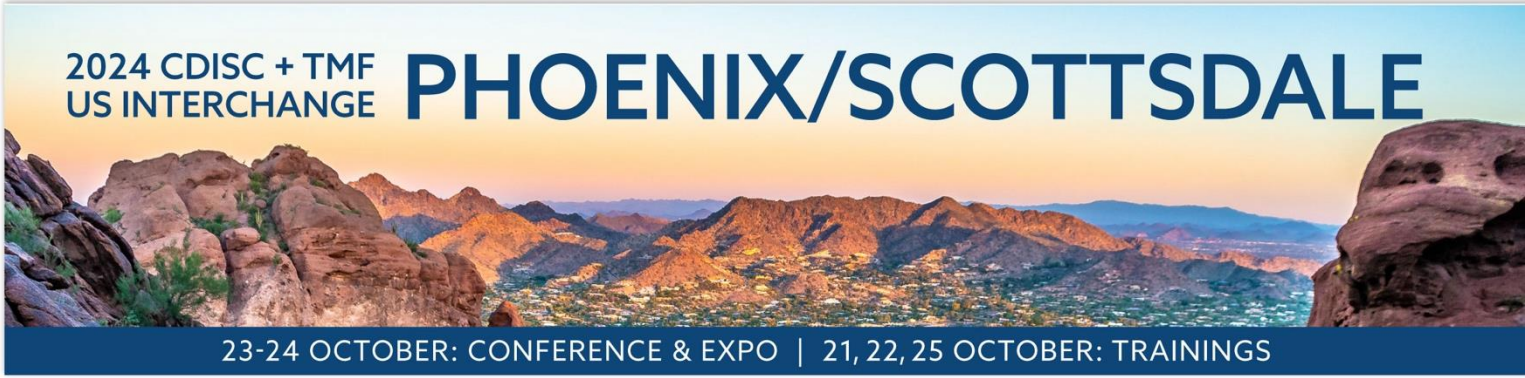


Lilly's Metadata-Driven Innovation Journey

Keith Hibbetts
Andy Miskell
Eli Lilly and Company

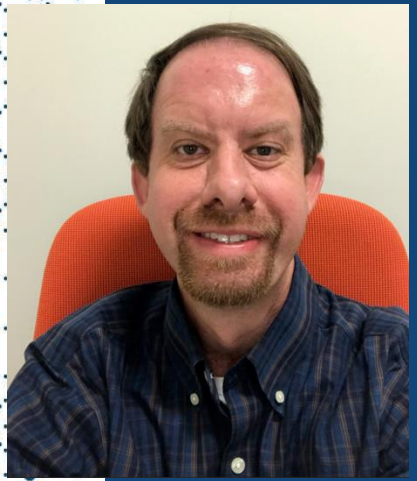
Meet the Speakers

Keith Hibbetts

Title: Senior Director – Analysis Data Standards

Organization: Eli Lilly and Company

Keith Hibbetts has 24 years of industry experience in roles that include programming, leadership, and data standards. He currently leads a team focused on data standards for SDTM, ADaM, and TFLs and leads initiatives in automation for SDTM and ADaM.



Andy Miskell

Title: Advisor Computational Statistician

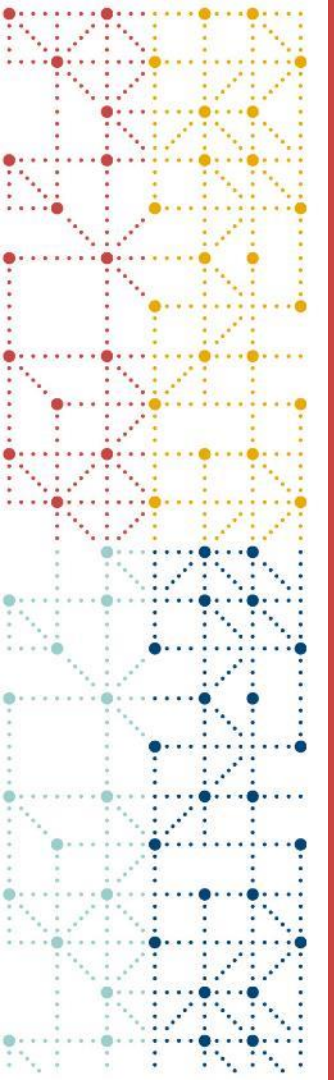
Organization: Eli Lilly and Company

Andy Miskell has been a Statistical Programmer for 25 years working for both CROs and Pharma companies. He currently leads a team of programmers to develop centralized macros for cross-study implementation of Analysis Results Datasets, TFLs, and other elements in the clinical data flow using both SAS and R. He is part of teams devising strategy for process automation including Analysis Results Datasets as well as migrating to use of open-source coding languages in clinical data reporting.

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.*

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



Agenda

1. Overview
2. Data Collection Standards
3. SDTM
4. ADaM
5. ARDS

From:
Mostly
manual
effort

To:
Mostly
automated
effort

Robust Data Standards

Metadata

Generic Automation Code

Metadata Repository

Programming Process

Metadata Repository

- STAMP (STandards Automation Metadata Process)
- TCS ADD Platform – significant amount of customization

Data Collection Standards

- Robust data collection standards are critical
  - Use needs to be governed
- Go-Live in STAMP was Oct. 2022
  - Over 750 data collection standards were migrated
    - Grown to over 900 today

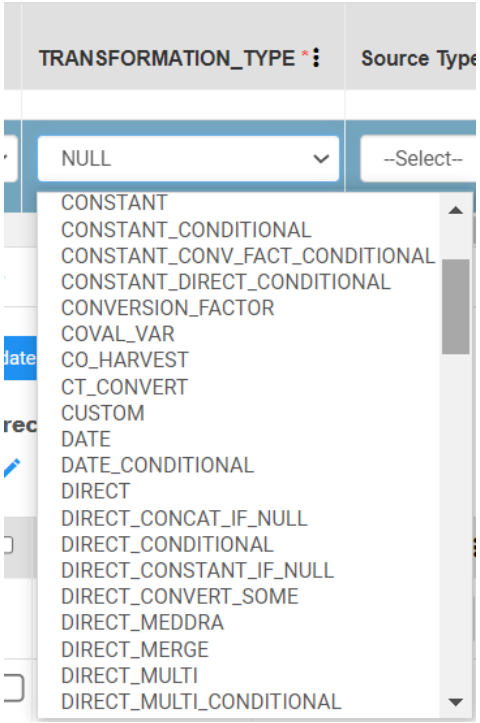
SDTM

- Robust SDTM standards were in place at beginning of the journey
  - Changes were necessary to achieve automation
  - Instead of text-based algorithms, we needed granular transformation metadata
    - This metadata feeds into generic automation code
    - This hierarchal approach to programming is different than past processes.
- Go-Live was Feb. 2023
  - ~60 studies have flowed through this process
  - Consistently seeing 95-97% of variables automated

Transformation Metadata

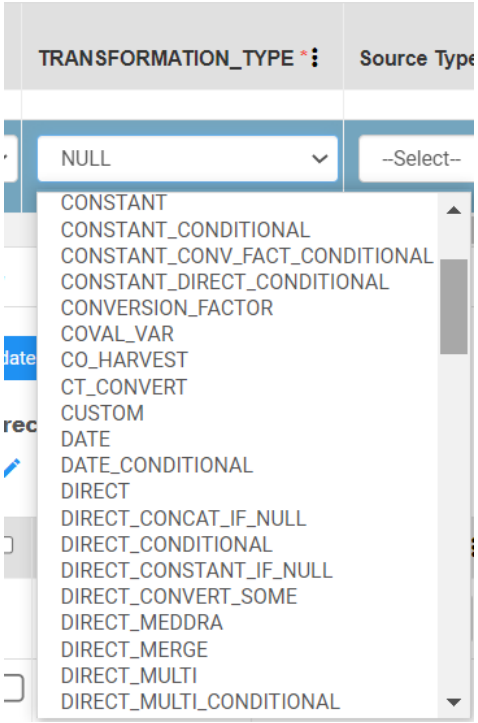
- 1 row per SDTM variable per input source dataset
- Defines how exactly how that SDTM variable is created for records from that source

Transformation Metadata



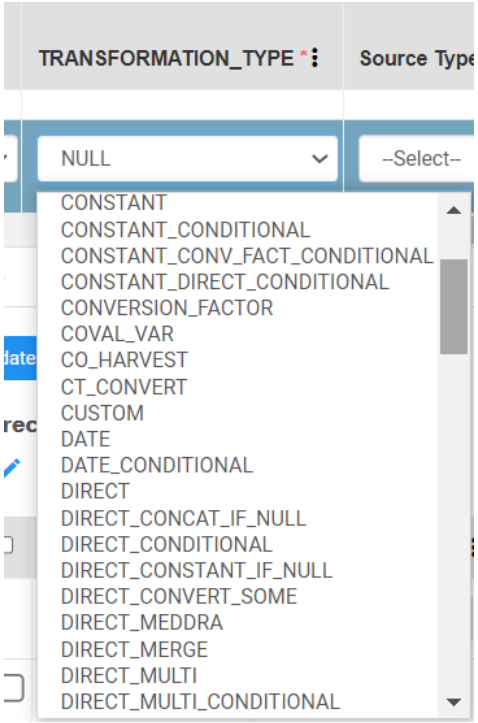
- Transformation Type is the key field
- Defines the type of logic to be used to create the variable
- Approximately 100 different transformation types

Transformation Metadata



- Some are highly reusable
  - DIRECT -> Used when an input variable has the exact value needed in the output
  - CONSTANT -> Used when each output record should have the same constant value
  - CONCAT -> Used when input variables should be concatenated into an output value
  - NULL -> null value on all records

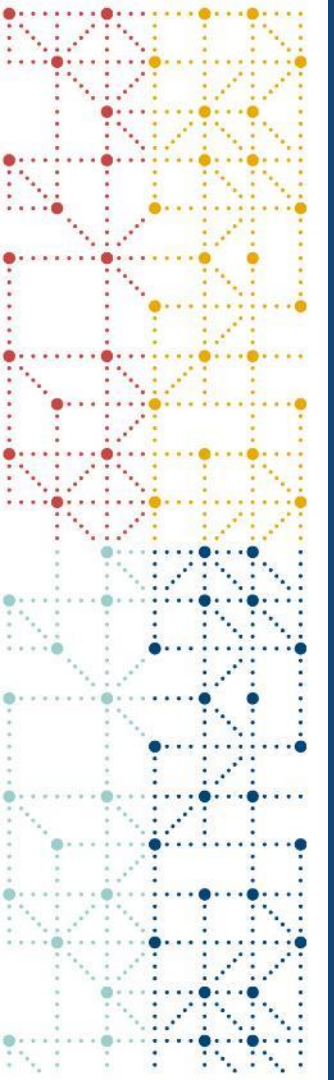
Transformation Metadata



- Some are specific to a type of variable
  - SEQ -> Creates the xxSEQ variable
  - STUDYDAY -> Calculates time from the subject's reference start date
- Some are highly specific to particular variables/input sources

ADaM

- ADaM functionality build is in-progress
  - Anticipated early 2025 launch
- Same basic metadata concepts as SDTM
  - Granular transformation metadata
  - Drive use of internal standards and automate as much as possible



Analysis Results DataSets (ARDS)

Re-imagine analysis process flow to save out analysis into re-usable files and then make display (TFL) purely a formatting step with no extra analysis performed.

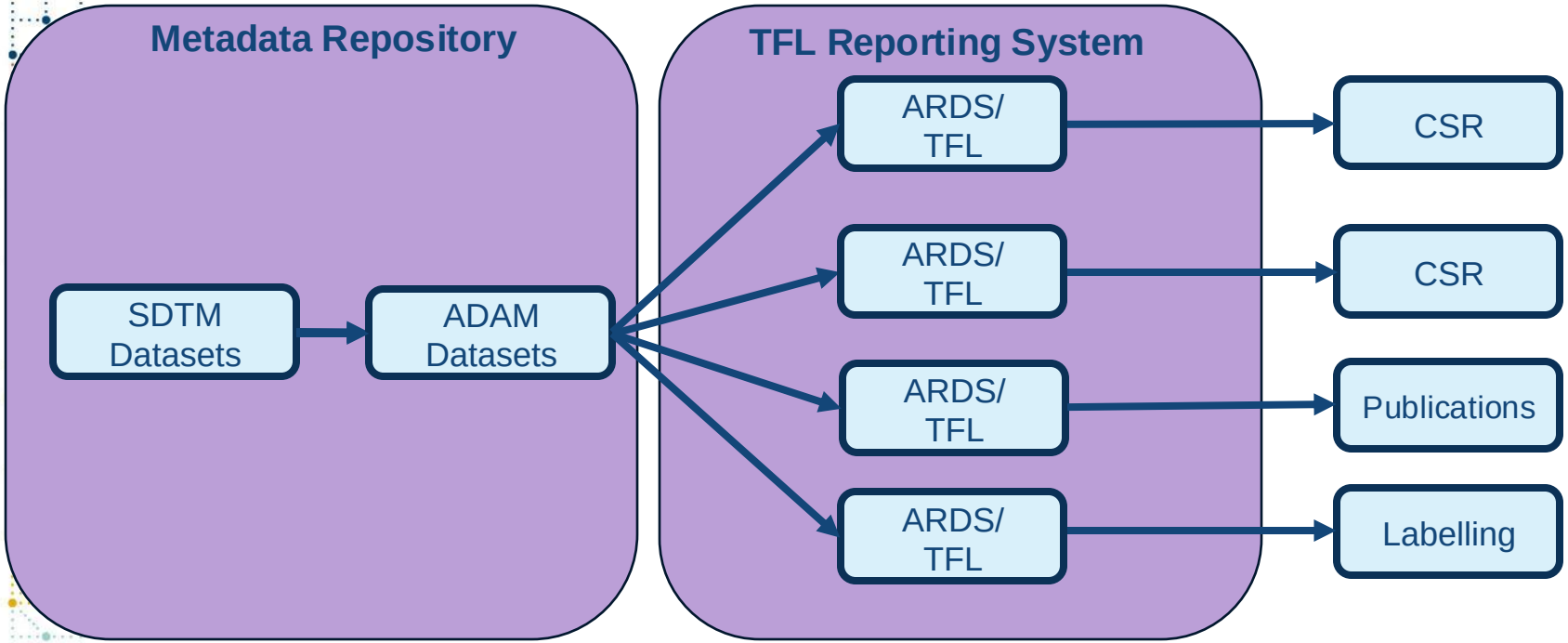
Include in MDR to facilitate dynamic end-to-end transparency of data flow

Robust ARDS Model with vertical structure

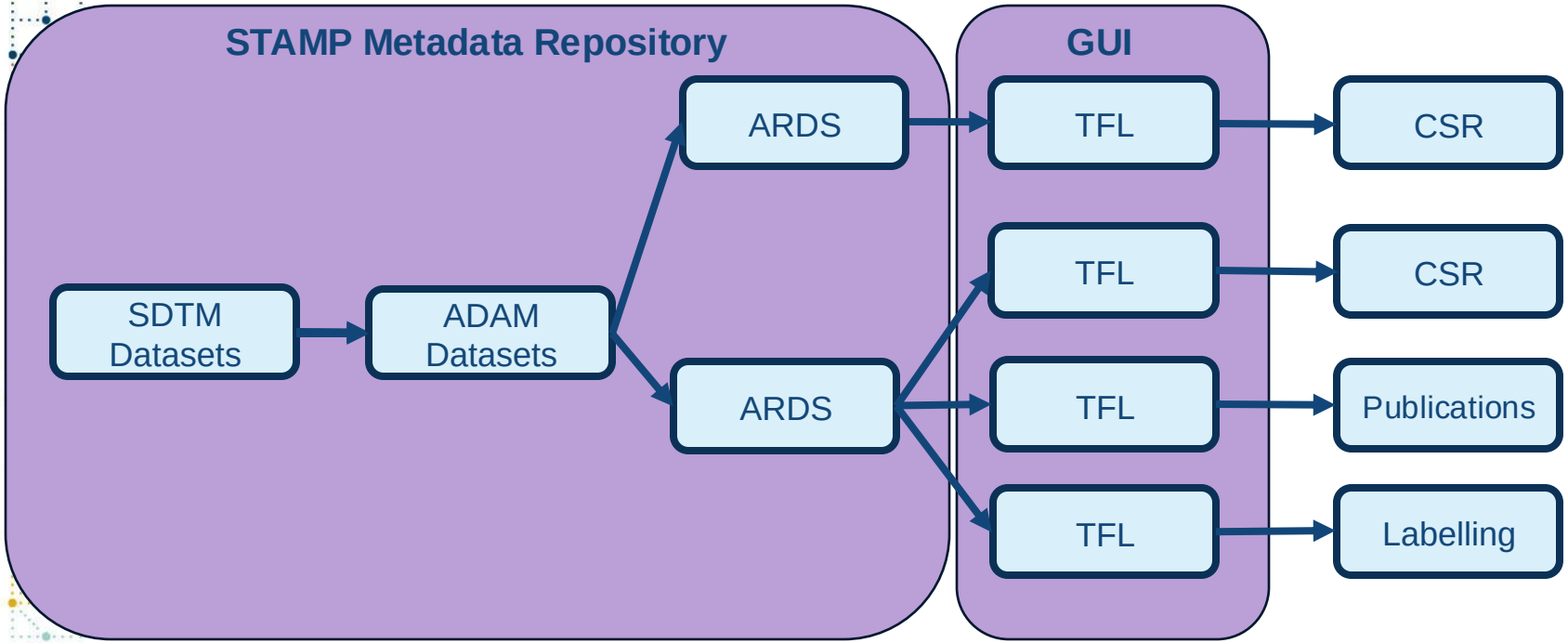
- ARDS structure built out and tested across myriad of standards to ensure robustness
- Vertical
- Unrounded values
- Any analysis and input structure
- Standard naming conventions and controlled terminology which can be used by downstream tools consuming files

POPFLAG	POPULATION	CAT1CDVAR	CAT1CD	CAT1VAR	CAT1N	CAT1	CAT2VAR	CAT2N	CAT2	CAT3VAR	CAT3N	CAT3	TRTN	TRT	REFTRTN	REFTRT	ANALYSISTYPE	RESULTTYPE	RESULTTYPE2	RESULTTYPE3	RESULTTYPE4	RESULT
ITTFL	INTENT TO TREAT POPULATION	PARAMCD	XXXXXXX	PARAM	1	xxxxx	AVISIT	1	Baseline				1	Trt A			Baseline	Count	Observed		MMRM	12345
ITTFL	INTENT TO TREAT POPULATION	PARAMCD	XXXXXXX	PARAM	1	xxxxx	AVISIT	1	Baseline				2	Trt B			Baseline	Count	Observed		MMRM	12345
ITTFL	INTENT TO TREAT POPULATION	PARAMCD	XXXXXXX	PARAM	1	xxxxx	AVISIT	1	Baseline				3	Trt C			Baseline	Count	Observed		MMRM	12345
ITTFL	INTENT TO TREAT POPULATION	PARAMCD	XXXXXXX	PARAM	1	xxxxx	AVISIT	1	Baseline				4	Pooled 1			Baseline	Count	Observed		MMRM	12345
ITTFL	INTENT TO TREAT POPULATION	PARAMCD	XXXXXXX	PARAM	1	xxxxx	AVISIT	1	Baseline				5	Pooled 2			Baseline	Count	Observed		MMRM	12345
ITTFL	INTENT TO TREAT POPULATION	PARAMCD	XXXXXXX	PARAM	1	xxxxx	AVISIT	1	Baseline				999	Total			Baseline	Count	Observed		MMRM	12345

Clinical Data Flow (Current)



Clinical Data Flow (Future)



ARDS Instantiation

- Robust ARDS Model with vertical structure
- Modularize specifications in system
- Flexibility and Tailored options for each analysis
- Metadata-driven code produced to create files from requirements allowing for greater standardization
- Dynamic traceability back through ADAM and SDTM to Data Collection

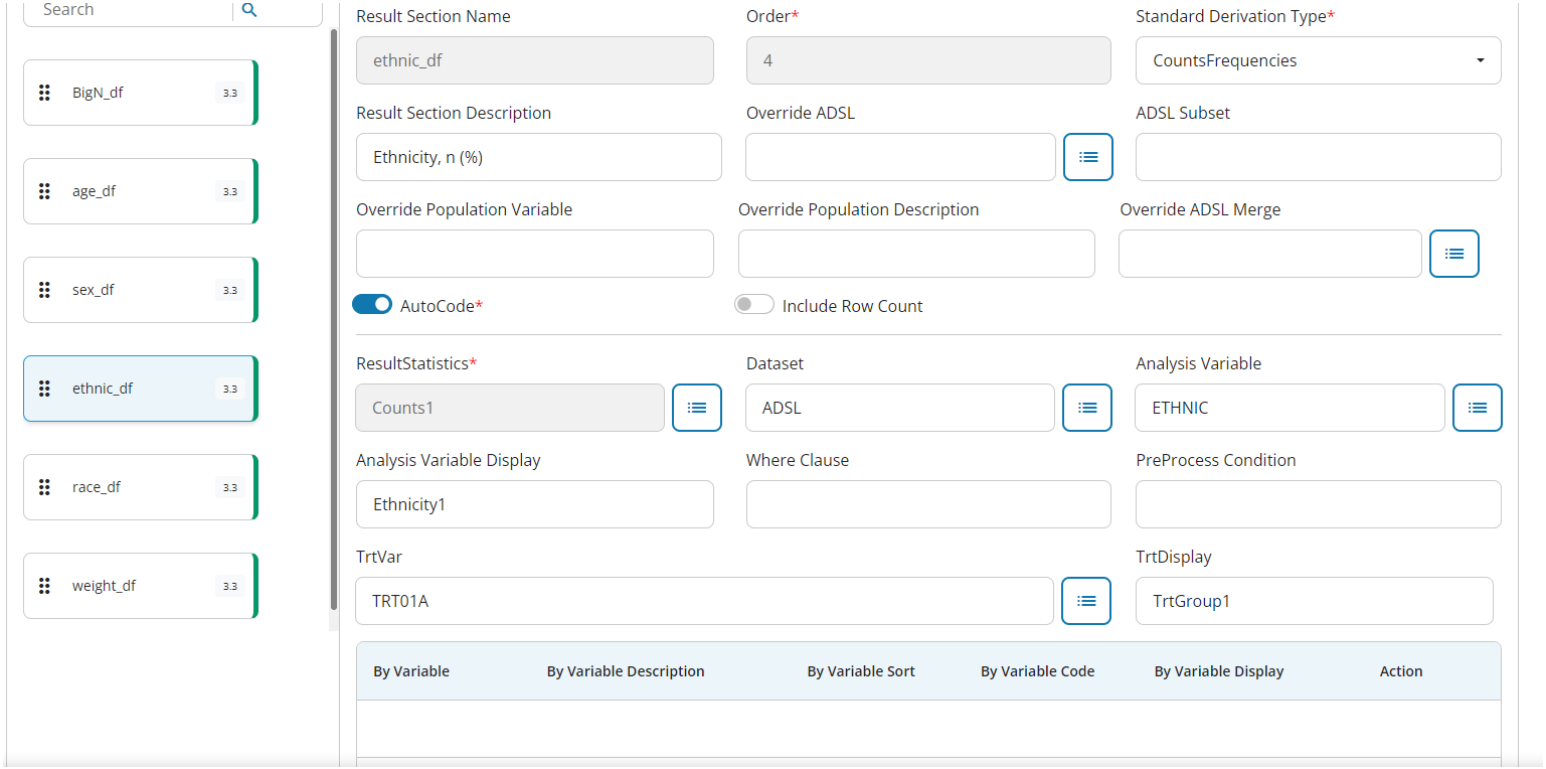
Robust ARDS Model with vertical structure

Statistics Macros
(This list will grow)

- summaryStatistics
- countsFrequencies
- pvaluesOverall
- pvaluesPairwise
- pvaluesPairwiseContinuous
- oddsRatios
- riskDifference
- lsMeanStatistics
- lsMeanDiffStatistics
- fullModelStatisticsMMRM
- fullModelStatisticsANCOVA
- nonStandard
- survfit
- linearRegression
- correlation
- icc_Psychometric
- t_Statistic
- npar1way

Result Section Name	Status	SDTM Version	Standard Derivation	Result Section Description
BigN_df	NoChange	3.3	CountsFrequencies	Big N Treatment Counts
age_df	NoChange	3.3	SummaryStatistics	Age (years)
sex_df	NoChange	3.3	CountsFrequencies	Gender, n (%)
ethnic_df	NoChange	3.3	CountsFrequencies	Ethnicity, n (%)
race_df	NoChange	3.3	CountsFrequencies	Race, n (%)
weight_df	NoChange	3.3	SummaryStatistics	Weight (kg)

Modularize specifications in system



Flexibility and Tailored options for each analysis

Summary Statistics

Display Pattern*: SummStats1

Standard Derivation Type*: SummaryStatistics

Search

Stat Value*	Stat Label*	Derivation
n	n	
mean	Mean	
sd	Std Dev	
median	Median	
minimum	Minimum	
maximum	Maximum	

ADD +

MMRM

Model Parameters:	
Model:	TRT Base*TRT AVISIT
Class Variables	AVISIT
Covariate (dropdown list and can be re-ordered):	UN TOEP ARH(1)
lsmeans	TRT*AVISIT
lsmeans options	diff cl
Random	USUBJID
Degrees of Freedom	Satterwaite
Confidence Interval Width	95
Repeated	AVISITN/SUBJECT=USUBJID un
Requested Statistics	
LS Mean	LS Mean|LS Mean
Std Err	Std Err|LS Mean
CI Low	95% CI Low|LS Mean
CI High	95% CI High|LS Mean
p-value	p-value|LS Mean

Metadata-driven code to create files from requirements

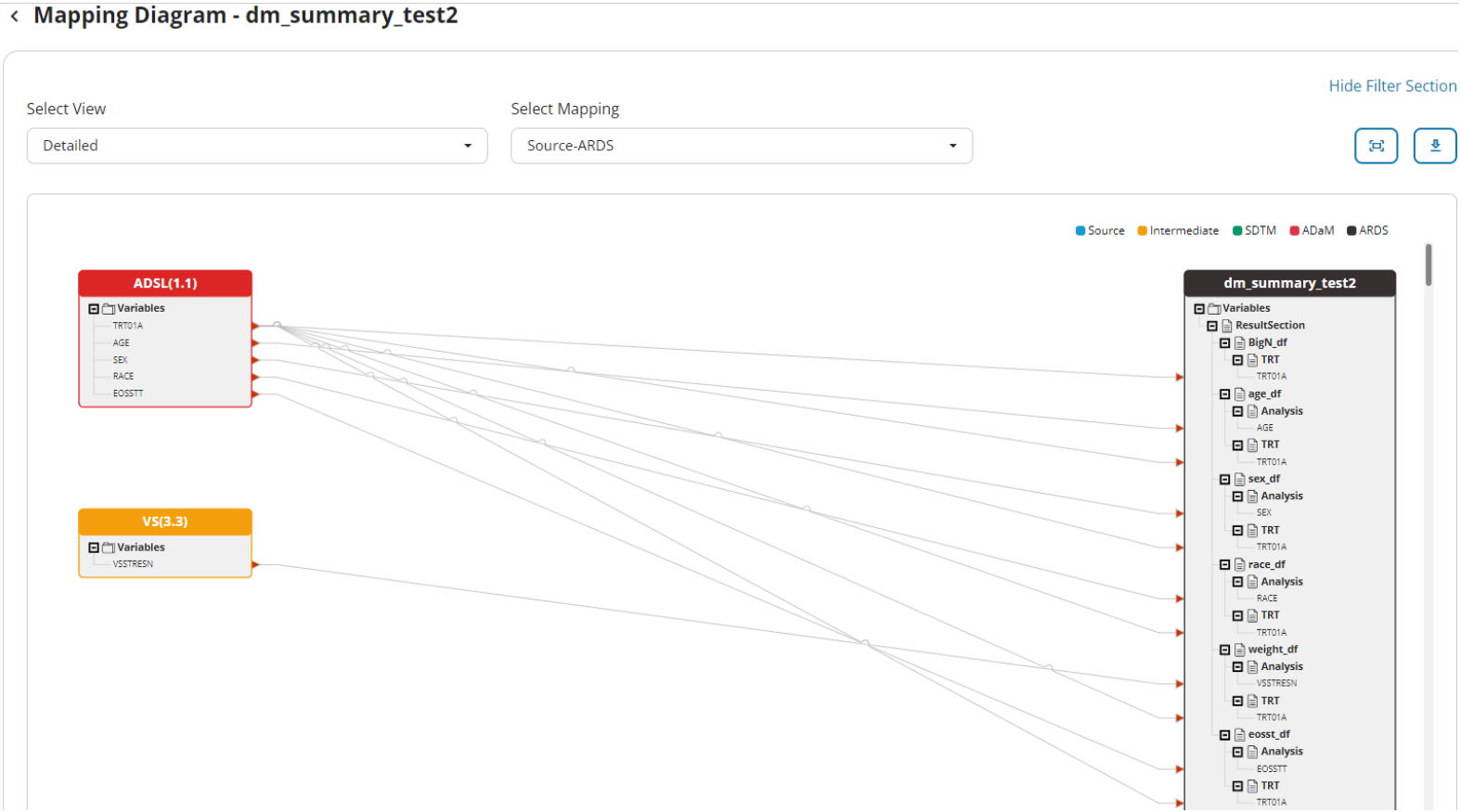
Modular approach enables increase of standardization with code from 70% in current reporting system to anticipated 90%

FileName	IsAutoCode	IncludeRowCount	Name	Order	DerivationType	TrtVar	TrtDisplay	DataSetTypeName	ResultStatisticsName	OverrideADSLMerge	SDTMVersionName	Description
dm_summary_test1	Yes	No	BigN_df	1	CountsFrequencies	TRT01A	TrtGroup1		BigN		3.3	Big N Treatment Counts
dm_summary_test1	Yes	No	age_df	2	SummaryStatistics	TRT01A	TrtGroup1	ADSL	SummStats1		3.3	Age (years)
dm_summary_test1	Yes	No	sex_df	3	CountsFrequencies	TRT01A	TrtGroup1	ADSL	Counts1		3.3	Gender, n (%)
dm_summary_test1	Yes	No	ethnic_df	4	CountsFrequencies	TRT01A	TrtGroup1	ADSL	Counts1		3.3	Ethnicity, n (%)
dm_summary_test1	Yes	No	race_df	5	CountsFrequencies	TRT01A	TrtGroup1	ADSL	Counts1		3.3	Race, n (%)
dm_summary_test1	Yes	No	weight_df	6	SummaryStatistics	TRT01A	TrtGroup1	VS	SummStats1		3.3	Weight (kg)
dm_summary_test2	Yes	No	BigN_df	1	CountsFrequencies	TRT01A	TrtGroup1		BigN		3.3	Big N Treatment Counts

ARDS Output File

POPFLAG	POPULATION	TRTN	TRT	CAT1VARLABEL	CAT1VAR	CAT1	CAT1N	CAT1LABEL	RESULTTYPE	RESULT
SAFFL	Safety Population	4	Trt D	Age (years)	AGE				n	141
SAFFL	Safety Population	4	Trt D	Age (years)	AGE				Std Dev	10.9300252
SAFFL	Safety Population	999	Total	Age (years)	AGE				Maximum	85
SAFFL	Safety Population	999	Total	Age (years)	AGE				Mean	53.98367347
SAFFL	Safety Population	999	Total	Age (years)	AGE				Median	55
SAFFL	Safety Population	999	Total	Age (years)	AGE				Minimum	18
SAFFL	Safety Population	999	Total	Age (years)	AGE				n	490
SAFFL	Safety Population	999	Total	Age (years)	AGE				Std Dev	11.27549321
SAFFL	Safety Population	2	Trt B	Gender, n (%)	SEX	F	1	Female	Count	59
SAFFL	Safety Population	4	Trt D	Gender, n (%)	SEX	F	1	Female	Count	115
SAFFL	Safety Population	3	Trt C	Gender, n (%)	SEX	F	1	Female	Count	116
SAFFL	Safety Population	1	Trt A	Gender, n (%)	SEX	F	1	Female	Count	114
SAFFL	Safety Population	999	Total	Gender, n (%)	SEX	F	1	Female	Count	404
SAFFL	Safety Population	2	Trt B	Gender, n (%)	SEX	M	2	Male	Count	10
SAFFL	Safety Population	4	Trt D	Gender, n (%)	SEX	M	2	Male	Count	26
SAFFL	Safety Population	3	Trt C	Gender, n (%)	SEX	M	2	Male	Count	25
SAFFL	Safety Population	1	Trt A	Gender, n (%)	SEX	M	2	Male	Count	25
SAFFL	Safety Population	999	Total	Gender, n (%)	SEX	M	2	Male	Count	86
SAFFL	Safety Population	2	Trt B	Gender, n (%)	SEX	F	1	Female	Percentage	85.50724638
SAFFL	Safety Population	4	Trt D	Gender, n (%)	SEX	F	1	Female	Percentage	81.56028369
SAFFL	Safety Population	3	Trt C	Gender, n (%)	SEX	F	1	Female	Percentage	82.26950355
SAFFL	Safety Population	1	Trt A	Gender, n (%)	SEX	F	1	Female	Percentage	82.01438849
SAFFL	Safety Population	999	Total	Gender, n (%)	SEX	F	1	Female	Percentage	82.44897959
SAFFL	Safety Population	2	Trt B	Gender, n (%)	SEX	M	2	Male	Percentage	14.49275362
SAFFL	Safety Population	4	Trt D	Gender, n (%)	SEX	M	2	Male	Percentage	18.43971631
SAFFL	Safety Population	3	Trt C	Gender, n (%)	SEX	M	2	Male	Percentage	17.73049645
SAFFL	Safety Population	1	Trt A	Gender, n (%)	SEX	M	2	Male	Percentage	17.98561151
SAFFL	Safety Population	999	Total	Gender, n (%)	SEX	M	2	Male	Percentage	17.55102041
SAFFL	Safety Population	2	Trt B	Ethnicity, n (%)	ETHNIC	MISSING	9999	Missing	Count	1
SAFFL	Safety Population	3	Trt C	Ethnicity, n (%)	ETHNIC	MISSING	9999	Missing	Count	1
SAFFL	Safety Population	999	Total	Ethnicity, n (%)	ETHNIC	MISSING	9999	Missing	Count	2
SAFFL	Safety Population	2	Trt B	Ethnicity, n (%)	ETHNIC	HISPANIC OR LATINO	1	HISPANIC OR LATINO	Count	32
SAFFL	Safety Population	4	Trt D	Ethnicity, n (%)	ETHNIC	HISPANIC OR LATINO	1	HISPANIC OR LATINO	Count	82

Dynamic traceability back through ADAM and SDTM to Data Collection



Thank You!

Keith Hibbetts: keith.hibbetts@lilly.com

Andy Miskell: andrew.miskell@lilly.com