PP12

Anatomy of a Table – Our Continued Journey to Automating Output Delivery Through Metadata

Edwin van Stein, GSK, The Netherlands

Principle

WORM – Write Once, Read Many

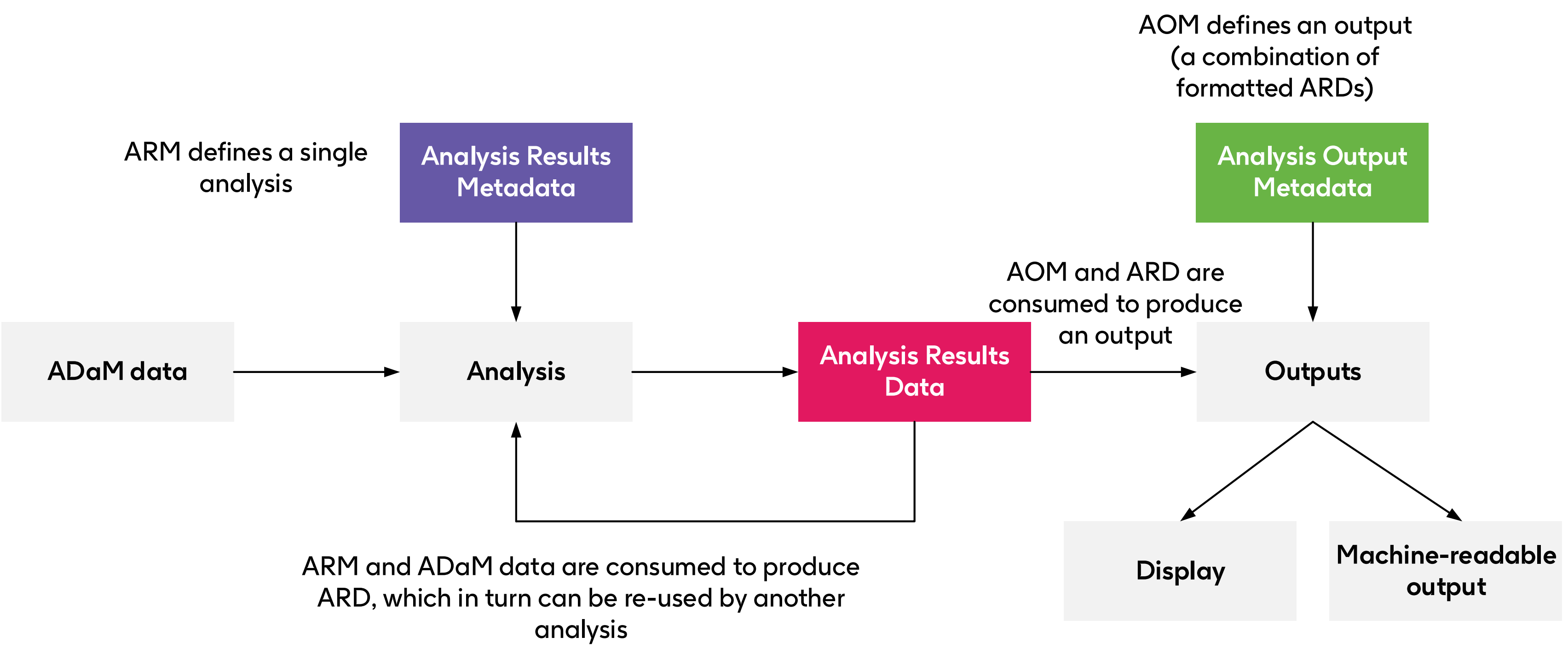
- Any analysis defined **once**
- Any analysis executed **once**
- Any analysis validate **once**
- **Re-use** analysis results across outputs
- **Re-use** analysis results across other analyses

Abstract

To improve traceability, transparency and consistency, GSK is implementing automation of analyses and output creation through standardised metadata. Tool-agnostic analysis results metadata (A highly-enriched adaptation of CDISC ARM) is used to produce analysis results data. Following the WORM principle (write once, read many), tool-agnostic analysis output metadata can then be used to render those results into one or more displays (tables, figures and listings, both post-text and in-text), other outputs (e.g. XML files for disclosure) and systems (in whatever format those systems require) without reproducing the numbers. This leads to a process where it's easier to validate and reuse results and to build tools in a multilingual setting.

This poster is a follow-up to my presentation from the PHUSE EU Connect 2023 and focuses on visualising the metadata for a single table and how it all fits together in our journey to standardising our results.

1. Vision



AOM defines an output (a combination of formatted ARDs)

ARM defines a single analysis

Analysis Results Metadata

Analysis Output Metadata

AOM and ARD are consumed to produce an output

ADaM data

Analysis

Analysis Results Data

Outputs

Display

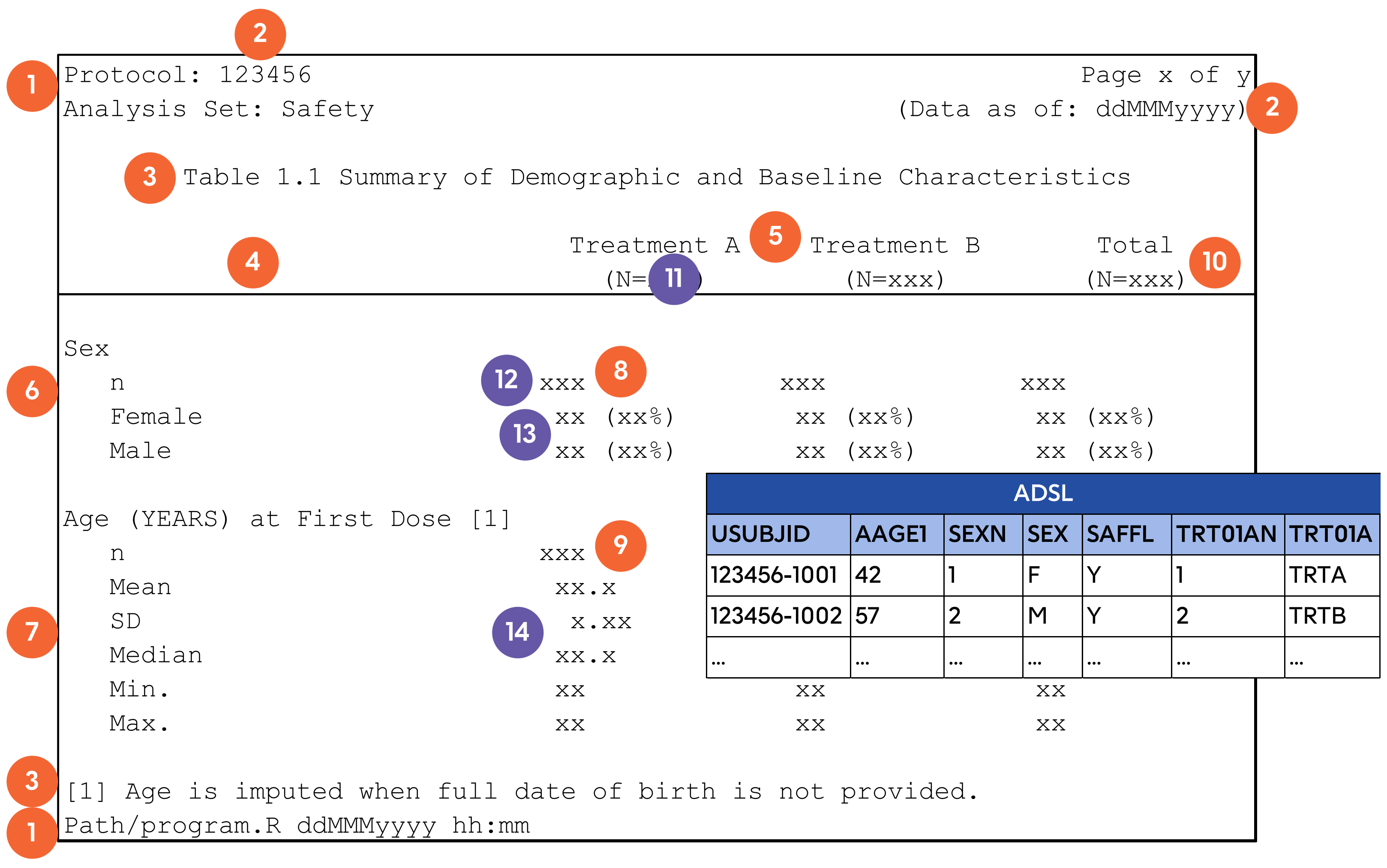
Machine-readable output

ARM and ADaM data are consumed to produce ARD, which in turn can be re-used by another analysis

All metadata that can be re-used across analyses is captured in **Analysis Concepts**. For instance:

- Safety analysis set is always the same WhereClause (SAFFL EQ "Y") with the same label (Safety)
- Actual treatment may be defined by TRTAN, TRTA and a corresponding codelist, where TRTAN is used for sorting and TRTA (or the codelist) for display purposes

2. Shell & Input



1
2
2
3
4
5
10
11
6
12
8
13
9
7
14
3
1

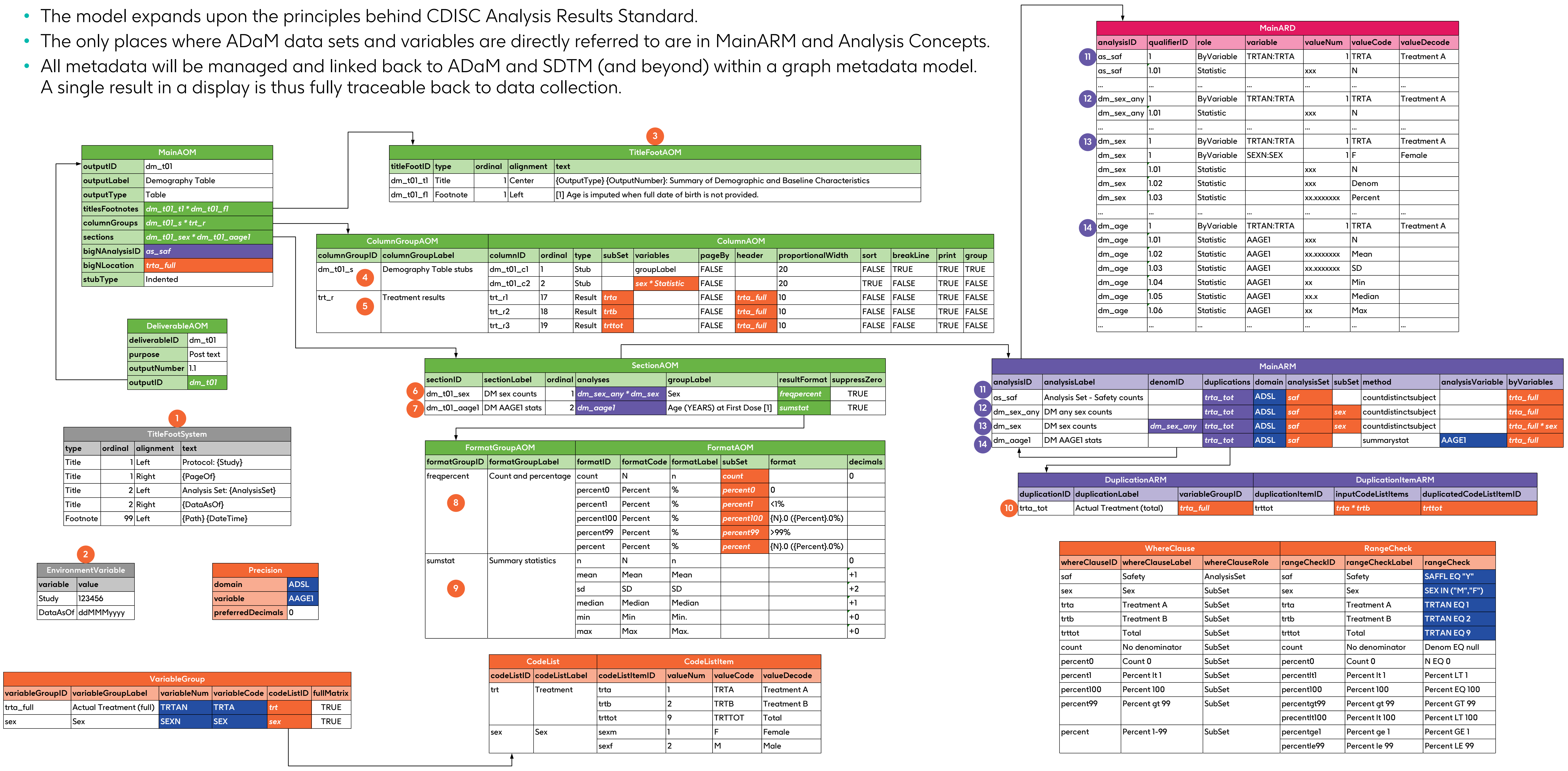
```
Protocol: 123456                                             Page x of y
Analysis Set: Safety                              (Data as of: ddMMMyyyy)

     Table 1.1 Summary of Demographic and Baseline Characteristics

                                  Treatment A   Treatment B       Total
                                     (N=   )       (N=xxx)      (N=xxx)

Sex
   n                                xxx             xxx             xxx
   Female                            xx  (xx%)       xx  (xx%)       xx  (xx%)
   Male                              xx  (xx%)       xx  (xx%)       xx  (xx%)

Age (YEARS) at First Dose [1]
   n                                xxx
   Mean                              xx.x
   SD                                 x.xx
   Median                            xx.x
   Min.                              xx              xx              xx
   Max.                              xx              xx              xx

[1] Age is imputed when full date of birth is not provided.
Path/program.R ddMMMyyyy hh:mm
```

ADSL

USUBJID	AAGE1	SEXN	SEX	SAFFL	TRT01AN	TRT01A
123456-1001	42	1	F	Y	1	TRTA
123456-1002	57	2	M	Y	2	TRTB
…	…	…	…	…	…	…

3. Metadata

- The model expands upon the principles behind CDISC Analysis Results Standard.
- The only places where ADaM data sets and variables are directly referred to are in MainARM and Analysis Concepts.
- All metadata will be managed and linked back to ADaM and SDTM (and beyond) within a graph metadata model. A single result in a display is thus fully traceable back to data collection.



MainAOM

outputID	dm_t01
outputLabel	Demography Table
outputType	Table
titlesFootnotes	dm_t01_t1 * dm_t01_f1
columnGroups	dm_t01_s * trt_r
sections	dm_t01_sex * dm_t01_aage1
bigNAnalysisID	as_saf
bigNLocation	trta_full
stubType	Indented

TitleFootAOM

titleFootID	type	ordinal	alignment	text
dm_t01_t1	Title	1	Center	{OutputType} {OutputNumber}: Summary of Demographic and Baseline Characteristics
dm_t01_f1	Footnote	1	Left	[1] Age is imputed when full date of birth is not provided.

ColumnGroupAOM / ColumnAOM

columnGroupID	columnGroupLabel	columnID	ordinal	type	subSet	variables	pageBy	header	proportionalWidth	sort	breakLine	print	group
dm_t01_s	Demography Table stubs	dm_t01_c1	1	Stub		groupLabel	FALSE		20	FALSE	TRUE	TRUE	TRUE
		dm_t01_c2	2	Stub		sex * Statistic	FALSE		20	TRUE	FALSE	TRUE	FALSE
trt_r	Treatment results	trt_r1	17	Result	trta		FALSE	trta_full	10	FALSE	FALSE	TRUE	FALSE
		trt_r2	18	Result	trtb		FALSE	trta_full	10	FALSE	FALSE	TRUE	FALSE
		trt_r3	19	Result	trttot		FALSE	trta_full	10	FALSE	FALSE	TRUE	FALSE

DeliverableAOM

deliverableID	dm_t01
purpose	Post text
outputNumber	1.1
outputID	dm_t01

SectionAOM

sectionID	sectionLabel	ordinal	analyses	groupLabel	resultFormat	suppressZero
dm_t01_sex	DM sex counts	1	dm_sex_any * dm_sex	Sex	freqpercent	TRUE
dm_t01_aage1	DM AAGE1 stats	2	dm_aage1	Age (YEARS) at First Dose [1]	sumstat	TRUE

TitleFootSystem

type	ordinal	alignment	text
Title	1	Left	Protocol: {Study}
Title	1	Right	{PageOf}
Title	2	Left	Analysis Set: {AnalysisSet}
Title	2	Right	{DataAsOf}
Footnote	99	Left	{Path} {DateTime}

EnvironmentVariable

variable	value
Study	123456
DataAsOf	ddMMMyyyy

Precision

domain	ADSL
variable	AAGE1
preferredDecimals	0

FormatGroupAOM / FormatAOM

formatGroupID	formatGroupLabel	formatID	formatCode	formatLabel	subSet	format	decimals
freqpercent	Count and percentage	count	N	n	count		0
		percent0	Percent	%	percent0	0	
		percent1	Percent	%	percent1	<1%	
		percent100	Percent	%	percent100	{N}.0 ({Percent}.0%)	
		percent99	Percent	%	percent99	>99%	
		percent	Percent	%	percent	{N}.0 ({Percent}.0%)	
sumstat	Summary statistics	n	N	n			0
		mean	Mean	Mean			+1
		sd	SD	SD			+2
		median	Median	Median			+1
		min	Min	Min.			+0
		max	Max	Max.			+0

VariableGroup

variableGroupID	variableGroupLabel	variableNum	variableCode	codeListID	fullMatrix
trta_full	Actual Treatment (full)	TRTAN	TRTA	trt	TRUE
sex	Sex	SEXN	SEX	sex	TRUE

CodeList / CodeListItem

codeListID	codeListLabel	codeListItemID	valueNum	valueCode	valueDecode
trt	Treatment	trta	1	TRTA	Treatment A
		trtb	2	TRTB	Treatment B
		trttot	9	TRTTOT	Total
sex	Sex	sexm	1	F	Female
		sexf	2	M	Male

MainARD

analysisID	qualifierID	role	variable	valueNum	valueCode	valueDecode
as_saf	1	ByVariable	TRTAN:TRTA	1	TRTA	Treatment A
as_saf	1.01	Statistic		xxx	N	
…	…	…	…	…	…	…
dm_sex_any	1	ByVariable	TRTAN:TRTA	1	TRTA	Treatment A
dm_sex_any	1.01	Statistic		xxx	N	
…	…	…	…	…	…	…
dm_sex	1	ByVariable	TRTAN:TRTA	1	TRTA	Treatment A
dm_sex	1	ByVariable	SEXN:SEX	1	F	Female
dm_sex	1.01	Statistic		xxx	N	
dm_sex	1.02	Statistic		xxx	Denom	
dm_sex	1.03	Statistic		xx.xxxxxxx	Percent	
…	…	…	…	…	…	…
dm_age	1	ByVariable	TRTAN:TRTA	1	TRTA	Treatment A
dm_age	1.01	Statistic	AAGE1	xxx	N	
dm_age	1.02	Statistic	AAGE1	xx.xxxxxxx	Mean	
dm_age	1.03	Statistic	AAGE1	xx.xxxxxxx	SD	
dm_age	1.04	Statistic	AAGE1	xx	Min	
dm_age	1.05	Statistic	AAGE1	xx.x	Median	
dm_age	1.06	Statistic	AAGE1	xx	Max	
…	…	…	…	…	…	…

MainARM

analysisID	analysisLabel	denomID	duplications	domain	analysisSet	subSet	method	analysisVariable	byVariables
as_saf	Analysis Set - Safety counts		trta_tot	ADSL	saf		countdistinctsubject		trta_full
dm_sex_any	DM any sex counts		trta_tot	ADSL	saf	sex	countdistinctsubject		trta_full
dm_sex	DM sex counts	dm_sex_any	trta_tot	ADSL	saf	sex	countdistinctsubject		trta_full * sex
dm_aage1	DM AAGE1 stats		trta_tot	ADSL	saf		summarystat	AAGE1	trta_full

DuplicationARM / DuplicationItemARM

duplicationID	duplicationLabel	variableGroupID	duplicationItemID	inputCodeListItems	duplicatedCodeListItemID
trta_tot	Actual Treatment (total)	trta_full	trttot	trta * trtb	trttot

WhereClause / RangeCheck

whereClauseID	whereClauseLabel	whereClauseRole	rangeCheckID	rangeCheckLabel	rangeCheck
saf	Safety	AnalysisSet	saf	Safety	SAFFL EQ "Y"
sex	Sex	SubSet	sex	Sex	SEX IN ("M","F")
trta	Treatment A	SubSet	trta	Treatment A	TRTAN EQ 1
trtb	Treatment B	SubSet	trtb	Treatment B	TRTAN EQ 2
trttot	Total	SubSet	trttot	Total	TRTAN EQ 9
count	No denominator	SubSet	count	No denominator	Denom EQ null
percent0	Count 0	SubSet	percent0	Count 0	N EQ 0
percent1	Percent lt 1	SubSet	percentlt1	Percent lt 1	Percent LT 1
percent100	Percent 100	SubSet	percent100	Percent 100	Percent EQ 100
percent99	Percent gt 99	SubSet	percentgt99	Percent gt 99	Percent GT 99
			precentlt100	Percent lt 100	Percent LT 100
percent	Percent 1-99	SubSet	percentge1	Percent ge 1	Percent GE 1
			percentle99	Percent le 99	Percent LE 99

Abbreviations

ADaM	Analysis Data Model
ADSL	ADaM Subject-level Analysis
AOM	Analysis Output Metadata
ARD	Analysis Results Data
ARM	Analysis Results Metadata
ARS	Analysis Results Standard
CDISC	Clinical Data Interchange Standards Consortium
SDTM	Study Data Tabulation Model
WORM	Write Once, Read Many
XML	Extensible Markup Language

Acknowledgements

This metadata model is designed by the author, but heavily relies on previous and ongoing input by and discussions with the following people (in alphabetical order): Alexey Kuznetsov, Dean Grundy, Fabian Noelle, Grzegorz Lach, Jeremy Teoh, Roza Antonyan, Stephen Law, and Warwick Benger.

Author LinkedIn
for comments
and questions

Presenting author: Edwin van Stein, edwin.j.stein@gsk.com