



Automated Code Generation: Evaluating Deterministic and Probabilistic approaches

Bloemfontein PHUSE SDE

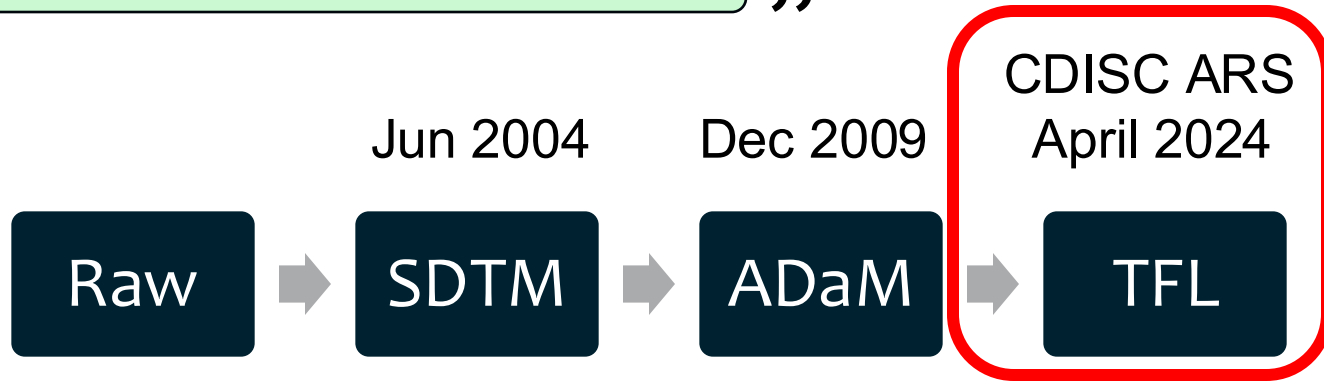
7 November 2025

Malan Bosman, Lead Data Automation Engineer

Agile!

“ The clinical development industry is experiencing rapid change, which presents opportunities for dynamic and positive transformation of our procedures and processes ”

ChatGPT 3.5
November 2022



Vision: Automated code generation



CDISC Analysis Result Standards – Released April 19, 2024!



Analysis Results Standard (ARS) v1.0



Large trials generate many analysis results in the form of tables, figures, and written reports, yet these results are rarely output in a form that is machine-readable. Previously, there has been no standard way of describing and organizing these results, making it difficult to automate their generation, make them reproducible, trace their origin, or enable them to be reused in other outputs.

To address these inefficiencies, CDISC has developed the [Analysis Results Standard \(ARS\)](https://cdisc-org.github.io/analysis-results-standard/), which aim to facilitate automation, reproducibility, reusability, and traceability of analysis results data.

Features of ARS v1.0

- A Logical Data Model that describes analysis results and associated metadata.
- A User Guide to illustrate and exercise the model with common safety displays.

<https://cdisc-org.github.io/analysis-results-standard/>

Analysis Results Standard (ARS)

Logical model to support both the prospective specification of analyses and the fully contextualized representation of the results of the analyses.

URI: <https://www.cdisc.org/ars/1-0> Name: ars_idm

Schema Diagram

Classes

Classes provide templates for organizing data. Data objects instantiate class slots (aka fields, attributes) that are applicable to it. See [LinkML documentation](#).

Class	Description
NamedObject	An object with a name
ReportingEvent	A set of analyses and outputs created for a specific requirement...
ListOfContents	A structured list of analyses and outputs

cdisc

Analysis Results Standard User Guide

Version 1.0 (Final)

Prepared by the
Analysis Results Standard Team

Notes to Readers

- This is the final Version 1.0 of the Analysis Results Standard User Guide.
- This document is based on ADaM v2.1 and Analysis Results Metadata (ARM) v1.0 for Define-XML v2.0

Revision History

Date	Version
2024-04-19	Final

© 2024 Clinical Data Interchange Standards Consortium, Inc. All rights reserved.

<https://wiki.cdisc.org/display/ARSP/Analysis+Results+Standard+User+Guide+v1.0>

CDISC ARS Metadata

Analysis Set

Data Subset

Analysis Grouping

Analysis Method

Summary of Demographics

Study - CDISC 360 Page x of y

Table 14.1.1
Summary of Demographics
Safety Population

Characteristics	Placebo (N=XX)	Xanomeline Low Dose (N=XX)	Xanomeline High Dose (N=XX)
Age (years)			
n	XX	XX	XX
Mean (SD)	XX.X (XX.XX)	XX.X (XX.XX)	XX.X (XX.XX)
Median	XX.X	XX.X	XX.X
Q1, Q3	XX.X, XX.X	XX.X, XX.X	XX.X, XX.X
Min, Max	XX, XX	XX, XX	XX, XX
Age Group, n (%)			
< 65 years	XX (XX.X)	XX (XX.X)	XX (XX.X)
≥ 65 years	XX (XX.X)	XX (XX.X)	XX (XX.X)
Gender, n (%)			
Male	XX (XX.X)	XX (XX.X)	XX (XX.X)
Female	XX (XX.X)	XX (XX.X)	XX (XX.X)
Ethnicity, n (%)			
Hispanic or Latino	XX (XX.X)	XX (XX.X)	XX (XX.X)
Not Hispanic or Latino	XX (XX.X)	XX (XX.X)	XX (XX.X)

Source dataset: adsl, Generated on: DDMONYYYY:HH:MM
Program: <pid>.sas, Output: <pid><oid>.rtf, Generated on: DDMONYYYY:HH:MM

Summary of TEAE by SOC and PT

Study - CDISC 360 Page x of y

Table 14.3.1.1
Summary of TEAE by System Organ Class and Preferred Term
Safety Population

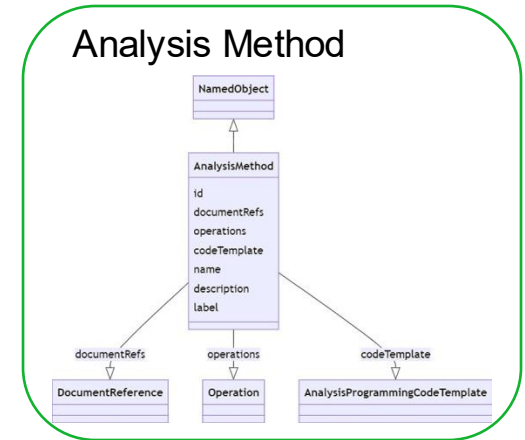
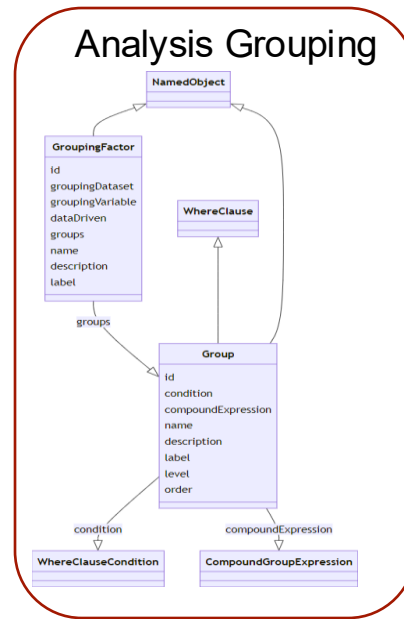
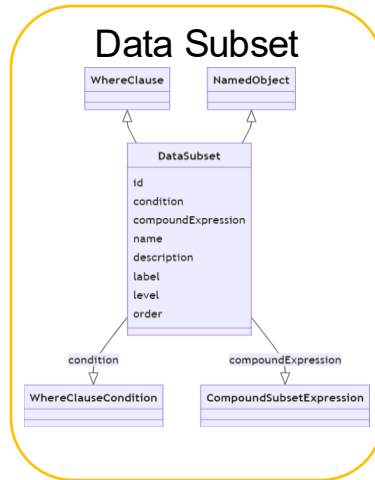
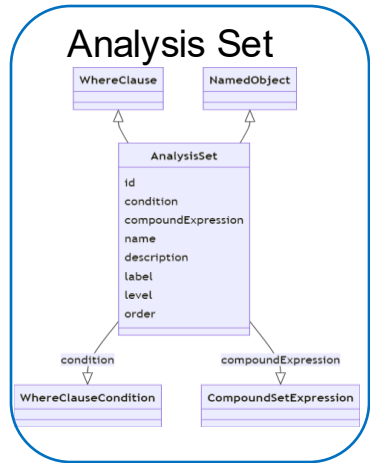
System Organ Class Preferred Term [a], n (%)	Placebo (N=XX)	Xanomeline Low Dose (N=XX)	Xanomeline High Dose (N=XX)
Number of subjects with at least one event	XX (XX.X)	XX (XX.X)	XX (XX.X)
<SOC 1>	XX (XX.X)	XX (XX.X)	XX (XX.X)
<Preferred Term 1>	XX (XX.X)	XX (XX.X)	XX (XX.X)
...	XX (XX.X)	XX (XX.X)	XX (XX.X)
<Preferred Term n>	XX (XX.X)	XX (XX.X)	XX (XX.X)
<SOC 2>	XX (XX.X)	XX (XX.X)	XX (XX.X)
<Preferred Term 1>	XX (XX.X)	XX (XX.X)	XX (XX.X)
...	XX (XX.X)	XX (XX.X)	XX (XX.X)
<Preferred Term n>	XX (XX.X)	XX (XX.X)	XX (XX.X)

Notes: TEAE=Treatment-Emergent Adverse Events.
Subjects are counted once within each system organ class and preferred term.
[a] All investigators adverse events were coded using MedDRA version xx.x.

Source dataset: adae, Generated on: DDMONYYYY:HH:MM
Program: <pid>.sas, Output: <pid><oid>.rtf, Generated on: DDMONYYYY:HH:MM

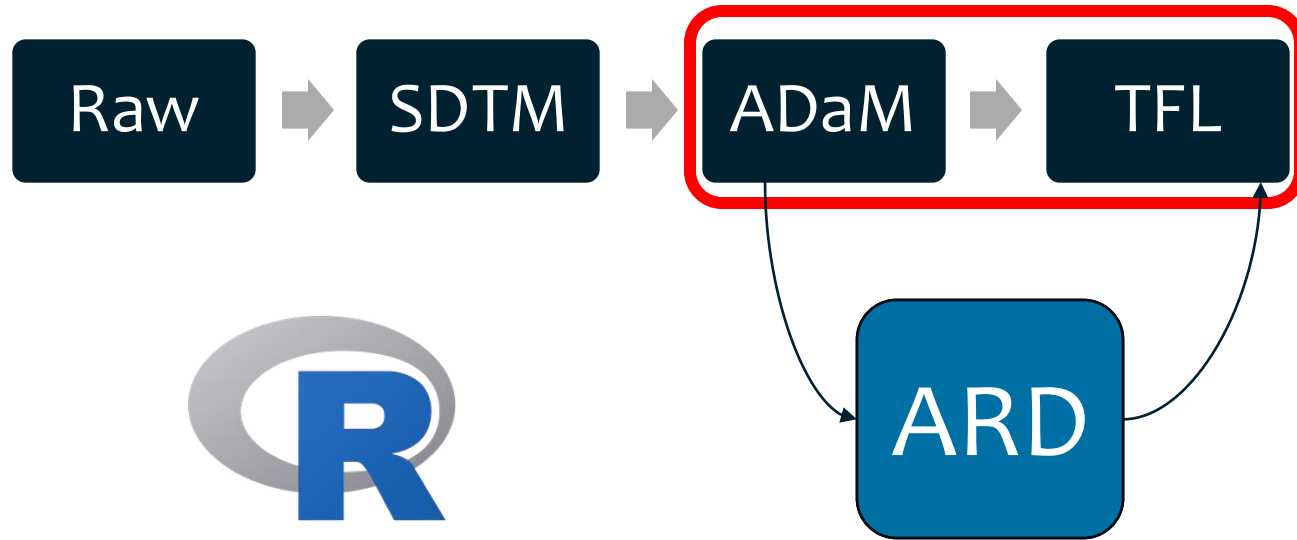
CDISC ARS Metadata

One **Output** has many **Analyses**. One **Analysis** has the following:



One **Analysis Method** consists of one or more **Operations**

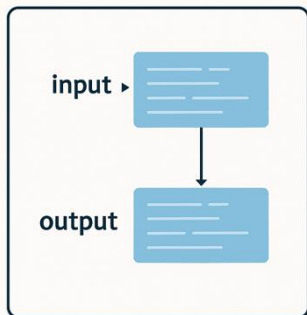
Scope



Approaches

Deterministic

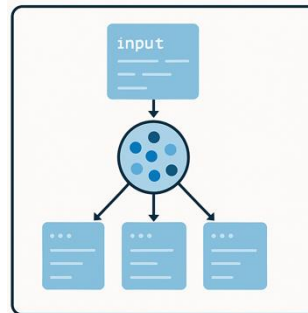
- Think: Macros, functions, logic



- Explicit, rule-based logic
- Same input -> same output
- No randomness; high reproducibility

Probabilistic

- Think: Generative AI



- Statistical inference
- Same input -> possible variation in output
- Prediction-based; introduces randomness

What is siera?

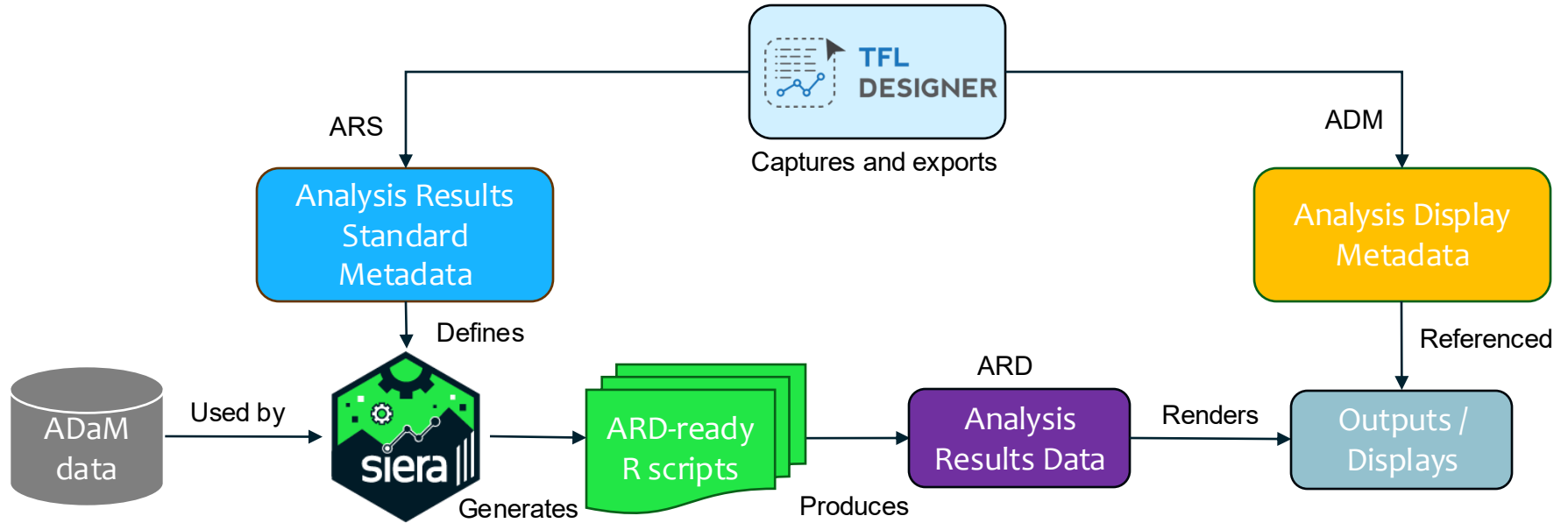
siera* is an open-source **R package** that **ingests ARS metadata** for a reporting event and converts the metadata to **produce executable R scripts**. These R scripts can be run as-is (combined with the relevant ADaM dataset) to **produce each output's ARD**.

<https://github.com/clymbclinical/siera>

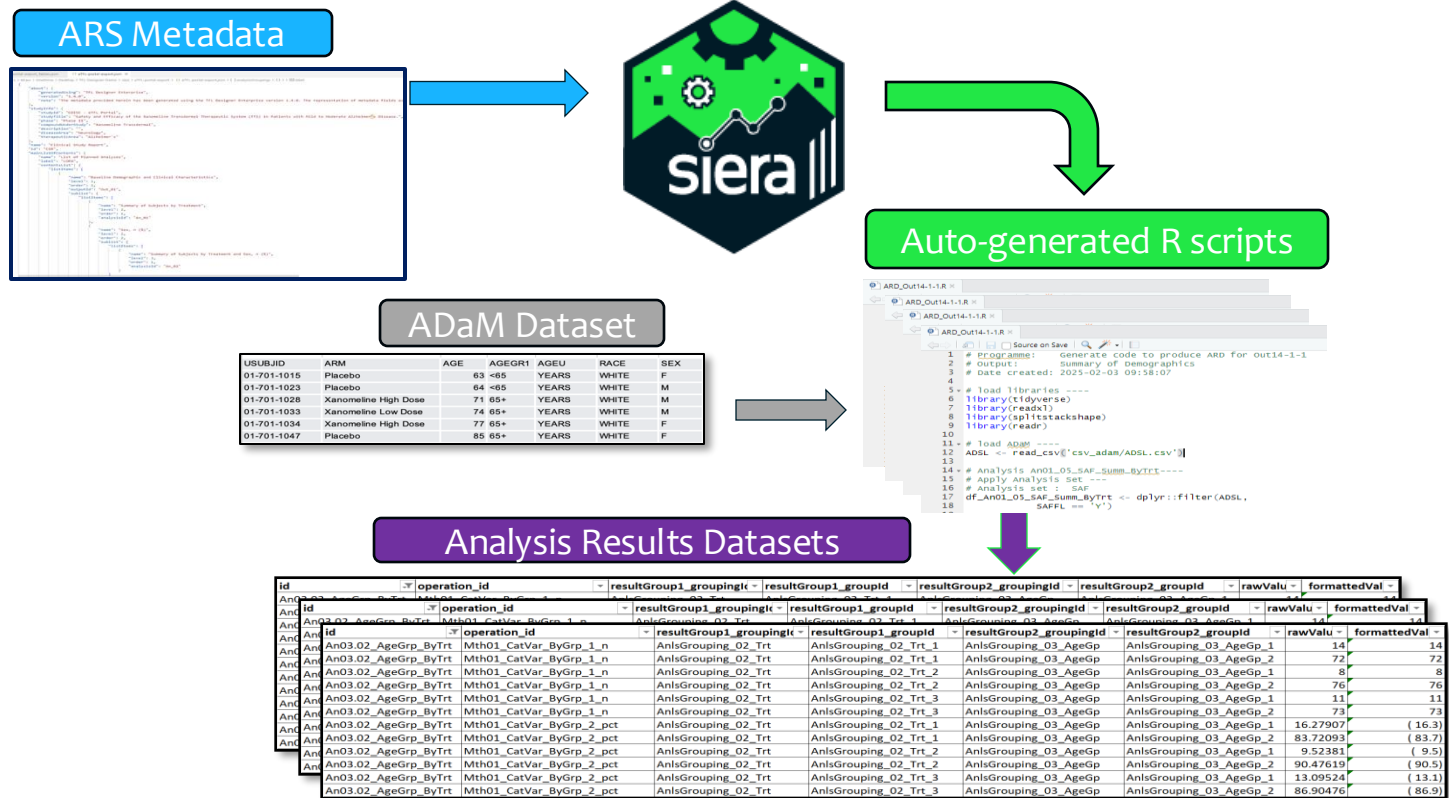


* on CRAN and approved by COSA!

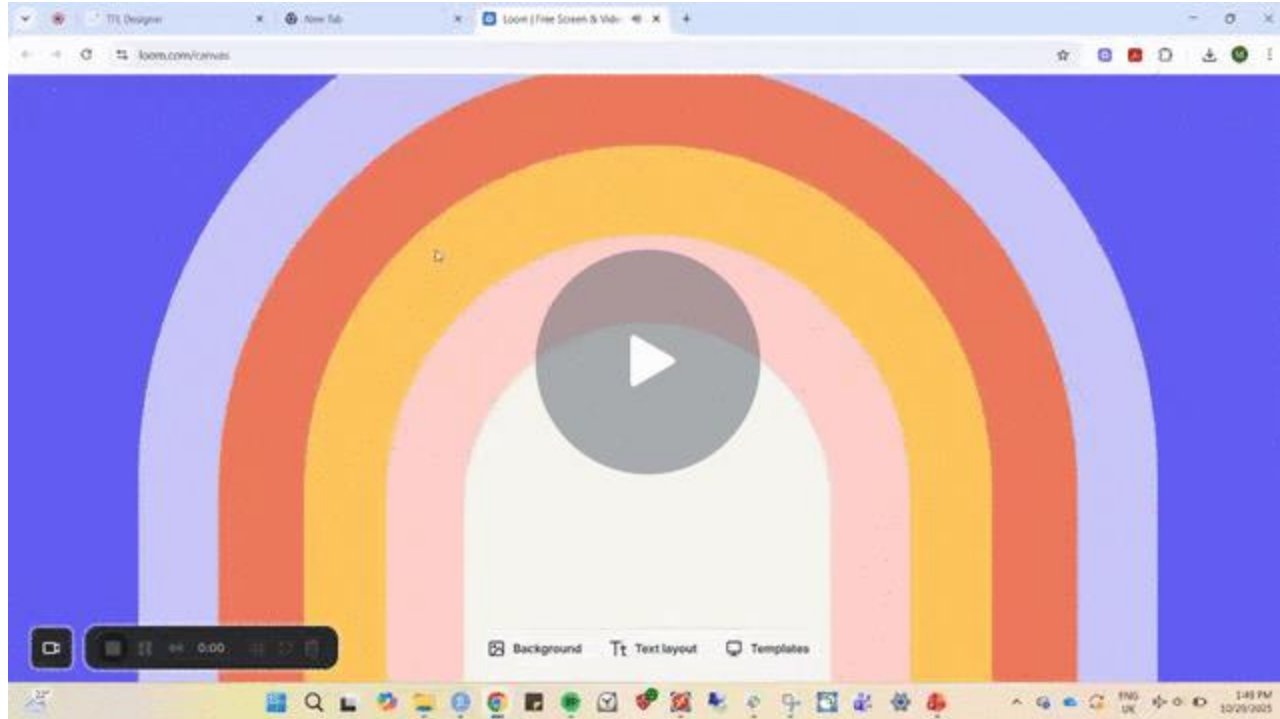
Evaluation: Deterministic approach



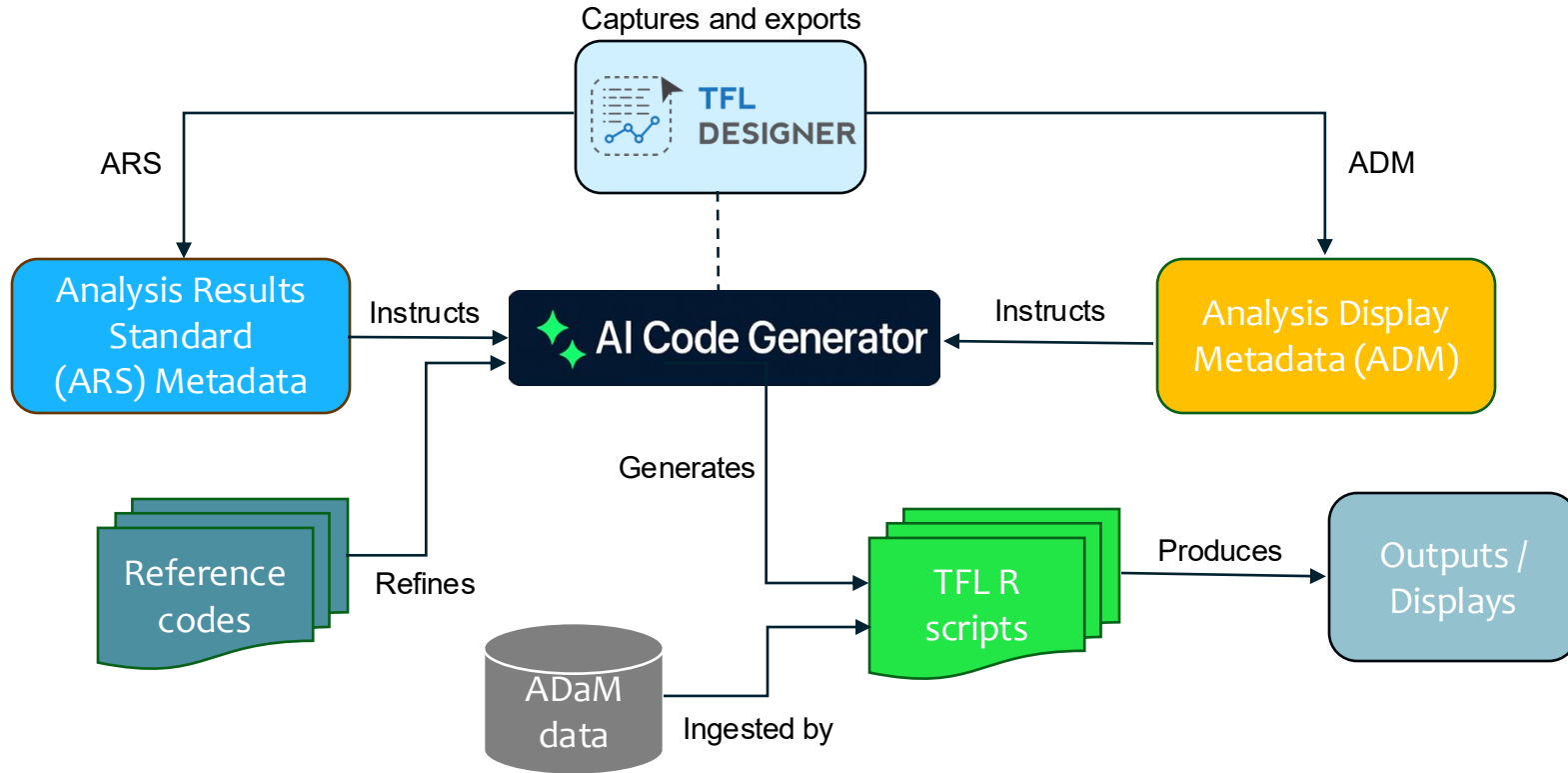
Evaluation: Deterministic approach



Evaluation: Deterministic approach - video



Evaluation: Probabilistic approach



Evaluation: Probabilistic approach

[TFL Designer - 29 October 2025 - Watch Video](#)

TFL DESIGNER File Edit View Help Neurology / Alzheimer's / Bloem SDE 2025

Table of Contents: Baseline Demographic and Clinical Characteristics

Baseline Demographic and Clinical Characteristics

Safety Population

Characteristics	Xanomeline Low Dose (N=XX) n (%)	Xanomeline High Dose (N=XX) n (%)	Placebo (N=XX) n (%)	Total Population (N=XX) n (%)
Sex, n (%)				
Female	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)
Male	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)
Age, Years				
XX	XX	XX	XX	XX
XX.X (XX.XX)	XX.X (XX.XX)	XX.X (XX.XX)	XX.X (XX.XX)	XX.X (XX.XX)

Properties Outline

Connected Study: ☒ **Female**

Where Clause for Female: (ADSLASEX EQ Female)

Result Properties

Result Description: Sex, n (%)

Dataset: **ADSL**

Variable: **ADSLASEX**

Where Clause for Sex, n (%): **Configure**

Advanced

Statistics 1

Findings: Implementation requirements

Deterministic

ARS metadata

Functions / macros
for statistical
operations

Internal processes
for transforming
ARD to display

Probabilistic

ARS + ADM
metadata

Reference code
for statistical
operations

Internal processes
for human-in-the-
loop

Findings: Workability

Deterministic

Predictable
results

Rigid setup

Easy to
debug

Probabilistic

Prone to
variation

Flexible
setup

Debugging
time-
consuming

Findings: Success factors

Deterministic

Good understanding of ARS

Strong test suite for functions/ macro's

Well-defined Internal standards

Probabilistic

Thoughtful prompt design

Feedback loop

Quality reference code



Malan Bosman

Lead Data Automation Engineer

mbosman@clymbclinical.com



EMAIL

info@clymbclinical.com



WEBSITE

www.clymbclinical.com



PHONE

781-692-3613