



The Clinical Data
Science Conference

US 2025

16–19 March

Renaissance Orlando at SeaWorld Florida

Advancing TFL Automation: R-Based Meta-Programming Powered by ARS

Paper & Presentation # OS20,

PHUSE US Connect 2025

March 19th, 2025

Malan Bosman, Lead Data Automation Engineer

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/>

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

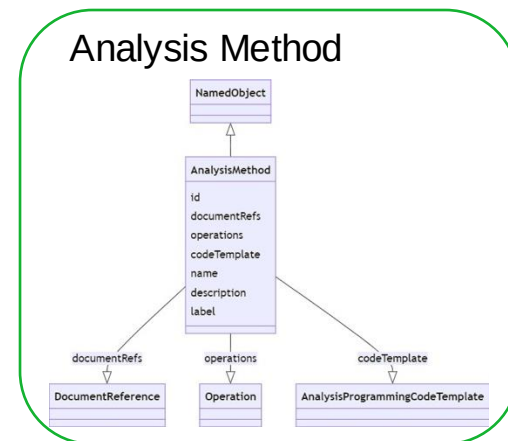
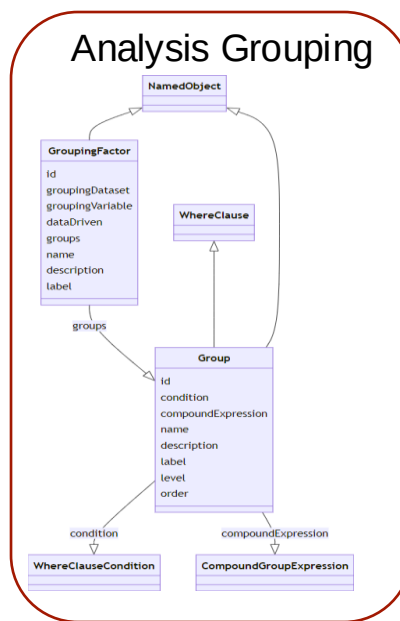
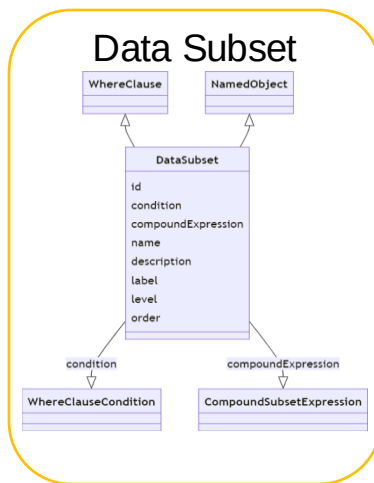
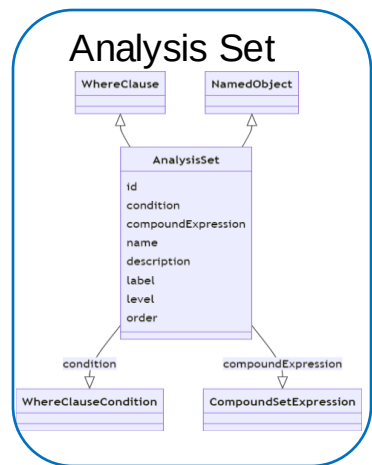
Date	Version
2024-04-19	Final

© 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

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



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

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

Table 14.3.1.1

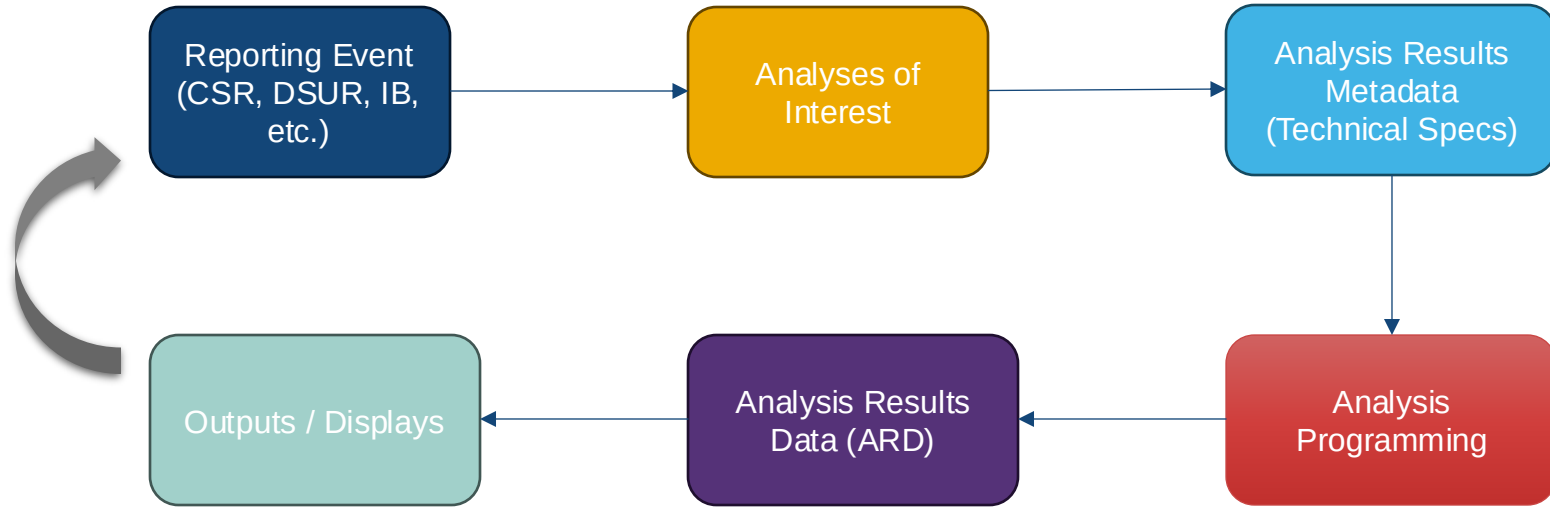
Page x of y

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

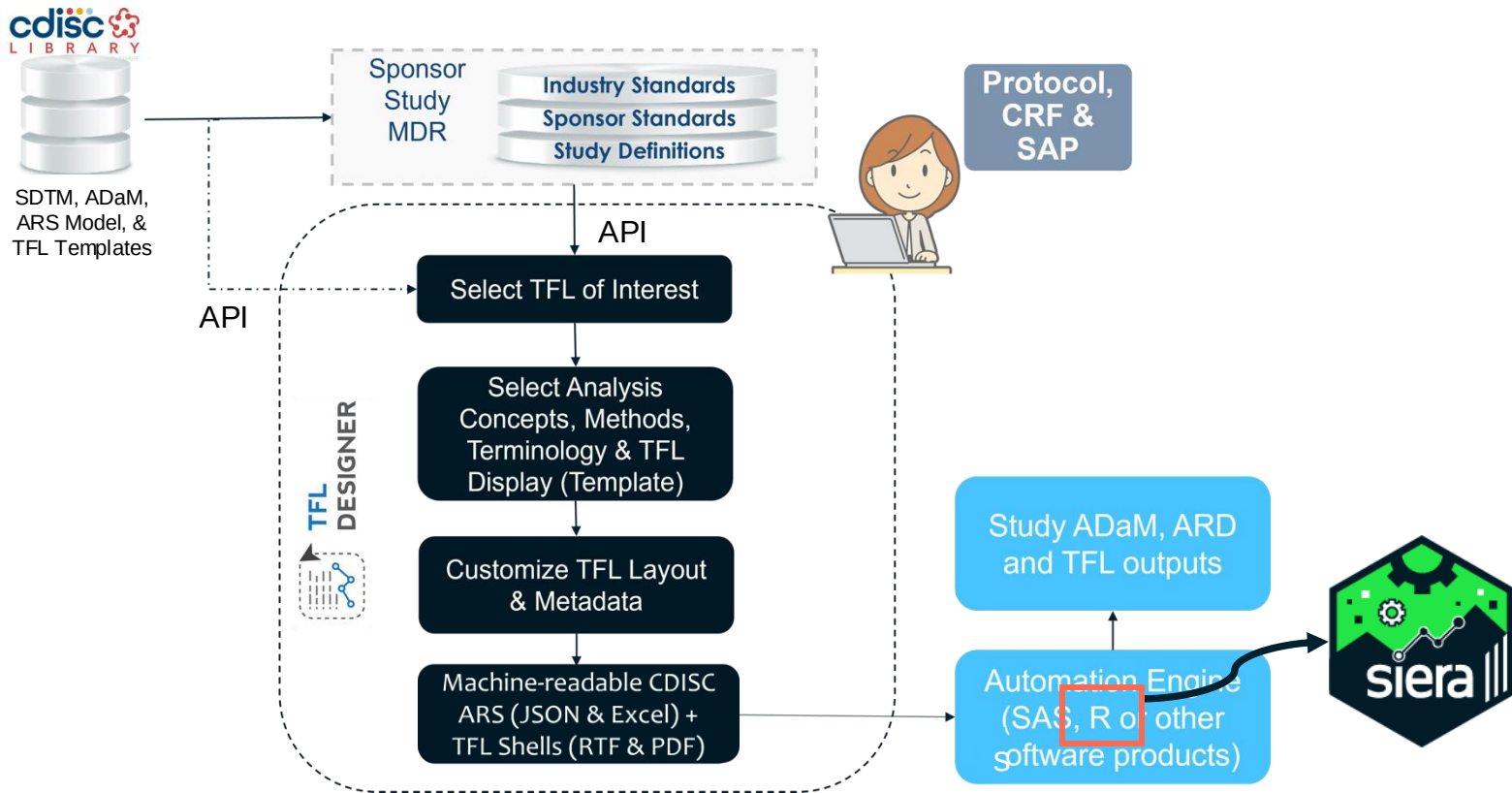
System Organ Class Preferred Term [a], n (%)	Placebo (N=XX)	Xanomeline Low Dose (N=XX)	Xanomeline High Dose (N=XX)
Number of subjects with at least one event	XX (XX.X)	XX (XX.X)	XX (XX.X)
<SOC 1>	XX (XX.X)	XX (XX.X)	XX (XX.X)
<Preferred Term 1>	XX (XX.X)	XX (XX.X)	XX (XX.X)
...	XX (XX.X)	XX (XX.X)	XX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

ARS Model Supported Workflow and Entry Points



Metadata-driven TFL Deliverable Workflow



What is siera?

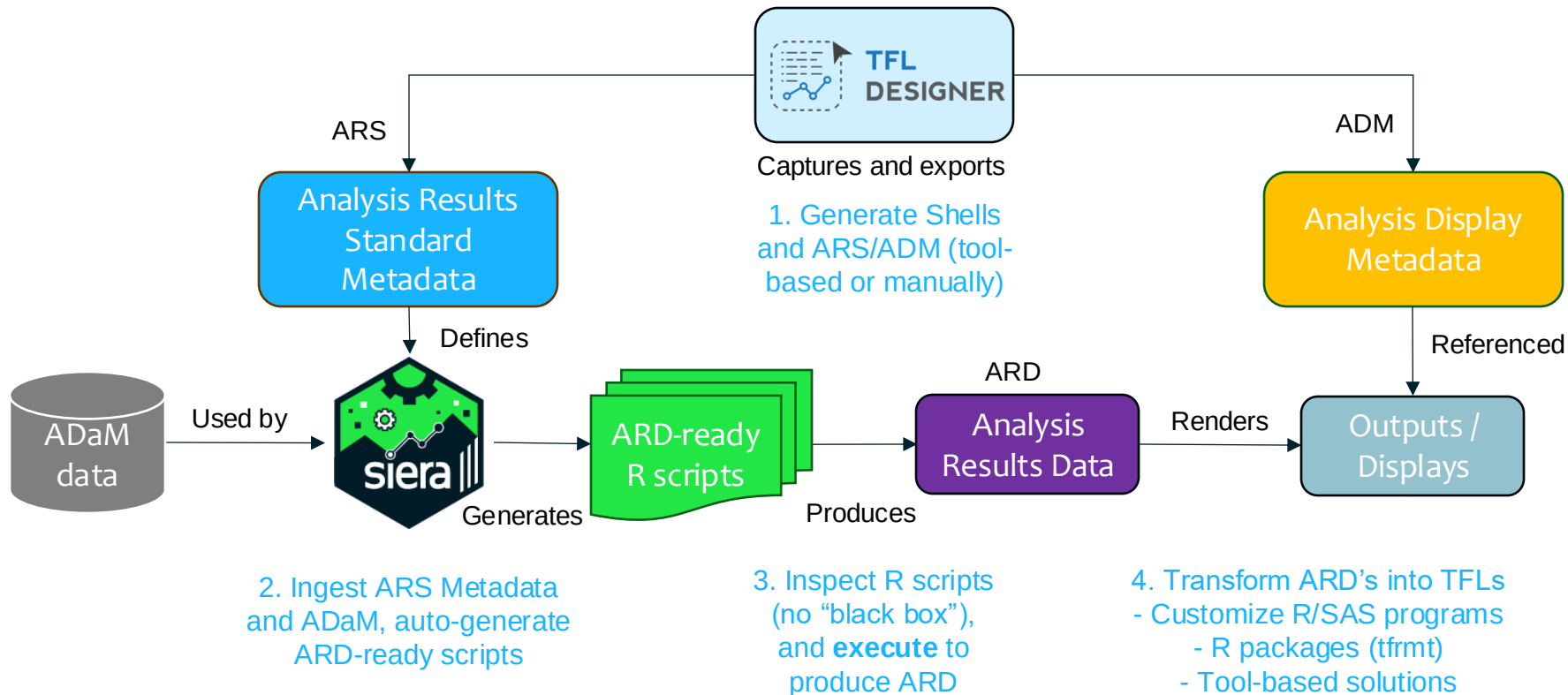
siera* is an open-source **R package** that **ingests ARS metadata** for a reporting event and converts the metadata to **produce executable R scripts**. These R scripts can be run as-is (combined with the relevant ADaM dataset) to **produce each output's ARD**.

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

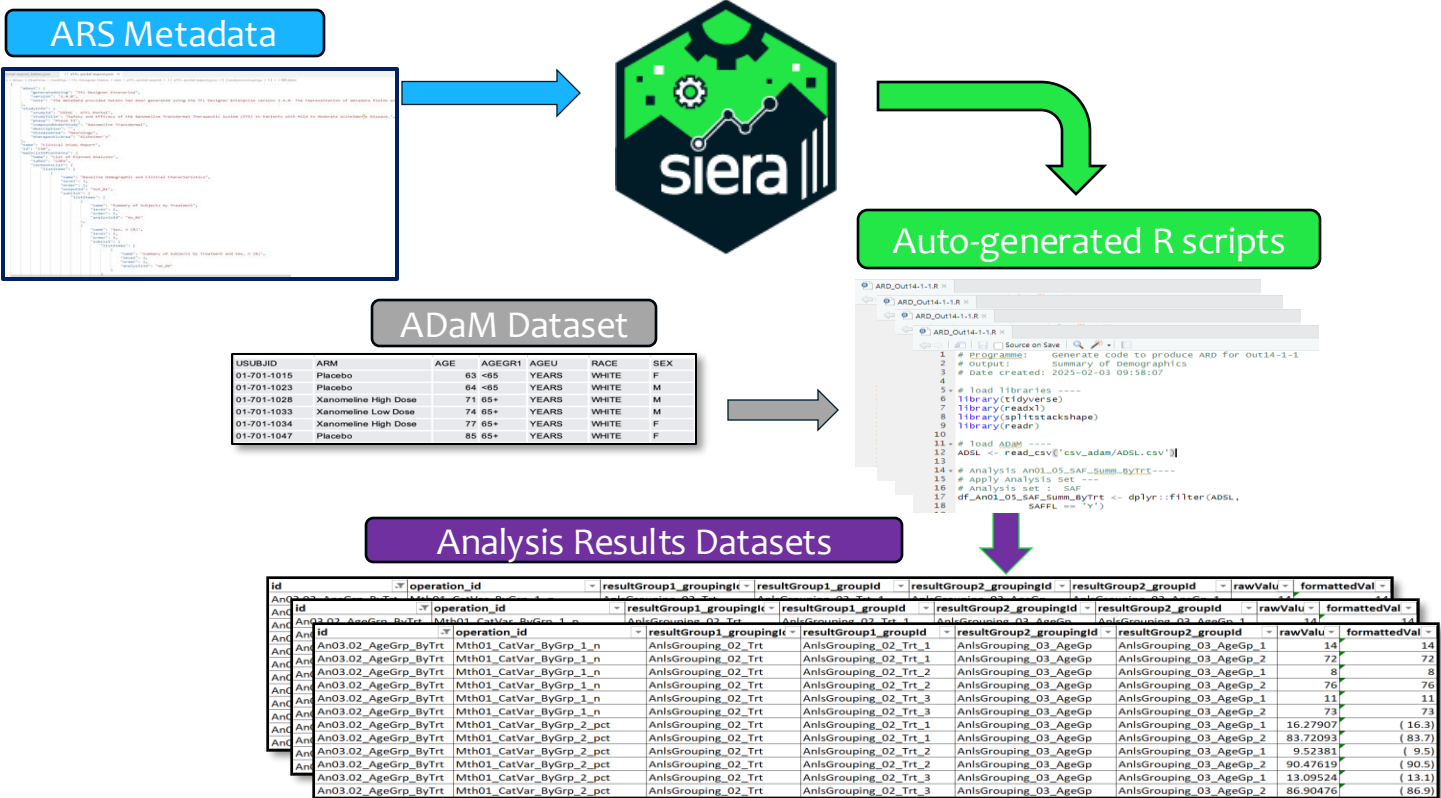


* on CRAN and approved by COSA!

siera Workflow [Design to Execution]



siera Workflow [ARS Metadata to ARD]



ARS Metadata [from TFL Designer]: JSON or Excel

JSON

```
portal-export-Tables.json 1 eTFL-portal-export.json x
> bhavi > OneDrive > Desktop > TFL-Designer-Demo > Jazz > eTFL-portal-export > 1 eTFL-portal-export.json > 1 analysisGroupings > 1 1 > label
{
  "about": {
    "generatedUsing": "TFL Designer Enterprise",
    "version": "1.4.0",
    "note": "The metadata provided herein has been generated using the TFL Designer Enterprise version 1.4.0. The representation of metadata fields are as follows."
  },
  "studyInfo": {
    "studyId": "CDISC - eTFL Portal",
    "studyTitle": "Safety and Efficacy of the Xanomeline Transdermal Therapeutic System (TTS) in Patients with Mild to Moderate Alzheimer's Disease.",
    "phase": "Phase II",
    "compoundUnderStudy": "Xanomeline Transdermal",
    "description": "",
    "diseaseArea": "Neurology",
    "therapeuticArea": "Alzheimer's"
  },
  "name": "Clinical Study Report",
  "id": "CSR",
  "mainListofContents": {
    "name": "List of Planned Analyses",
    "label": "LOPA",
    "contentList": {
      "listItems": [
        {
          "name": "Baseline Demographic and Clinical Characteristics",
          "level": 1,
          "order": 1,
          "outputId": "Out_01",
          "subList": {
            "listItems": [
              {
                "name": "Summary of Subjects by Treatment",
                "level": 2,
                "order": 1,
                "analysisId": "An_01"
              },
              {
                "name": "Sex, n (%)",
                "level": 2,
                "order": 2,
                "subList": {
                  "listItems": [
                    {
                      "name": "Summary of Subjects by Treatment and Sex, n (%)"
                    }
                  ]
                }
              }
            ]
          }
        }
      ]
    }
  }
}
```

Excel

1	id	version	name
2	An_01	1	Summary of Subjects by Treatment
3	An_02	1	Summary of Subjects by Treatment and Sex, n (%)
4	An_02_Total	1	Summary of Subjects by Sex, n (%)
5	An_03	1	Summary of Age, Years by Treatment
6	An_03_Total	1	Summary of Age, Years
7	An_04	1	Summary of Subjects by Treatment and Age groups (years), n (%)
8	An_04_Total	1	Summary of Subjects by Age groups (years), n (%)
9	An_05	1	Summary of Subjects by Treatment and Race, n (%)
10	An_05_Total	1	Summary of Subjects by Race, n (%)
11	An_06	1	Summary of Subjects by Treatment and Ethnicity, n (%)
12	An_06_Total	1	Summary of Subjects by Ethnicity, n (%)
13	An_07	1	Summary of Subjects by Treatment and Country of participation, n (%)
14	An_07_Total	1	Summary of Subjects by Country of participation, n (%)
15	An_08	1	Summary of Baseline Weight (kg) by Treatment
16	An_08_Total	1	Summary of Baseline Weight (kg)
17	An_09	1	Summary of Baseline Height (cm) by Treatment
18	An_09_Total	1	Summary of Baseline Height (cm)
19	An_10	1	Summary of Baseline BMI (kg/m^2) by Treatment
20	An_10_Total	1	Summary of Baseline BMI (kg/m^2)
21	An_11	1	Summary of Duration of Disease (Months) by Treatment
22	An_11_Total	1	Summary of Duration of Disease (Months)

Calling **siera**

```
library(siera)
```

```
# the ARS JSON File:
```

```
json_path <- system.file("extdata", "Common_Safety_Displays.json", package = "siera")
```

```
# store ARD scripts in this folder:
```

```
output_folder <- file.path(Rscripts)
```

```
# this folder contains ADaM datasets to produce ARD:
```

```
ADaM_folder <- file.path(ADaM)
```

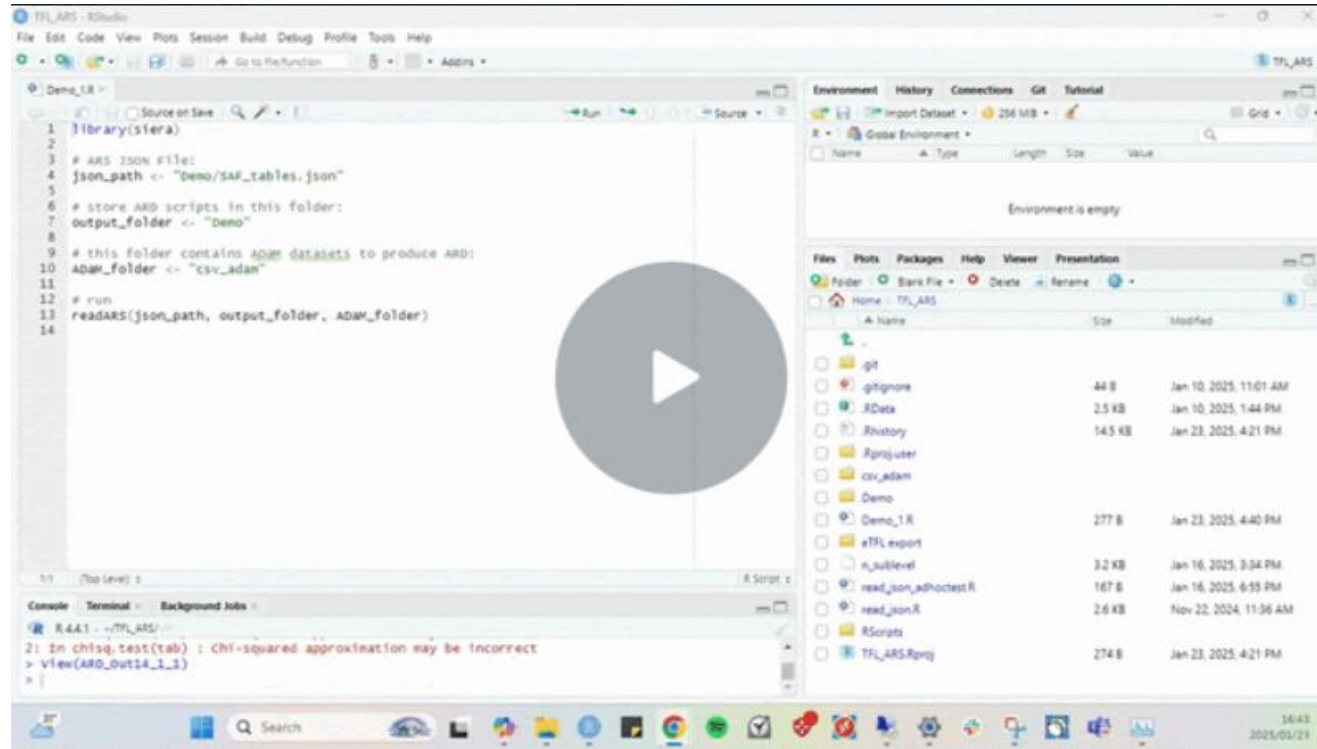
```
# run the readARS function with these 3 parameters.
```

```
readARS(json_path, output_folder, ADaM_folder)
```

siera generated ARD-ready R scripts

```
1 # Programme: Generate code to produce ARD for Out14-1-1
2 # Output: Summary of Demographics
3 # Date created: 2025-02-03 09:58:07
4
5 # load libraries ----
6 library(tidyverse)
7 library(readxl)
8 library(splitstackshape)
9 library(readr)
10
11 # load ADaM ----
12 ADSL <- read_csv('csv_adam/ADSL.csv')
13
14 # Analysis An01_05_SAF_Summ_ByTrt ----
15 # Apply Analysis Set --
16 # Analysis set : SAF
17 df_An01_05_SAF_Summ_ByTrt <- dplyr::filter(ADSL,
18   SAFFL == 'Y')
19
20
21
22 #Apply Analysis Grouping ---|
23 df1_An01_05_SAF_Summ_ByTrt <- df_An01_05_SAF_Summ_ByTrt %>%
24   dplyr::group_by(TRT01A)
```

Let's see **siera** in operation!



Video Presentation



Video link: <https://www.loom.com/share/1fc670d20224ef2bbc7db722e7de090?sid=713e3dac-7ded-4e3d-bea8-adfd223fec85>

Impact, Learnings and Next Steps

Impact

↑ Metadata setup
↓ Manual programming

↑ Time spent on complex tasks
↓ Time spent on repetitive tasks

↑ Quality and efficiency
↓ Overall Cost

Learnings

- Need of a design tool to generate both ARS and ADM metadata
- Integrating ARD into the process requires change
- Avoid “black box” solution
- Provide flexibility to the user

Next Steps

- Integration into workflows
- More feedback, more updates





Malan Bosman

Lead Data Automation Engineer

mbosman@clymbclinical.com



EMAIL

info@clymbclinical.com



WEBSITE

www.clymbclinical.com



PHONE

781-692-3613