELREC Dataset

The Related Records (RELREC) special purpose dataset

- Used to describe relationships between records for a subject within or across domains, and relationships of records between datasets
- Collected relationships either by
  - Explicit references
  - Information that necessitates using multiple datasets

STUDYID *Study Identifier*	**RDOMAIN** *Related Domain Abbreviation*	**USUBJID** *Unique Subject Identifier*	**IDVAR** *Identifying Variable*	**IDVARVAL** *Identifying Variable Value*	**RELTYPE** *Relationship Type*	**RELID** *Relationship Identifier*
STUDY123	AE	STUDY123/SUBJ01	AESEQ	5		AECM1
STUDY123	CM	STUDY123/SUBJ01	CMSEQ	11		AECM1
STUDY123	CM	STUDY123/SUBJ01	CMSEQ	12		AECM1
STUDY123	AE	STUDY123/SUBJ01	AESEQ	4		AEEC1
STUDY123	EC	STUDY123/SUBJ01	ECSEQ	3		AEEC1
STUDY123	CE		CELNKID		[ONE/MANY]	CEFA
STUDY123	FA		FALNKID		[ONE/MANY]	CEFA
STUDY123	AE		AESPID		[ONE/MANY]	AECE
STUDY123	CE		CELNKGRP		[ONE/MANY]	AECE

RELREC Dataset

The Related Records (RELREC) special purpose dataset

1

- Used to describe relationships between records for a subject within or across domains, and relationships of records between datasets
- Collected relationships either by
  - Explicit references
  - Information that necessitates using multiple datasets

STUDYID *Study Identifier*	RDOMAIN *Related Domain Abbreviation*	USUBJID *Unique Subject Identifier*	IDVAR *Identifying Variable*	IDVARVAL *Identifying Variable Value*	RELTYPE *Relationship Type*	RELID *Relationship Identifier*
STUDY123	AE	STUDY123/SUBJ01	AESEQ	5		AECM1
STUDY123	CM	STUDY123/SUBJ01	CMSEQ	11		AECM1
STUDY123	CM	STUDY123/SUBJ01	CMSEQ	12		AECM1
STUDY123	AE	STUDY123/SUBJ01	AESEQ	4		AEEC1
STUDY123	EC	STUDY123/SUBJ01	ECSEQ	3		AEEC1
STUDY123	CE		CELNKID		[ONE/MANY]	CEFA
STUDY123	FA		FALNKID		[ONE/MANY]	CEFA
STUDY123	AE		AESPID		[ONE/MANY]	AECE
STUDY123	CE		CELNKGRP		[ONE/MANY]	AECE

RELREC Dataset

The Related Records (RELREC) special purpose dataset

- Used to describe relationships between records for a subject within or across domains, and relationships of records between datasets
- Collected relationships either by
  - Explicit references
  - Information that necessitates using multiple datasets

2

STUDYID *Study Identifier*	**RDOMAIN** *Related Domain Abbreviation*	**USUBJID** *Unique Subject Identifier*	**IDVAR** *Identifying Variable*	**IDVARVAL** *Identifying Variable Value*	**RELTYPE** *Relationship Type*	**RELID** *Relationship Identifier*
STUDY123	AE	STUDY123/SUBJ01	AESEQ	5		AECM1
STUDY123	CM	STUDY123/SUBJ01	CMSEQ	11		AECM1
STUDY123	CM	STUDY123/SUBJ01	CMSEQ	12		AECM1
STUDY123	AE	STUDY123/SUBJ01	AESEQ	4		AEEC1
STUDY123	EC	STUDY123/SUBJ01	ECSEQ	3		AEEC1
STUDY123	CE		CELNKID		[ONE/MANY]	CEFA
STUDY123	FA		FALNKID		[ONE/MANY]	CEFA
STUDY123	AE		AESPID		[ONE/MANY]	AECE
STUDY123	CE		CELNKGRP		[ONE/MANY]	AECE

RELREC Dataset
Dataset Relationships – Linking Variables

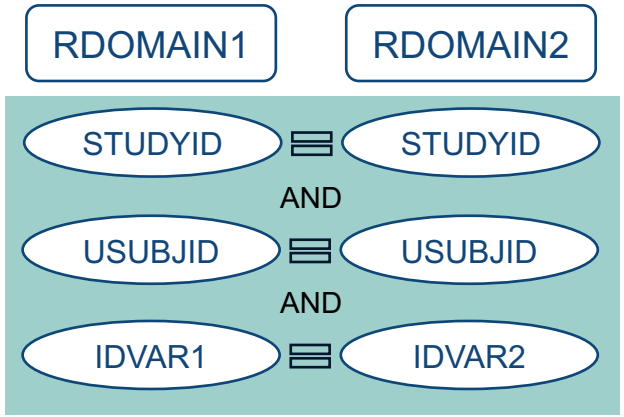
SDTM Variable	Purpose	Scope of values
--CAT/SCAT	For grouping records within one or more domains/CRFs. --SCAT requires --CAT	Values have a meaning across subjects and domains
--GRPID	For grouping records within a subject within a domain	Values have a meaning within the domain and subject
--SPID	Sponsor defined identifier	Value can be defined during collection to capture relationships across domains
--REFID	Optional internal or external identifier, for grouping records within a domain	Values can group multiple results of a test in the same domain or link procedures to results
--LNKID	Identifier used to link related records within or across domain in a one-to-one or one-to-many relationship	Values have a meaning within subjects across domains if used in RELREC
--LNKGRP	Identifier used for grouping records within or across domains, usually in a many-to-one relationship	Values have a meaning within subjects across domains if used in RELREC

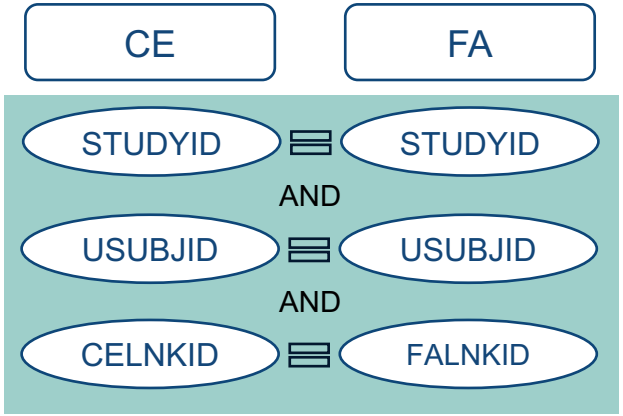
RELREC Dataset
Dataset Relationships – Linking Variables

SDTM Variable	Purpose	Scope of values
--CAT/SCAT	For grouping records within one or more domains/CRFs. --SCAT requires --CAT	Values have a meaning across subjects and domains
--GRPID	For grouping records within a subject within a domain	Values have a meaning within the domain and subject
--SPID	S	
--REFID	C	
--LNKID	I o	s
--LNKGRP	I	s
	to-one relationship	if used in RELREC

STUDYID *Study Identifier*	RDOMAIN *Related Domain Abbreviation*	USUBJID *Unique Subject Identifier*	IDVAR *Identifying Variable*	IDVARVAL *Identifying Variable Value*	RELTYPE *Relationship Type*	RELID *Relationship Identifier*
STUDY123	CE		CELNKID		[ONE/MANY]	CEFA
STUDY123	FA		FALNKID		[ONE/MANY]	CEFA
STUDY123	AE		AESPID		[ONE/MANY]	AECE
STUDY123	CE		CELNKGRP		[ONE/MANY]	AECE

RELREC Dataset
Dataset Relationships – Linking Variables





STUDYID *Study Identifier*	**RDOMAIN** *Related Domain Abbreviation*	**USUBJID** *Unique Subject Identifier*	**IDVAR** *Identifying Variable*	**IDVARVAL** *Identifying Variable Value*	**RELTYPE** *Relationship Type*	**RELID** *Relationship Identifier*
STUDY123	CE		CELNKID		[ONE/MANY]	CEFA
STUDY123	FA		FALNKID		[ONE/MANY]	CEFA

AstraZeneca Data Standards
RELREC Standardization

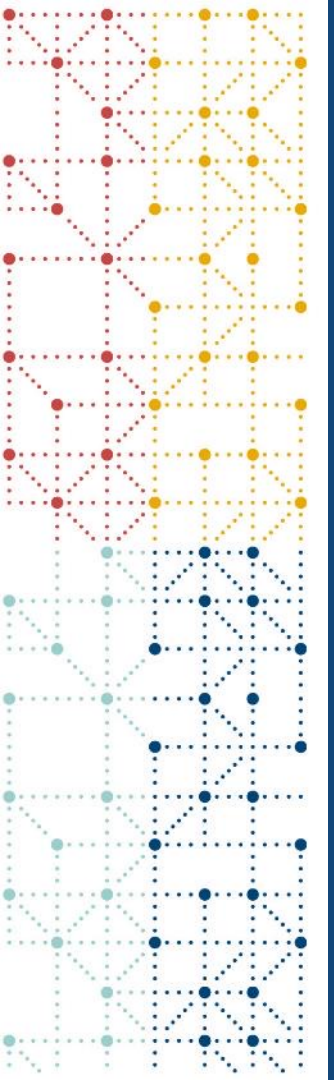
Simple rule, but complex requirements and implications:

- *The same variable needs to be selected for linking for all raw datasets utilizing a dataset relationship*
- *Regardless of the source raw datasets, the linking variable chosen as IDVAR needs to link the correct related records across the SDTM datasets*
- *Assigning new relationships, and adding to existing ones, needs to consider the'full picture'*

Complexities & Considerations

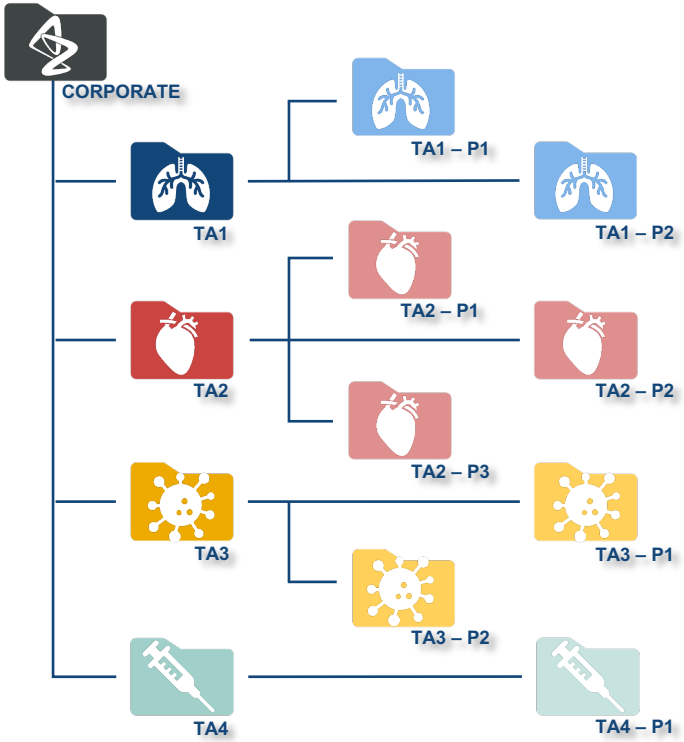
Standardizing the linking variables (RELREC.IDVAR) for each dataset relationship

Assigning values to linking variables ensuring uniqueness



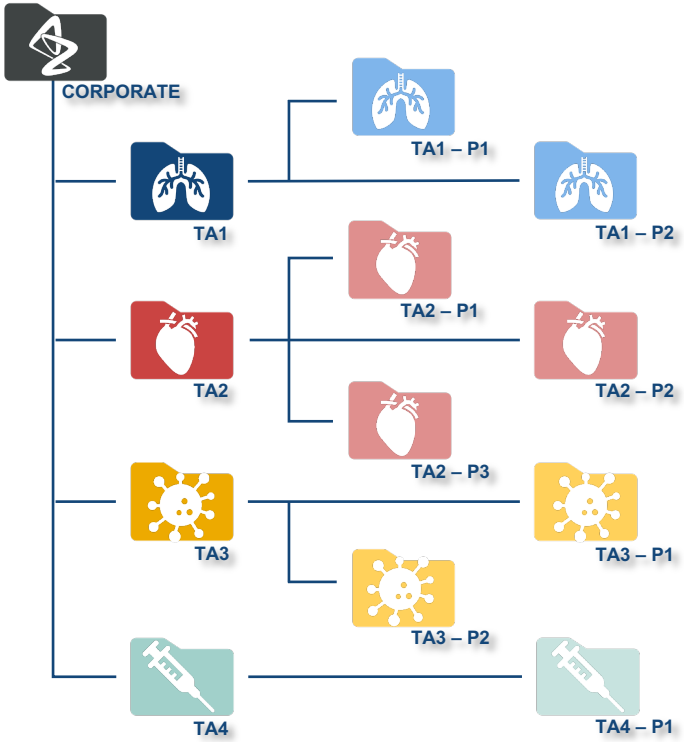
Sponsor (AstraZeneca) Data Standards
RELREC standardization

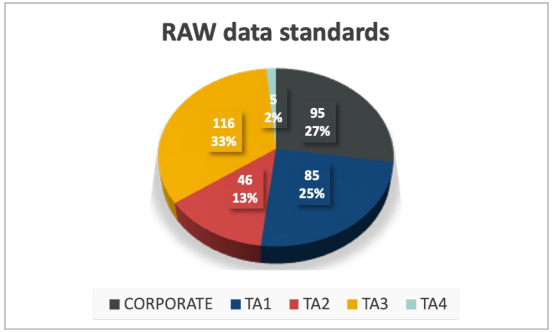
AstraZeneca Data Standards

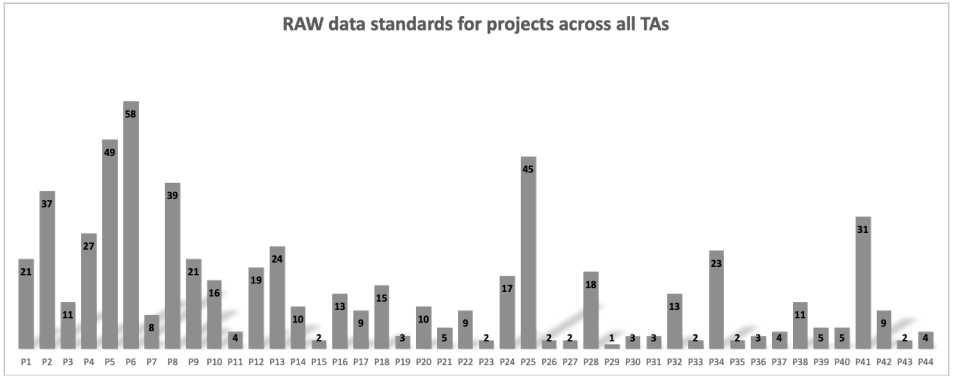


TAx – Therapeutic Area x Pn – Project n

AstraZeneca Data Standards

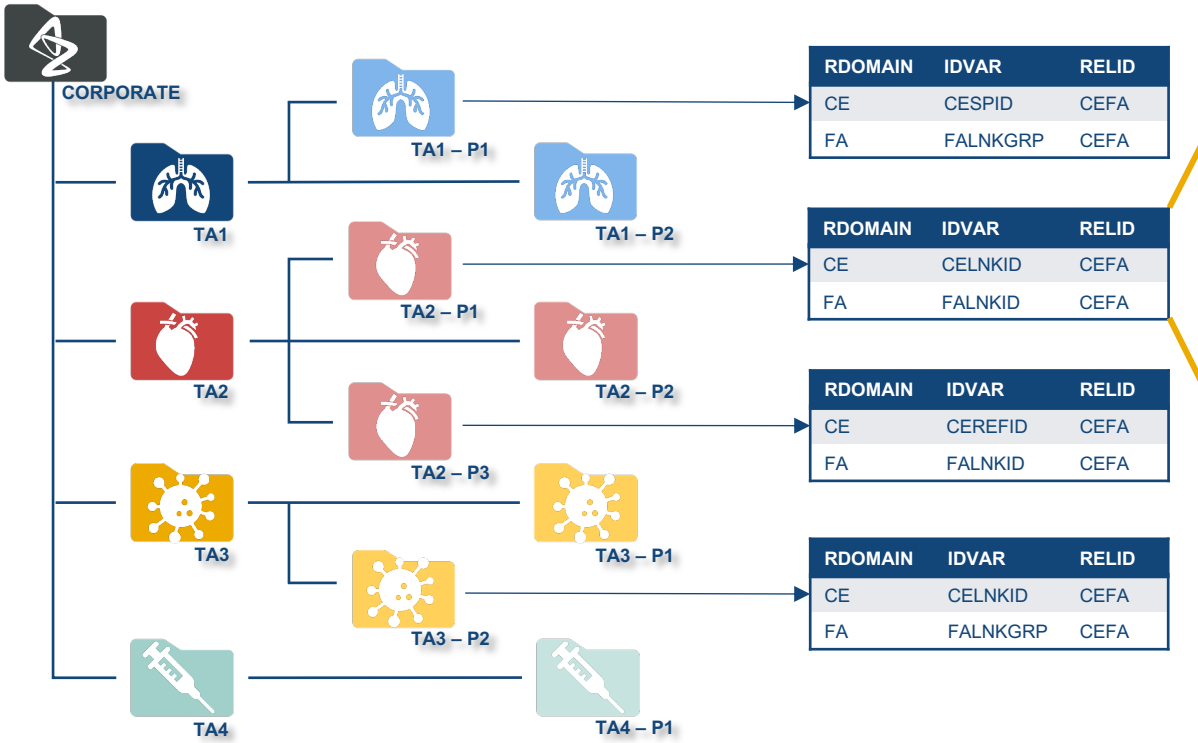






TAx – Therapeutic Area x Pn – Project n

AstraZeneca Data Standards



RDOMAIN	IDVAR	RELID
CE	CESPID	CEFA
FA	FALNKGRP	CEFA

RDOMAIN	IDVAR	RELID
CE	CELNKID	CEFA
FA	FALNKID	CEFA

RDOMAIN	IDVAR	RELID
CE	CEREFID	CEFA
FA	FALNKID	CEFA

RDOMAIN	IDVAR	RELID
CE	CELNKID	CEFA
FA	FALNKGRP	CEFA

Cerebrovascular Event

RDOMAIN	IDVAR	RELID
CE	CELNKID	CEFA
FA	FALNKID	CEFA

Heart Failure Event

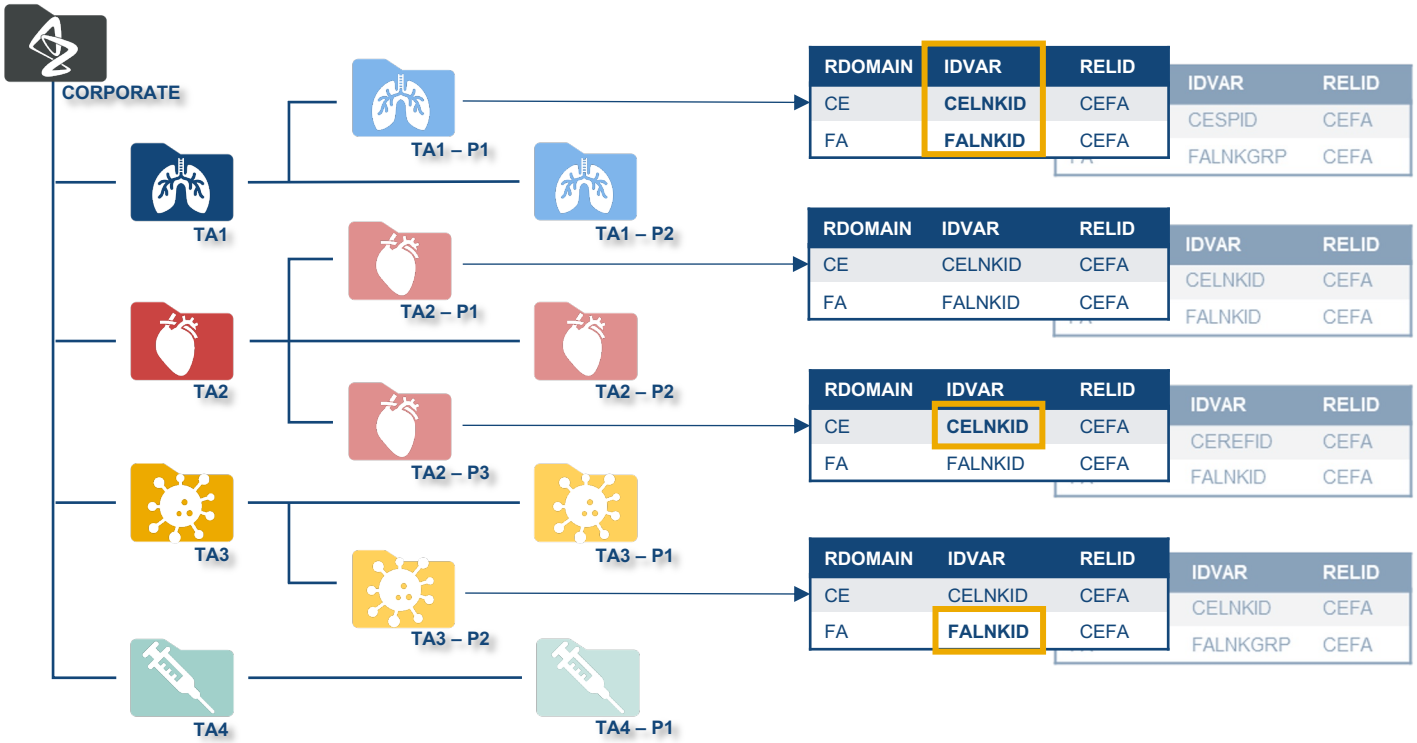
RDOMAIN	IDVAR	RELID
CE	CELNKID	CEFA
FA	FALNKID	CEFA

Hypoglycemic Event

RDOMAIN	IDVAR	RELID
CE	CELNKID	CEFA
FA	FALNKID	CEFA

TAx – Therapeutic Area x Pn – Project n

AstraZeneca Data Standards



TAx – Therapeutic Area x Pn – Project n

RELREC Dataset
Dataset Relationships – Linking Variables

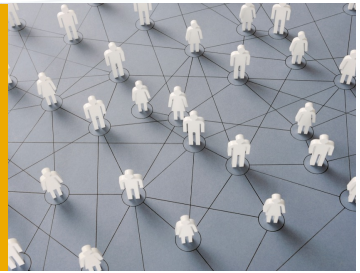
Simple rule, but complex requirements and implications:

- *The same variable needs to be selected for linking for all raw datasets utilizing a dataset relationship*
- *Regardless of the source raw datasets, the linking variable chosen as IDVAR needs to link the correct related records across the SDTM datasets*
- *Assigning new relationships, and adding to existing ones, needs to consider the'full picture'*

What values are appropriate to ensure unique links?

What are the shortcomings of designing linking variables based on each RAW dataset as a standalone entity?

How can a sponsor define and maintain relationships that function across studies?



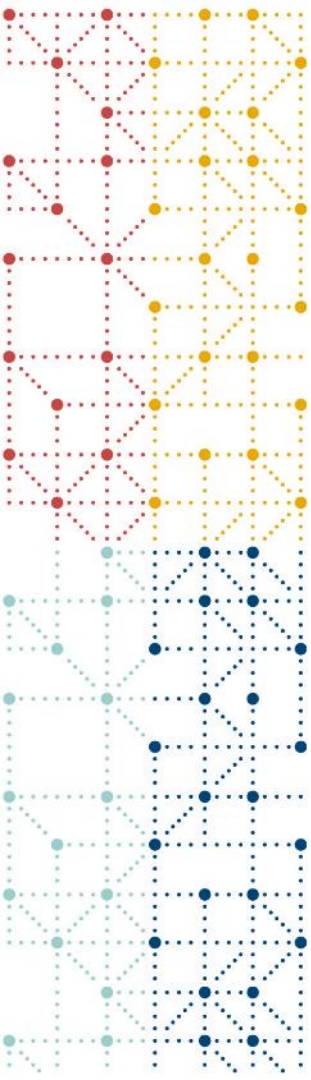
RELREC Linking Variables Approach

Standardizing linking variables (RELREC.IDVAR)

Standardizing Linking Variables

- RELREC IDVAR is defined as part of the SDTM standard
- For each dataset relationship – the raw dataset(s) utilizing a linking relationship (RELID) is specified
- Layer of complexity:
  Changing the linking variable for a relationship require an update to all raw data standards utilizing that relationship

Raw dataset (AZ standard)	RDOMAIN	IDVAR	RELID
CEREVENT HFEVT HYPOGLYC	CE	CELNKID	CEFA
CEREVENT HFEVT HYPOGLYC	FA	FALNKID	CEFA
AE CEREVENT HFEVT HYPOGLYC	AE	AESPID	AECE
AE CEREVENT HFEVT HYPOGLYC	CE	CELNKGRP	AECE
CLAB LBAESI	LB	LBGRPID	LBLC
CLAB LBAESI	LC	LCGRPID	LBLC



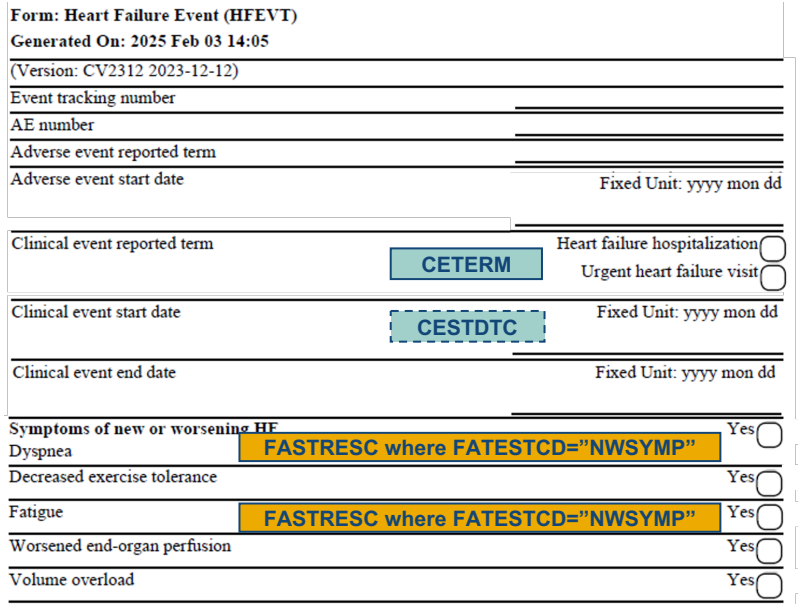
RELREC Linking Variables Approach

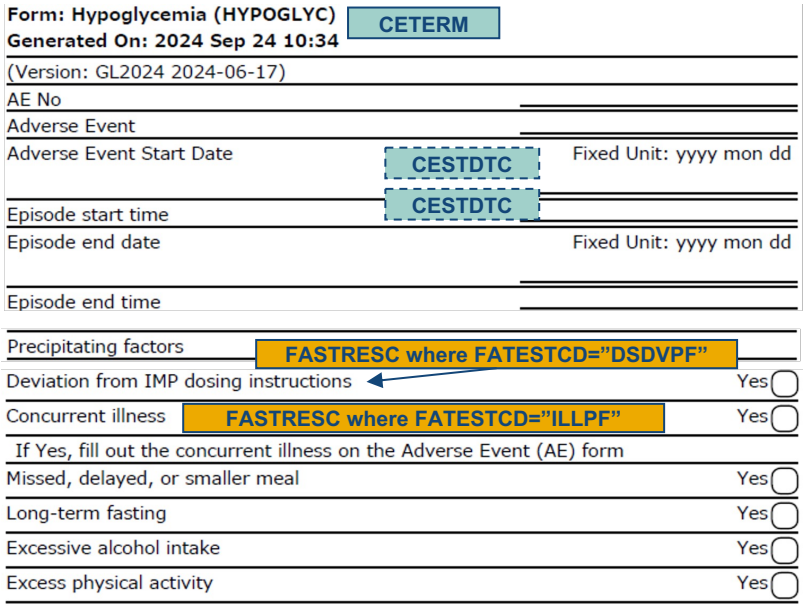
Standardizing linking variable values

CE = Clinical Events

FA = Findings About Events or Interventions

Standardizing Linking Variable Values

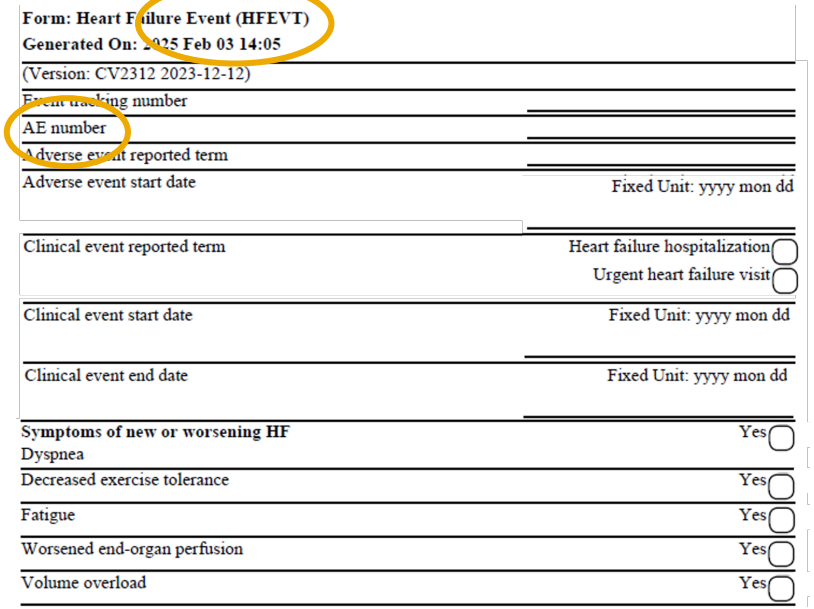




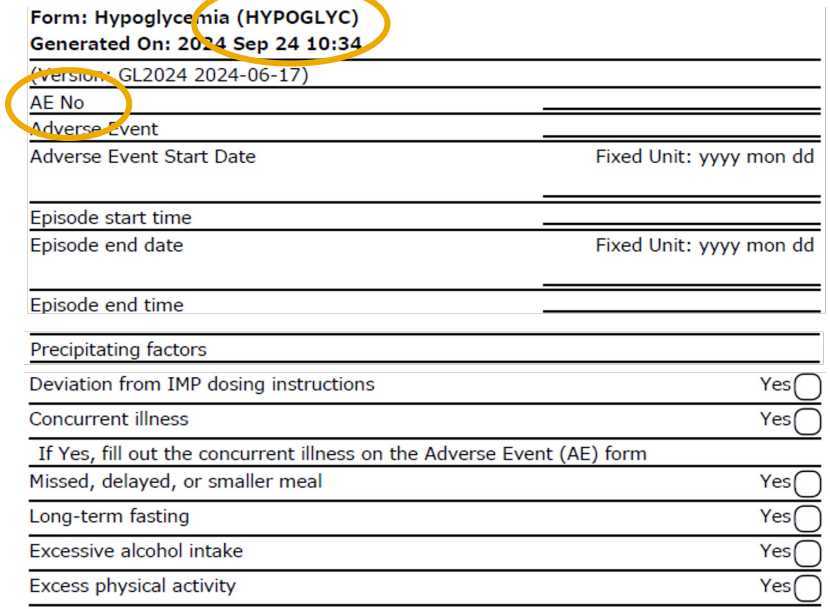
SUBJECT	MODULE	AENO	CETERM	DYSPNEA	FATIGUE
SUBJ01	HFEVT	1	Urgent heart failure visit	Y	
SUBJ01	HFEVT	2	Heart failure hospitalization	Y	Y

SUBJECT	MODULE	AENO	CETERM	DSDVPF	ILLPF
SUBJ01	HYPOGLYC	4	Hypoglycemia		Y
SUBJ02	HYPOGLYC	1	Hypoglycemia	Y	

Standardizing Linking Variable Values



Form: Heart Failure Event (HFEVT)
Generated On: 2025 Feb 03 14:05
(Version: CV2312 2023-12-12)
Event tracking number
AE number
Adverse event reported term
Adverse event start date — Fixed Unit: yyyy mon dd
Clinical event reported term — Heart failure hospitalization ☐ Urgent heart failure visit ☐
Clinical event start date — Fixed Unit: yyyy mon dd
Clinical event end date — Fixed Unit: yyyy mon dd
Symptoms of new or worsening HF — Yes ☐
Dyspnea
Decreased exercise tolerance — Yes ☐
Fatigue — Yes ☐
Worsened end-organ perfusion — Yes ☐
Volume overload — Yes ☐



Form: Hypoglycemia (HYPOGLYC)
Generated On: 2024 Sep 24 10:34
(Version: GL2024 2024-06-17)
AE No
Adverse Event
Adverse Event Start Date — Fixed Unit: yyyy mon dd
Episode start time
Episode end date — Fixed Unit: yyyy mon dd
Episode end time
Precipitating factors
Deviation from IMP dosing instructions — Yes ☐
Concurrent illness — Yes ☐
If Yes, fill out the concurrent illness on the Adverse Event (AE) form
Missed, delayed, or smaller meal — Yes ☐
Long-term fasting — Yes ☐
Excessive alcohol intake — Yes ☐
Excess physical activity — Yes ☐

SUBJECT	MODULE	AENO	CETERM	DYSPNEA	FATIGUE
SUBJ01	HFEVT	1	Urgent heart failure visit	Y	
SUBJ01	HFEVT	2	Heart failure hospitalization	Y	Y

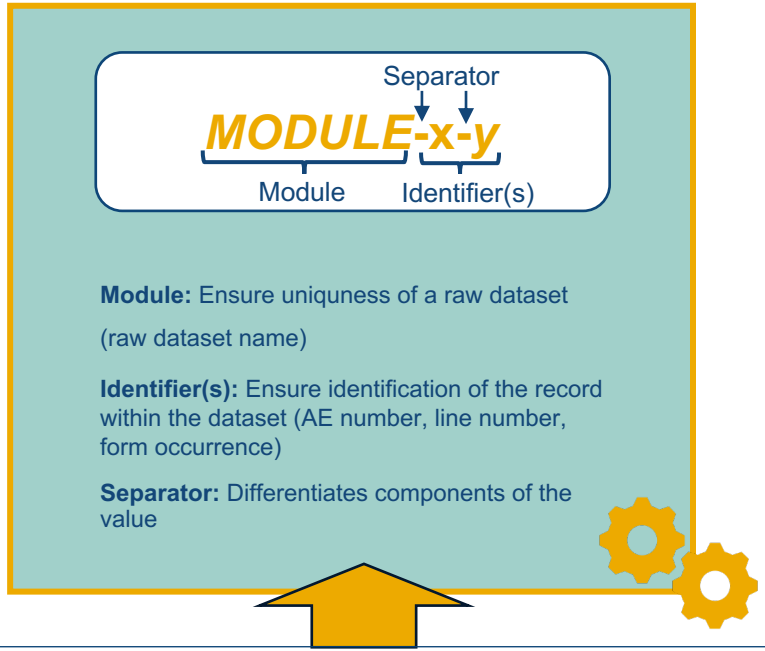
SUBJECT	MODULE	AENO	CETERM	DSDVPF	ILLPF
SUBJ01	HYPOGLYC	4	Hypoglycemia		Y
SUBJ02	HYPOGLYC	1	Hypoglycemia	Y	

Standardizing Linking Variable Values

SUBJECT	MODULE	AENO	CETERM	DYSPNEA	FATIGUE
SUBJ01	HFEVT	1	Urgent heart failure visit	Y	
SUBJ01	HFEVT	2	Heart failure hospitalization	Y	Y

SUBJECT	MODULE	AENO	CETERM	DSDVPF	ILLPF
SUBJ01	HYPOGLYC	4	Hypoglycemia		Y
SUBJ02	HYPOGLYC	1	Hypoglycemia	Y	

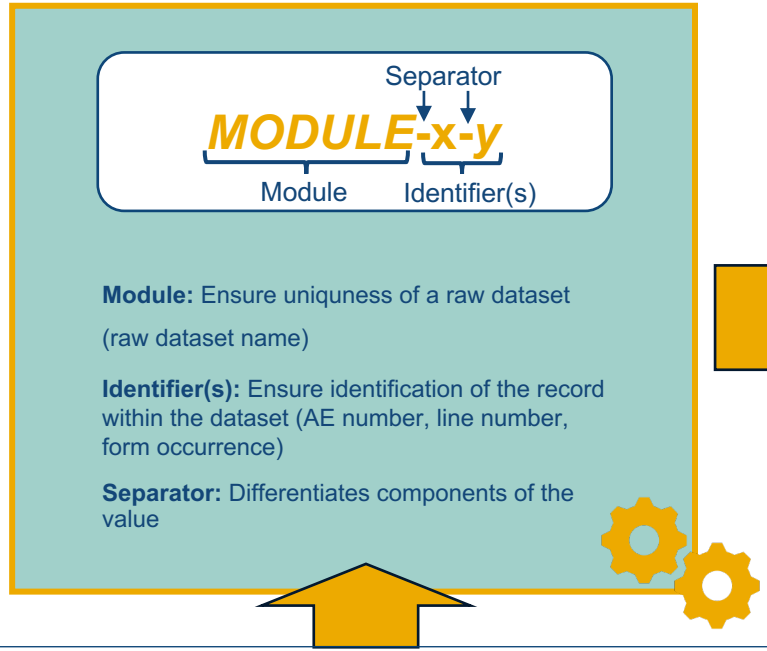
Standardizing Linking Variable Values



SUBJECT	MODULE	AENO	CETERM	DYSPNEA	FATIGUE
SUBJ01	HFEVT	1	Urgent heart failure visit	Y	
SUBJ01	HFEVT	2	Heart failure hospitalization	Y	Y

SUBJECT	MODULE	AENO	CETERM	DSDVPF	ILLPF
SUBJ01	HYPOGLYC	4	Hypoglycemia		Y
SUBJ02	HYPOGLYC	1	Hypoglycemia	Y	

Standardizing Linking Variable Values



ce.xpt

USUBJID	CELNKID	CETERM
STUDY123/SUBJ01	HFEVT-1	Urgent heart failure visit
STUDY123/SUBJ01	HFEVT-2	Heart failure hospitalization
STUDY123/SUBJ01	HYPOGLYC-4	Hypoglycemia
STUDY123/SUBJ02	HYPOGLYC-1	Hypoglycemia

fa.xpt

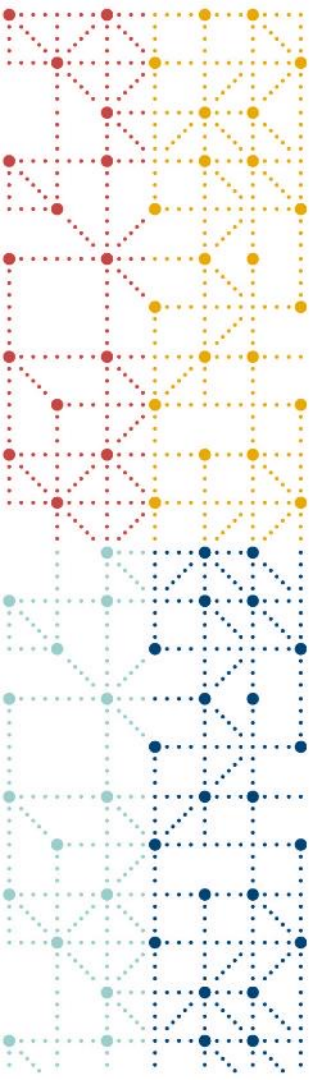
USUBJID	FALNKID	FATESTCD	FASTRESC
STUDY123/SUBJ01	HFEVT-1	NWSYMP	DYSPNEA
STUDY123/SUBJ01	HFEVT-2	NWSYMP	DYSPNEA
STUDY123/SUBJ01	HFEVT-2	NWSYMP	FATIGUE
STUDY123/SUBJ01	HYPOGLYC-4	ILLPF	Y
STUDY123/SUBJ02	HYPOGLYC-1	DSDVPF	Y

relrec.xpt

RDOMAIN	IDVAR	RELTYPE	RELID
CE	CELNKID	ONE	CEFA
FA	FALNKID	MANY	CEFA

SUBJECT	MODULE	AENO	CETERM	DYSPNEA	FATIGUE
SUBJ01	HFEVT	1	Urgent heart failure visit	Y	
SUBJ01	HFEVT	2	Heart failure hospitalization	Y	Y

SUBJECT	MODULE	AENO	CETERM	DSDVPF	ILLPF
SUBJ01	HYPOGLYC	4	Hypoglycemia		Y
SUBJ02	HYPOGLYC	1	Hypoglycemia	Y	



Automation and Testing Strategy

Automation and Testing Strategy Tracking: RELREC lookup template

	AG	AE	CE	DD	FA	HO	LB	LC	MH	PR	RS
AG	X						AGLNKID/LBLNKID	AGLNKID/LCLNKID			
AE		X	AESPID/CELNKGRP	AESPID/DDSPID	AESPID/FASPID	AESPID/HOLNKGRP	AESPID/LBLNKID	AESPID/LCLNKID			
CE		AESPID/CELNKGRP	X	CELNKID/DDLNKID	CELNKID/FALNKID					CELNKID/PRLNKID	
DD		AESPID/DDSPID	CELNKID/DDLNKID	X							
FA		AESPID/FASPID	CELNKID/FALNKID		X	FALNKID/HOLNKID			FALNKID/MHLNKID	FALNKID/PRLNKID	
HO		AESPID/HOLNKGRP			FALNKID/HOLNKID	X					
LB	AGLNKID/LBLNKID	AESPID/LBLNKID					X			LBLNKID/PRLNKID	
LC	AGLNKID/LCLNKID	AESPID/LCLNKID					LBGRPID/LCGRPID	X		LCLNKID/PRLNKID	
MH					FALNKID/MHLNKID				X		MHLNKID/RSLNKID
PR			CELNKID/PRLNKID		FALNKID/PRLNKID		LBLNKID/PRLNKID	LCLNKID/PRLNKID		X	
RS									MHLNKID/RSLNKID		X

File content

- Single source for existing defined relationships at Corporate and TA level
- Study or project specific needs for linking built onto the lookup template

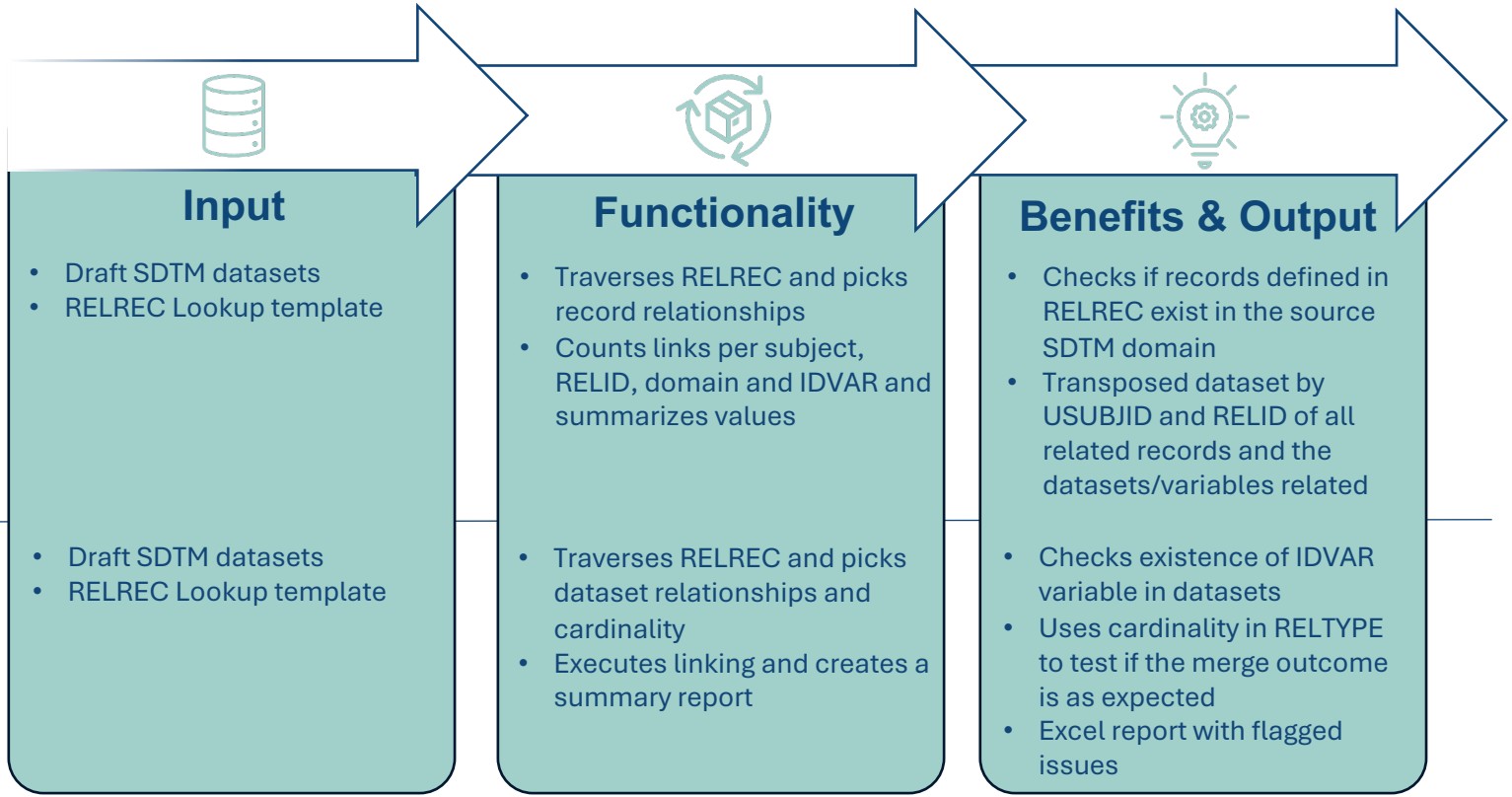
Benefits

- Existing defined relationships provide steer on what variables cannot be used
- Control ensured to not impact existing relationship while defining new relationships

Next steps

- Include cardinality of relationships in the template
- Provide drill down to the values of IDVARs, when derived
- Layer on raw values feeding into the IDVARs

Automation and Testing Strategy



Automation and Testing Strategy Macro: Future plans

For potential new relationships, provide a way to **test inputs** to the lookup template

Consume and translate the **record level relationships** output dataset to a **summary report**

Create a **summary tab for high level results** from testing the RELREC dataset

Add a pre-cursor module to consume the lookup table and SDTM RELREC specifications to **create RELREC dataset**

Implement a driver program to **switch-on/off various functionalities** in the RELREC automation and validation package

Thank You!

Emma Törner
emma.torner@astrazeneca.com

Parag Wani
parag.wani@astrazeneca.com