Analysis Concepts definition – Initial perspectives from the CDISC working group

Kirsten Walther Langendorf / Partner at data4knowledge

15-MAY-2025

Meet the Speaker

Kirsten Walther Langendorf

Title: Partner

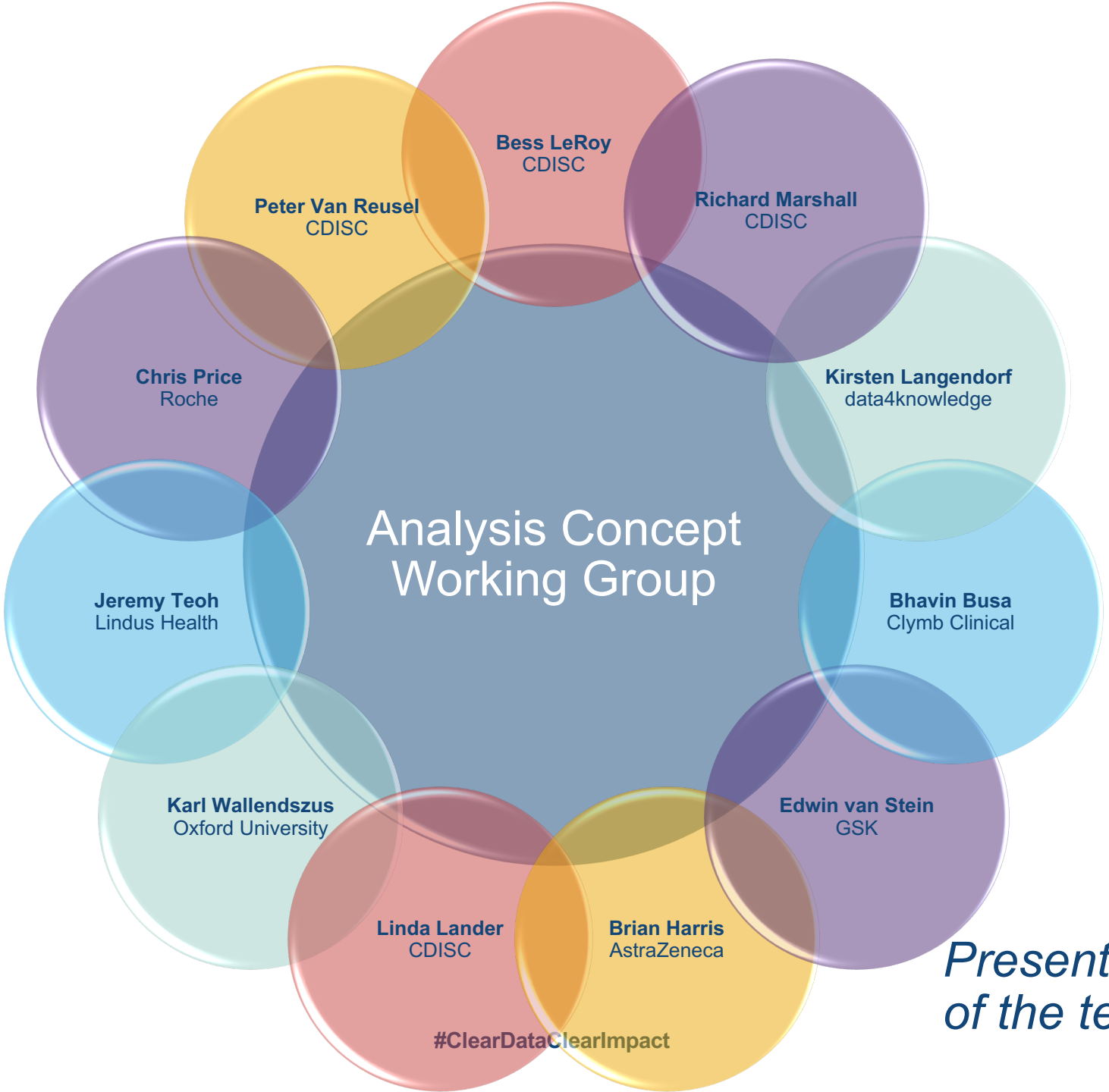
Organization: data4knowledge ApS

20+ years' experience in the pharmaceutical industry within programming, IT implementation & validation, process improvement, CDISC standards implementation, and statistics.

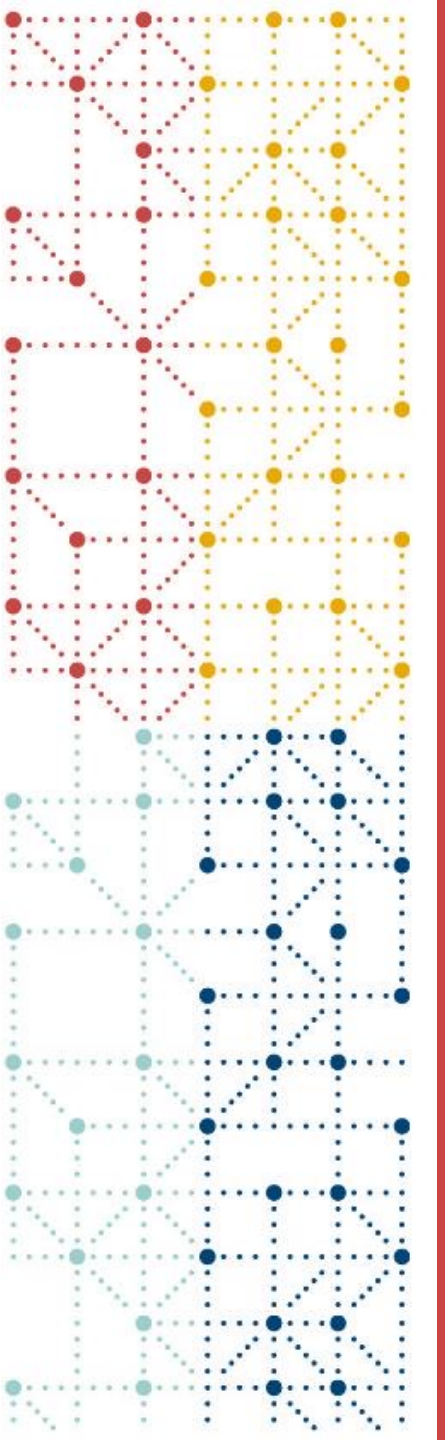
As partner at data4knowledge in Copenhagen, she has been involved in implementing various e2e metadata driven systems based on linked data technologies.

She actively contributes to the industry by volunteering with the CDISC Biomedical Concept curation team and the CDISC Analysis Concept team.

The team

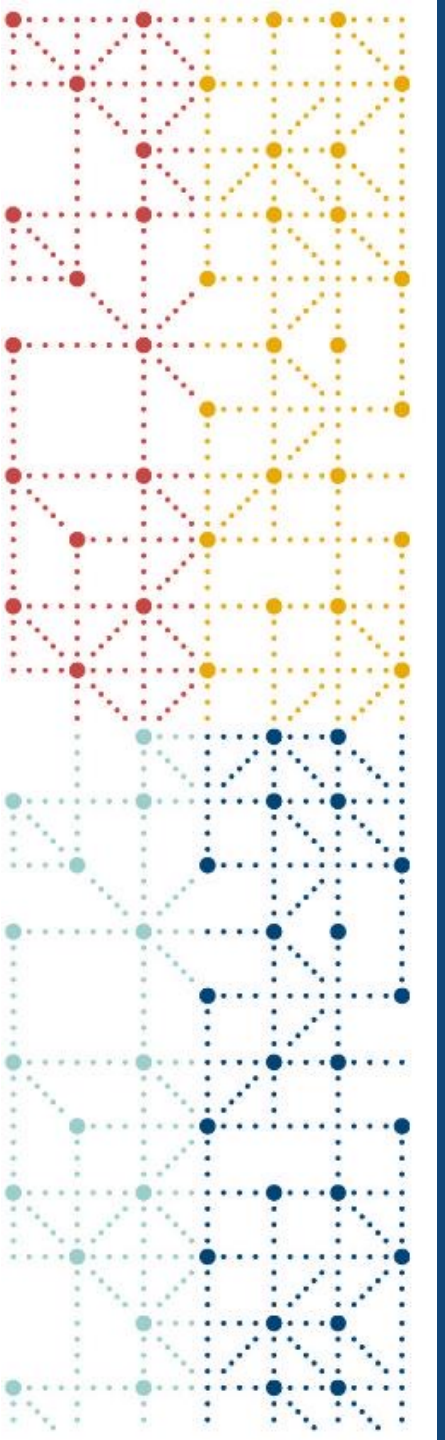


Presenting on behalf of the team



Agenda

1. Background
2. Why do we need Analysis Concepts
3. Status and considerations on Analysis Concepts use case and modelling
4. Looking ahead



Background

Analysis Concept Working group

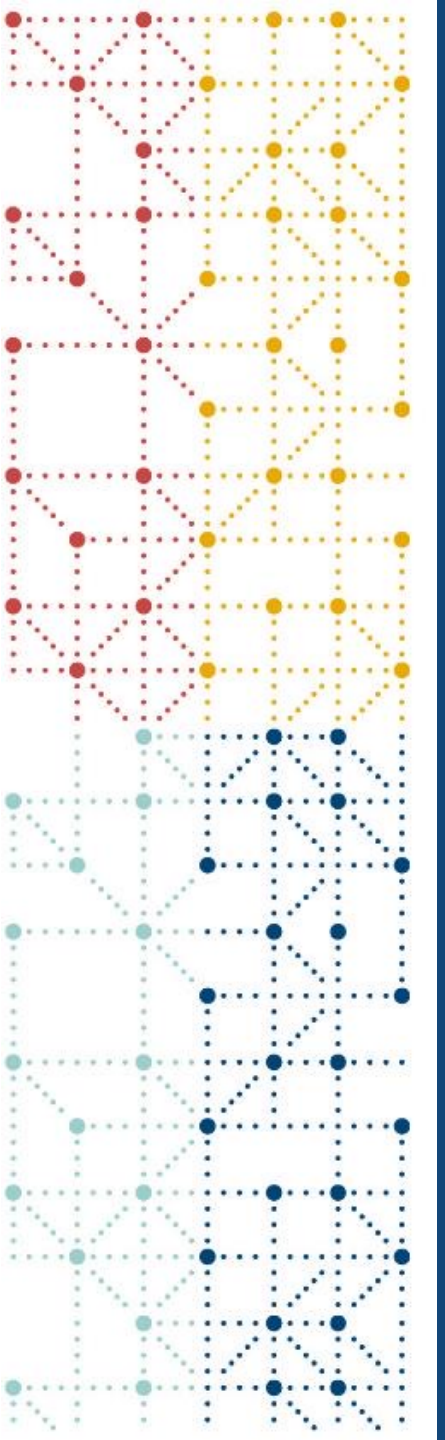
Statistics on the Analysis Concept Working group 😉

- First meeting: 22-Jan-2025
- # meetings so far: 8
- Duration/meeting: 1 hr
- Meet Wednesday every other week*

Time zone	Start	End
Central European	15:00	16:00
US Eastern Time	09:00	10:00
India	19:30	20:30

- Current meeting participants
  - Bess LeRoy (CDISC – BC and ARS team)
  - Bhavin Busa (Clymb Clinical)
  - Brian Harris (AstraZeneca)
  - Chris Price (Roche)
  - Edwin van Stein (GSK)
  - Jeremy Teoh (Lindus Health)
  - Karl Wallendszus (Oxford University)
  - Kirsten Langendorf (data4knowledge)
  - Linda Lander (CDISC – BC team)
  - Peter Van Reusel (CDISC – CSO)
  - Richard Marshall (CDISC – ARS team)

* Every week in April due to Interchange preparations



Why do we need Analysis Concepts

Today's specification of analysis

- Written in plain text
  - not machine readable
- Some details are left out
  - often (at best) retrievable from the analysis executable program/define.xml
  - reason why regulators requires SAS programs to be submitted
- Specifications are not software agnostic
  - SAS versus R syntax
  - different set of options

*know the document is old, but very likely the same is seen in newer documents

Primary analysis from CDISC pilot study*	
1	*The primary analysis of the ADAS-Cog (11) at Week 24 will use the efficacy population with LOCF imputation for any missing values at Week 24.*
2	*An ANCOVA model will be used with the **baseline score, site,** and **treatment** included as **independent variables**.* ***Treatment** will be included as a **continuous** variable, and results for a **test of dose response** will be produced.* ***Interaction terms will not be investigated**.*
3	*If the test for **dose response is statistically significant**, pairwise comparisons among the 3 groups will be performed and evaluated at a significance level of 0.05.*

1. The endpoint

- What was done (define.xml) versus text
  - CHG variable. The change from baseline at week 24 not the value at week 24
  - PARAMCD='ACTOT'
  - EFFFL='Y' and AVISIT='Week 24 – can be deduced from text

Primary analysis from CDISC pilot study	
1	*The primary analysis of the ADAS-Cog (11) at Week 24 will use the efficacy population with LOCF imputation for any missing values at Week 24.*

Analysis Variable(s)	CHG

Data References (incl. Selection Criteria)	ADQSADAS[EFFFL="Y" and ANL01FL="Y" and AVISIT="Week 24" and PARAMCD="ATOT"]

2. Specifying the model in SAS – how to translate?

From define.xml

SAS implementation

Display	Table 14-3.01 Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population)
AnalysisResult	**dose response analysis for ADAS-Cog changes from baseline**
Analysis Parameter(s)	ACTOT=Adas-Cog(11) Subscore
Analysis Variable(s)	CHG
Reason	Primary Endpoint Analysis; pre-specified in SAP
Data References (incl. Selection Criteria)	ADQSADAS[EFFFL="Y" and ANL01FL="Y" and AVISIT="Week 24" and PARAMCD="ATOT"]
Documentation	SAP Section 10.1.1Linear model analysis of CHG for dose response; using randomized dose (0 for placebo; 54 for low dose; 81 for high dose) and site group in model. Used PROC GLM in SAS to produce p-value (from Type III SS for treatment dose).

Programming Statements	proc glm data = ADQSADAS;
	where EFFFL='Y' and ANL01FL='Y' and AVISIT='Week 24' and PARAMCD='ATOT';
	class sitegr1;
	model CHG = trtpn sitegr1;
	run;

Primary analysis from CDISC pilot study	
2	*An ANCOVA model will be used with the **baseline score**, **site**, and **treatment** included as **independent variables**.* ***Treatment** will be included as a **continuous** variable, and results for a **test of dose response** will be produced.* ***Interaction terms will not be investigated**.*

R implementation

```r
adqsadas<-read_xpt("CDISC Pilot Study/updated-pilot-submission-package/900172/m5/datasets/cdiscpilot01/analysis
/adam/datasets/adqsadas.xpt")

#From define.xml : EFFFL='Y' and ANL01FL='Y' and AVISIT='Week 24' and PARAMCD="ATOT
adas_cog_11<-adqsadas%>%filter(PARAMCD=='ACTOT' & EFFFL=='Y' & AVISIT=='Week 24')
# from define.xml : class sitegr1;
# from define.xml: model CHG = trtpn sitegr1;
ancova_model <- lm(CHG ~ BASE + factor(SITEGR1) + TRTPN, data = adas_cog_11)

# Step 1: Test for dose response (significance of TRTPN coefficient)
df<-Anova(ancova_model, type = "III")
p_value_dose_response<-df['TRTPN',]$`Pr(>F)`
```

3. Pairwise estimates made – but test for DR > 0.05

Table 14-3.01
Primary Endpoint Analysis: ADAS Cog (11) - Change from Baseline to Week 24 - LOCF

	Placebo (N=79)	Xanomeline Low Dose (N=81)	Xanomeline High Dose (N=74)
Baseline			
n	79	81	74
Mean (SD)	24.1 (12.19)	24.4 (12.92)	21.3 (11.74)
Median (Range)	21.0 (5;61)	21.0 (5;57)	18.0 (3;57)
Week 24			
n	79	81	74
Mean (SD)	26.7 (13.79)	26.4 (13.18)	22.8 (12.48)
Median (Range)	24.0 (5;62)	25.0 (6;62)	20.0 (3;62)
Change from Baseline			
n	79	81	74
Mean (SD)	2.5 (5.80)	2.0 (5.55)	1.5 (4.26)
Median (Range)	2.0 (-11;16)	2.0 (-11;17)	1.0 (-7;13)
p-value(Dose Response) [1][2]			0.245
p-value(Xan - Placebo) [1][3]		0.569	0.233
Diff of LS Means (SE)		-0.5 (0.82)	-1.0 (0.84)
95% CI		(-2.1;1.1)	(-2.7;0.7)
p-value(Xan High - Xan Low) [1][3]			0.520
Diff of LS Means (SE)			-0.5 (0.84)
95% CI			(-2.2;1.1)

Pairwise comparison

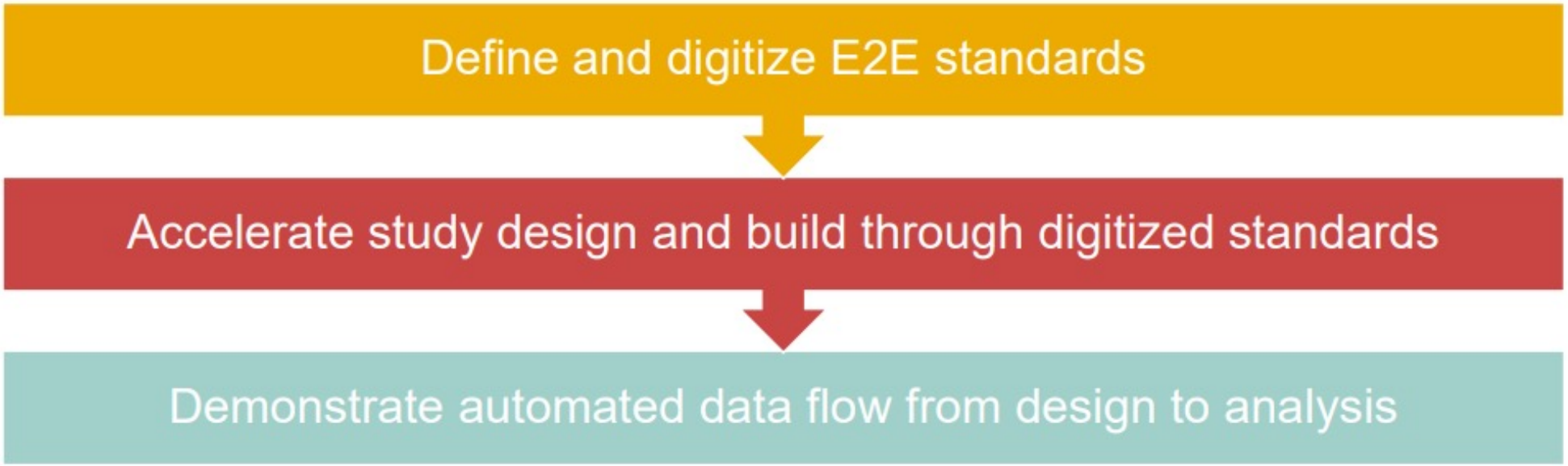
Primary analysis from CDISC pilot study	
3	*If the test for **dose response is statistically significant**, pairwise comparisons among the 3 groups will be performed and evaluated at a significance level of 0.05.*

We need better definitions

- Reducing ambiguity
  - Traditional narrative statistical analysis plans can contain ambiguities that lead to different interpretations. Analysis Concepts with standardized metadata structure ***enforce precision*** in specifying analysis settings and assumptions. Analysis Concepts create a structured way to document your statistical approaches, making regulatory review more efficient and reducing queries about your methodology.
- Enabling machine-readable analysis plans – automation
  - By structuring your analysis specifications as metadata rather than narrative text, you create ***machine-readable*** definitions that can directly link to statistical programming code. This reduces transcription errors and allows for automated validation of results against specifications.
- Supporting traceability
  - Analysis Concepts help maintain ***clear linkage between protocol objectives, endpoints***, and the specific analytical methods applied. This creates an audit trail showing how each study objective was addressed through specific statistical approaches.
- Streamlining collaboration
  - Analysis Concepts provide a ***common language*** between statisticians, clinicians, data managers, and other stakeholders. The structured format helps non-statisticians understand the planned analyses without needing to interpret complex statistical notation.

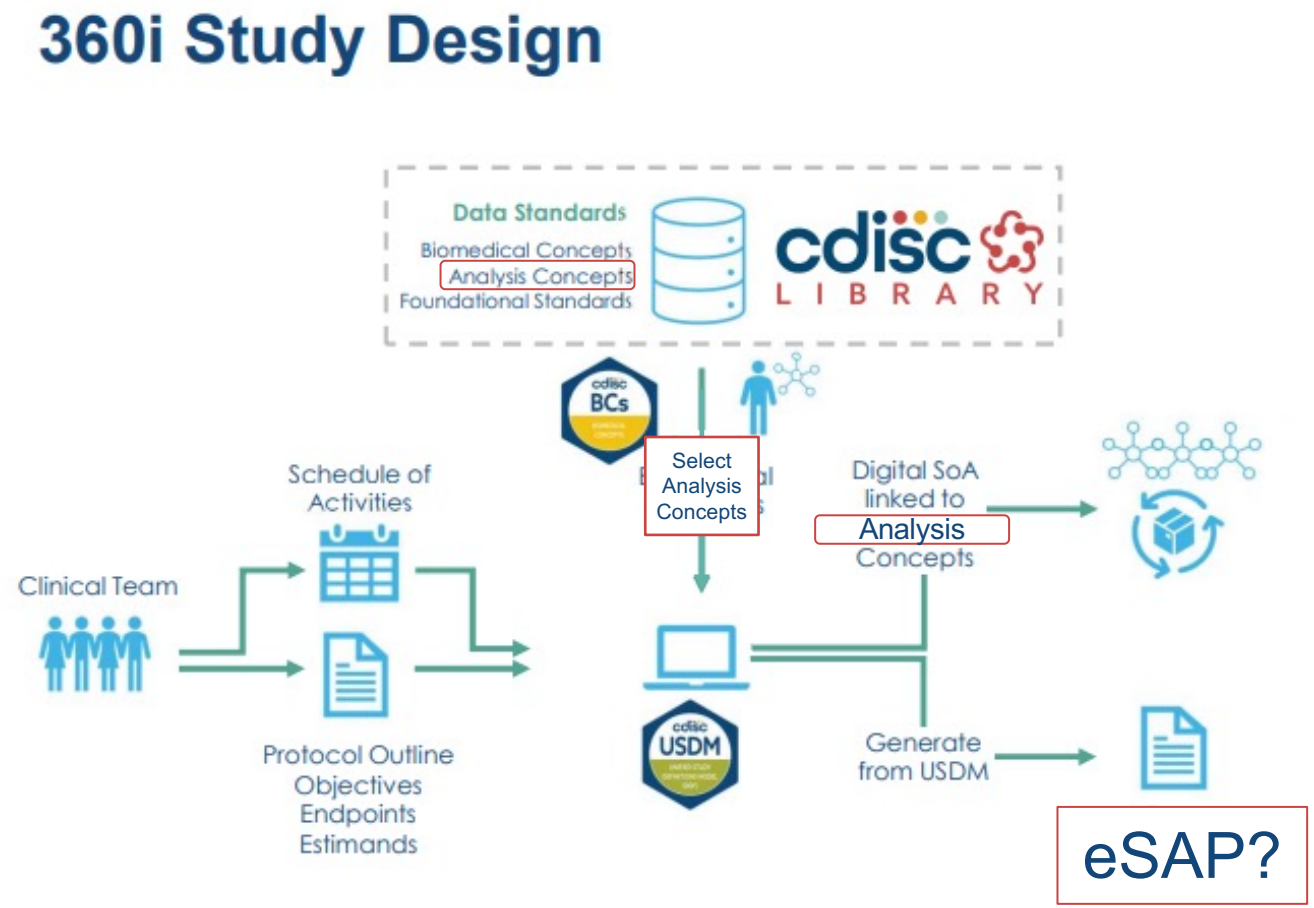
Analysis Concepts – scope of CDISC 360i

360i Journey: Ideas → Implementation → Common Practice



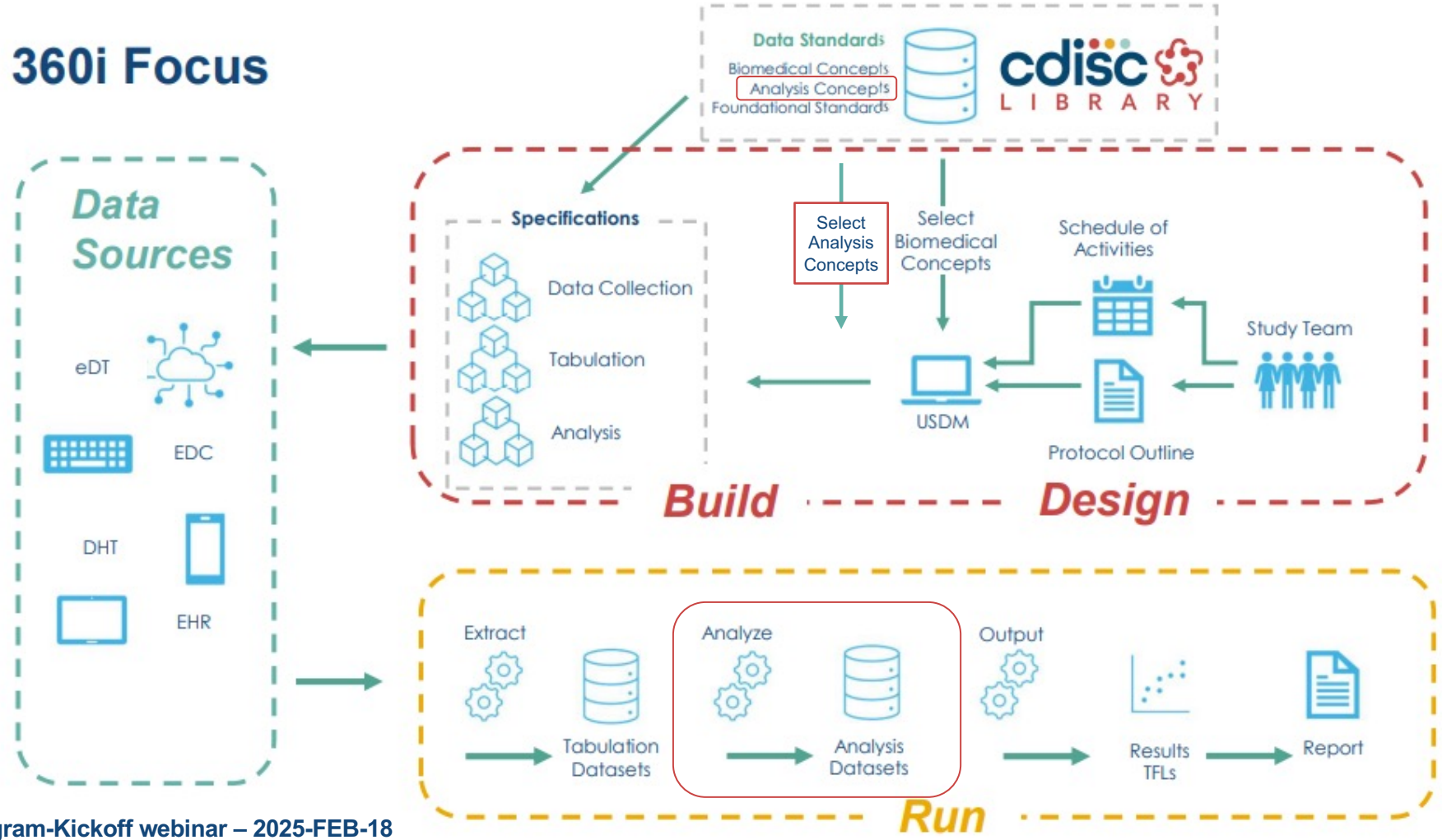
From 360i-Program-Kickoff webinar – 2025-FEB-18

Analysis Concepts – generate eSAP



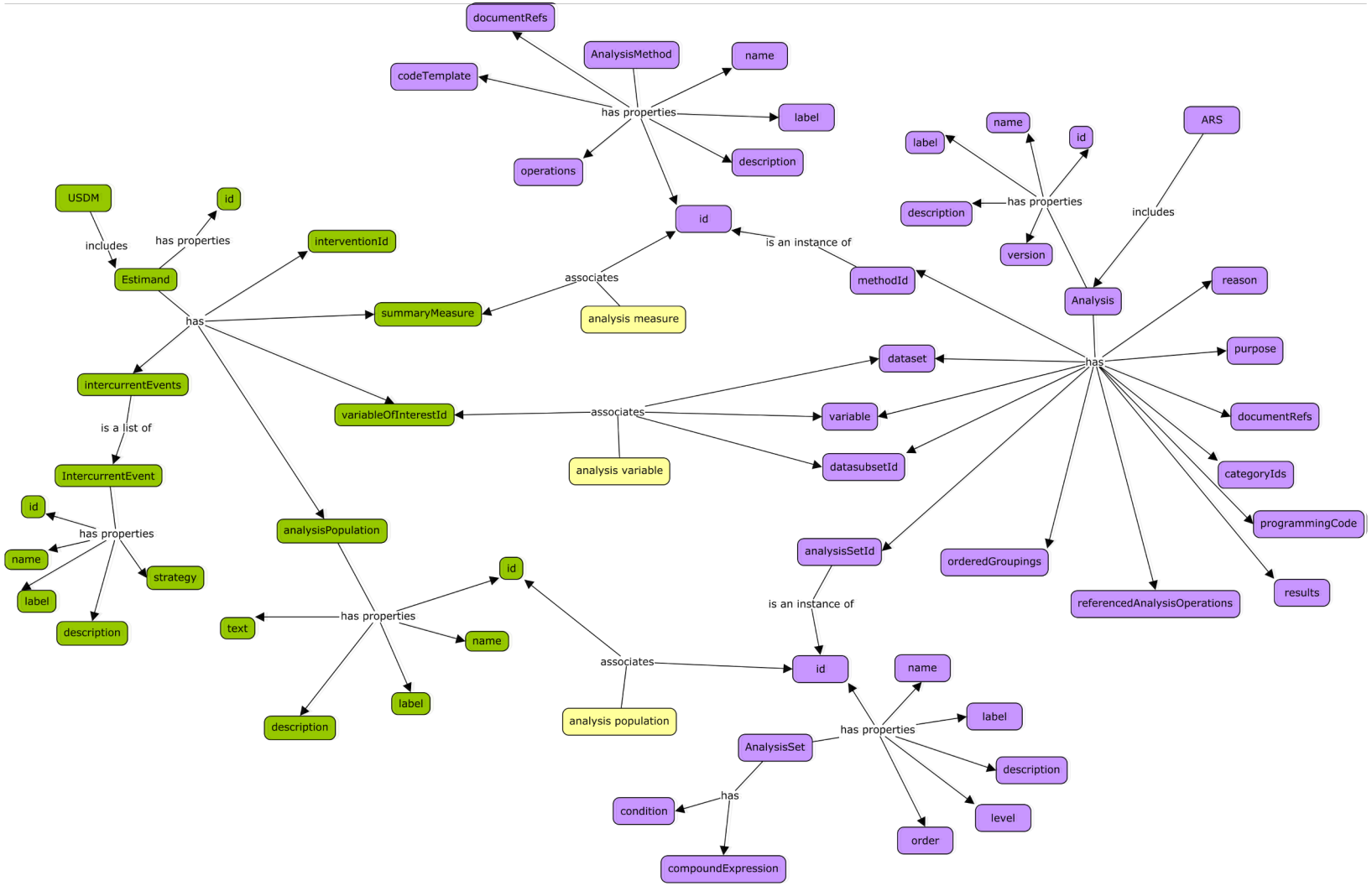
From 360i-Program-Kickoff webinar – 2025-FEB-18

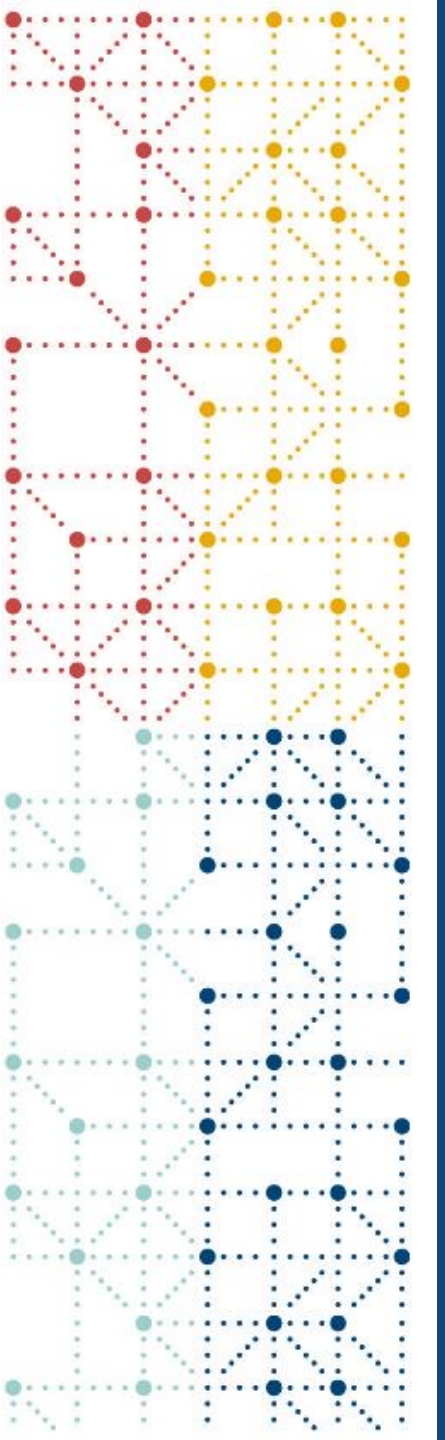
Analysis Concepts – use for downstream automation



From 360i-Program-Kickoff webinar – 2025-FEB-18

AC – the missing link between USDM and ARS





Status and considerations on Analysis Concepts use case and modelling

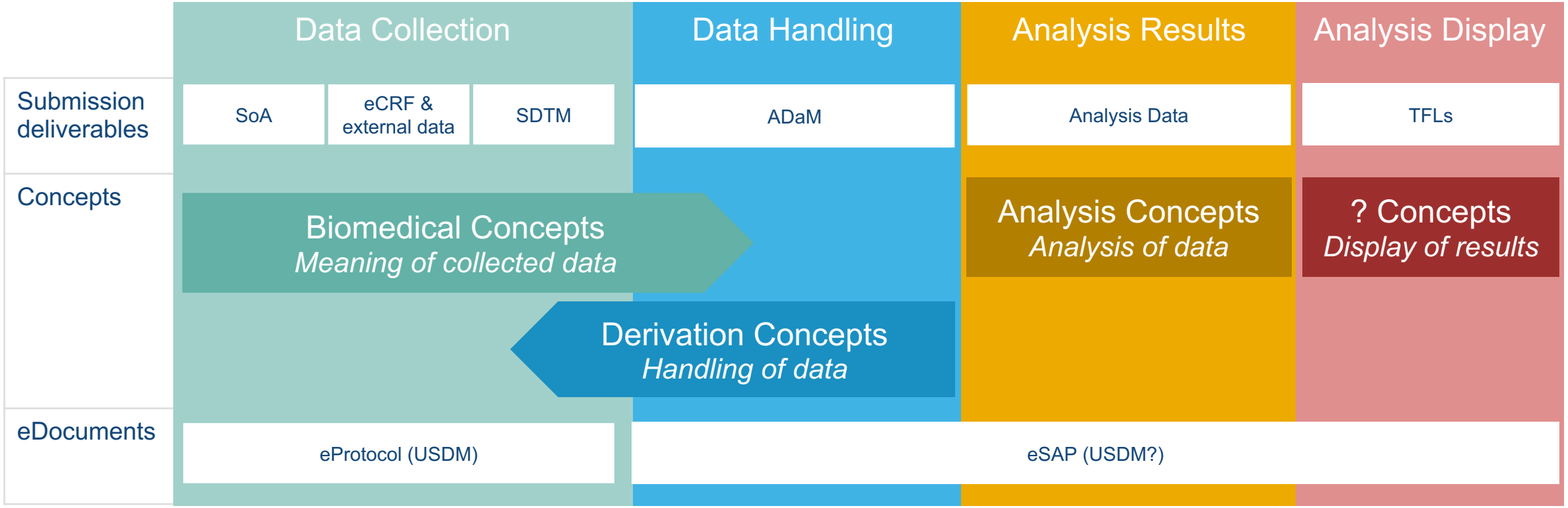
Use cases

- as a statistician, I can define the nuances of my analyses so that a human or AI programmer can accurately implement the analyses
- as a statistician, I can search for and reuse Analysis Methods
- as a statistician, I can connect methods that have different implementations, different implementation contexts
- as an academic, I can publish structured analysis design to journals
- as a programmer, I can understand the impact analysis of changes to SAP
- as a regulatory reviewer, I can understand unambiguously what analysis was done
- as a statistician, I can trace from the analysis conducted to the data point which contributed to the analysis

AC DC

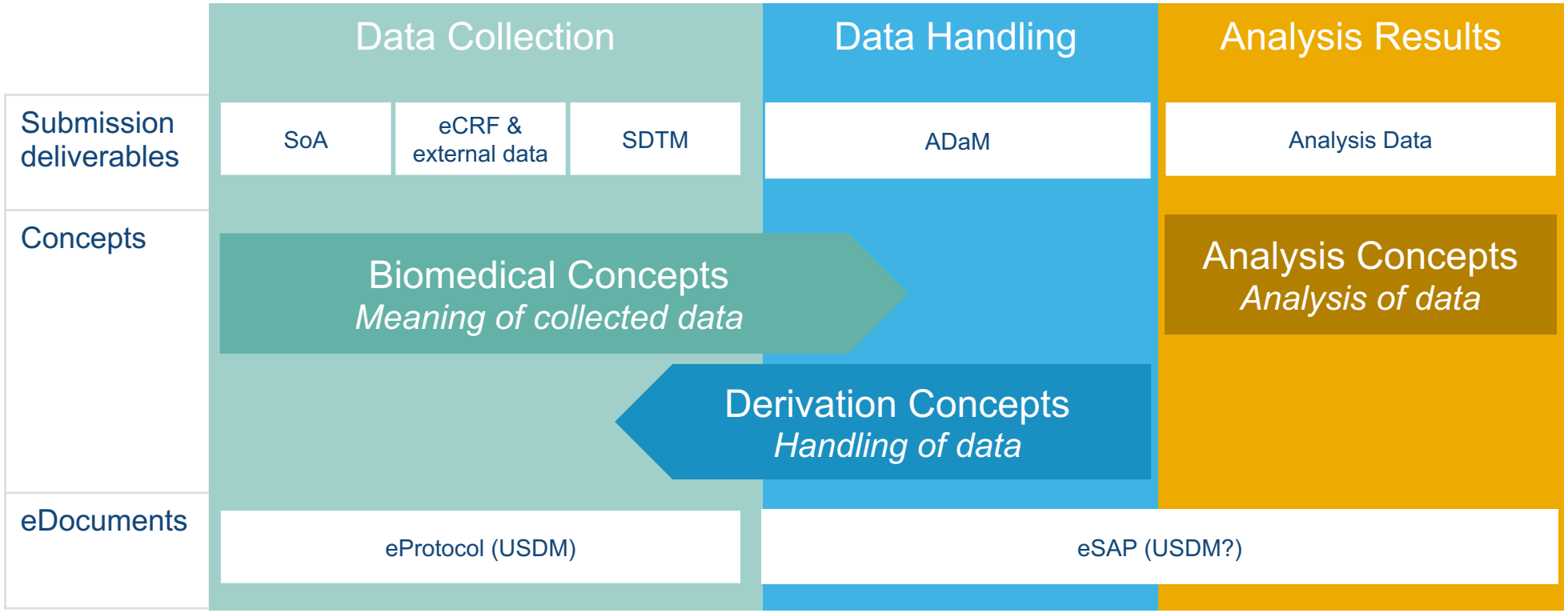
	Analysis Concept (AC)	Derivation Concept (DC)
Purpose	Examines existing data or information to draw conclusions, identify patterns, or test hypotheses	Generates new derived values from raw or derived data
Direction	Typically works with completed measurements or observations to extract meaning	Transforms or processes data to create new representations
Process	Involves applying statistical methods, critical thinking, and interpretative frameworks to understand data	Uses mathematical operations, formulas, or algorithms to calculate new quantities
Outcome	Produces insights, conclusions, or evaluations based on the data - aggregated data (not subject-level)	Creates derived data (subject level) that serve as inputs for subsequent analysis or derivations
Example 1	The p-value (from Type III Sums of Squares for treatment dose), based on linear model analysis of CHG for dose response; using randomized dose and site group in model.	CHG: Change from Baseline to Week 24 in ADAS Cog (11). Use LOCF is missing value at week 24.
Example 2	Mean value of CHG by visit	CHG: Change from BASELINE in ADAS Cog (11) by visit BASELINE = 'Y' if ADAS Cog (11) at visit 2

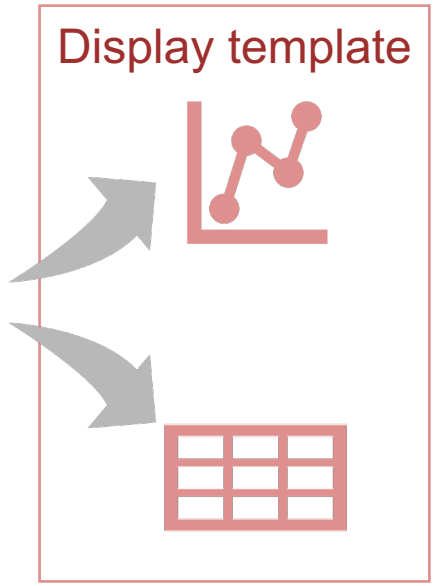
Concepts in the process



source: reproduced from similar slide made by Edwin van Stein, PHUSE SDE, Utrecht 2025

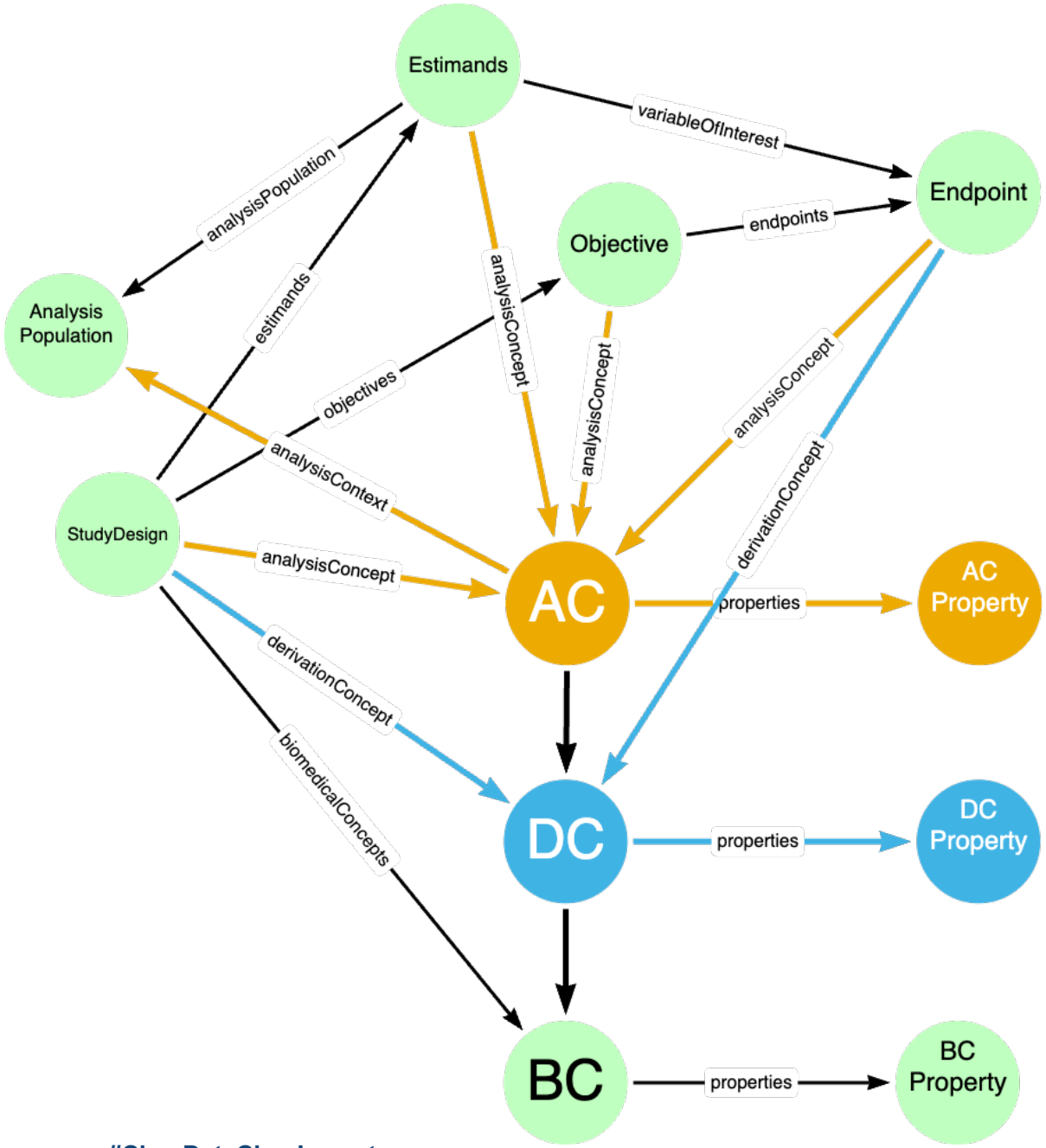
Concepts in the process – analysis data rendered for different displays





source: reproduced from similar slide made by Edwin van Stein, PHUSE SDE, Utrecht 2025

AC Basic model – like BC in USDM



Structuring the analysis text

The primary analysis of the ADAS-Cog (11) at Week 24 will use the efficacy population with LOCF imputation for any missing values at Week 24.

An ANCOVA model will be used with the baseline score, site, and treatment included as independent variables.

Treatment will be included as a continuous variable, and results for a test of dose response will be produced.

Interaction terms will not be investigated.

The primary analysis of the [ADAS-Cog (11)] at [Week 24] will use the [efficacy population] with [LOCF imputation] for any missing values at Week 24.

An [ANCOVA model] will be used with the [baseline score], [site], and [treatment] included as independent variables.

Treatment will be included as a [continuous variable], and results for a [test of dose response] will be produced.

[Interaction terms will not be investigated].

Statistical Analysis Mapping Table - example

The primary analysis of the [ADAS-Cog (11)] at [Week 24] will use the [efficacy population] with [LOCF imputation] for any missing values at Week 24.

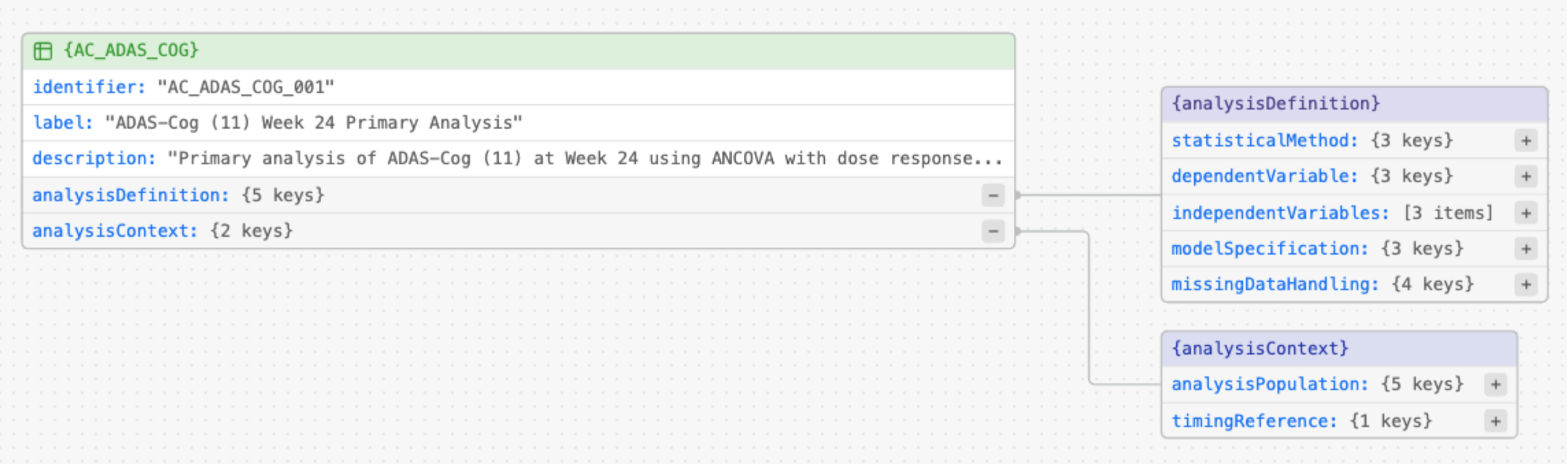
An [ANCOVA model] will be used with the [baseline score], [site], and [treatment] included as independent variables.

Treatment will be included as a [continuous variable], and results for a [test of dose response] will be produced.

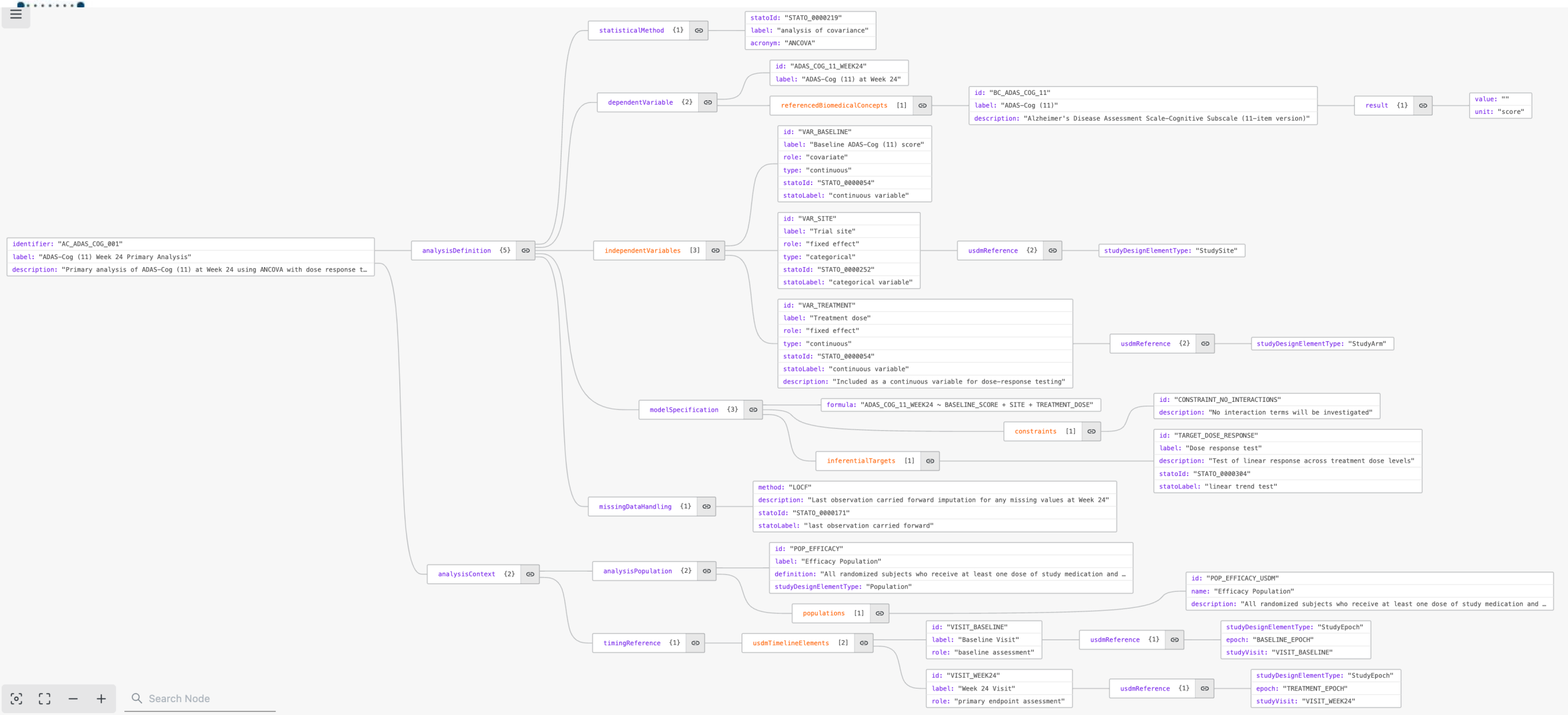
[Interaction terms will not be investigated].

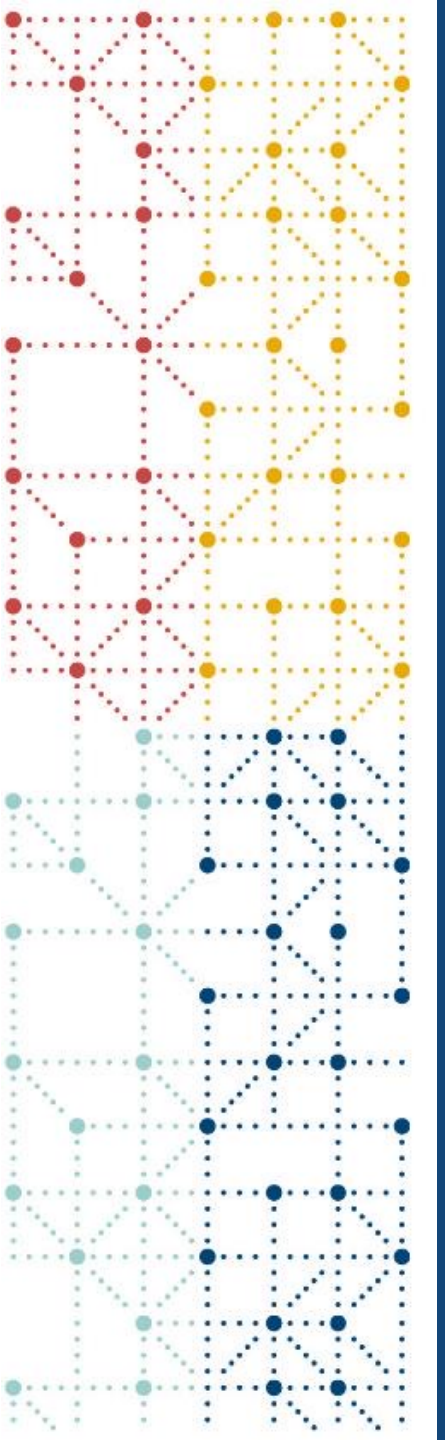
Tagged Text	AC Parent Property	AC Child Property	AC Property Detail/Reference	Value
ADAS-Cog (11)	analysisDefinition	dependentVariable	referencedBiomedicalConcepts.label	"ADAS-Cog (11)"
Week 24	analysisContext	timingReference	usdmTimelineElements.label	"Week 24 Visit"
efficacy population		analysisPopulation	label	"Efficacy Population"
LOCF imputation	analysisDefinition	missingDataHandling	method	"LOCF"
ANCOVA model		statisticalMethod	label	"analysis of covariance"
baseline score		independentVariables	label	"Baseline ADAS-Cog (11) score"
site		independentVariables	label	"Trial site"
treatment		independentVariables	label	"Treatment dose"
continuous variable		independentVariables	type	"continuous"
test of dose response		modelSpecification	inferentialTargets.label	"Dose response test"
Interaction terms will not be investigated		modelSpecification	constraints.description	"No interaction terms will be investigated"

Could the AC look like this?



Could the AC look like this – expanding details?





Looking ahead

Work ahead of us

- We need to work more on the AC model
  - Which structure and properties
- How does it fit into USDM
  - Endpoint
  - Objective
  - Estimands
- How Does it fit with ARS
- We should consider how we can 'templify' the ACs, since an analysis could be applied for many different endpoints
  - User will use a template and 'configure' it for study level
- We need to make a PoC (360i) for a USDM study with BC -> DC -> AC
  - We have not touched how derivations concepts are to be defined
  - How can we make the flow executable
- We need to investigate how/if to make eSAP

Thank You!

Contact:

Bess LeRoy bleroy@cdisc.org