CLYMB
CLINICAL

Let the Machine Do the Coding: A SAS-based Framework for Automating TFL Generation with ARS Metadata

Paper & Presentation # SM13

PHUSE US Connect 2025

March 18, 2025

Julie Wachara, Senior Statistical Programmer

CDISC Analysis Result Standards – Released April 19, 2024!

cdisc
Clear data. Clear impact

Analysis Results Standard (ARS) v1.0

cdisc ARS ANALYSIS RESULTS STANDARD

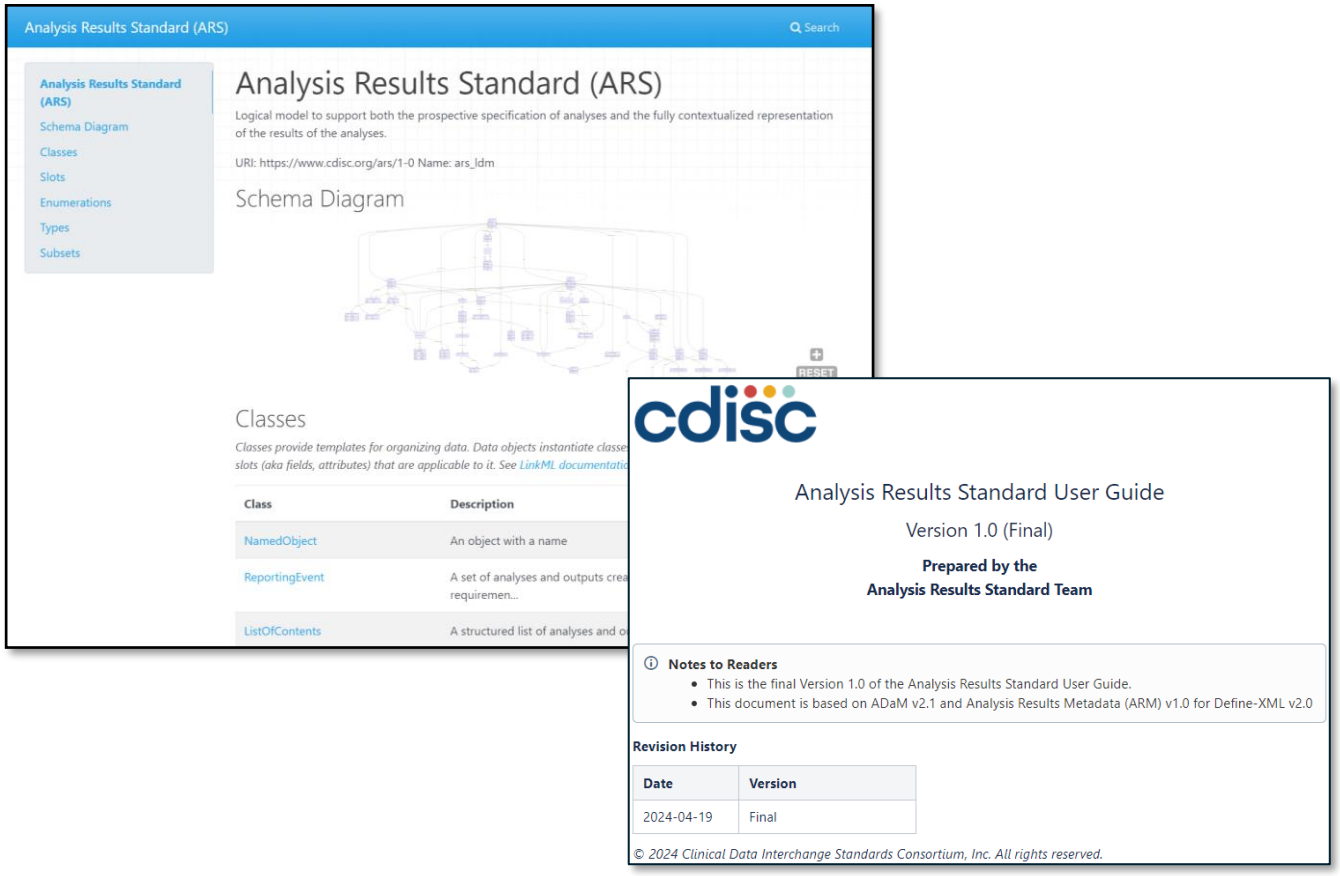
Large trials generate many analysis results in the form of tables, figures, and written reports, yet these results are rarely output in a form that is machine-readable. Previously, there has been no standard way of describing and organizing these results, making it difficult to automate their generation, 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), 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/

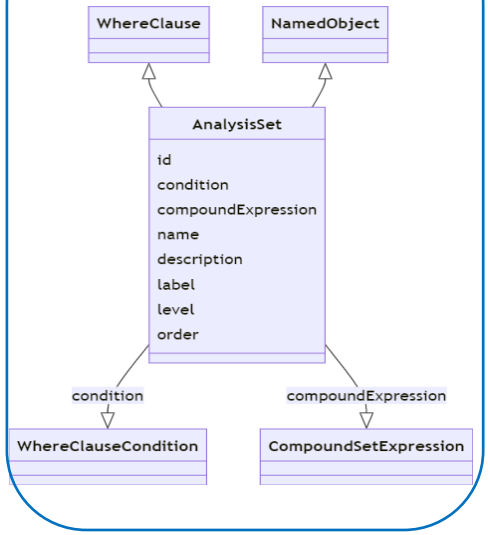


https://wiki.cdisc.org/display/ARSP/Analysis+Results+Standard+User+Guide+v1.0

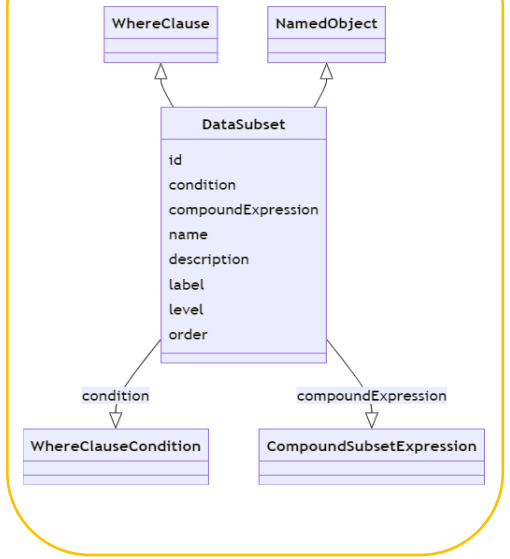
CDISC ARS Metadata

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

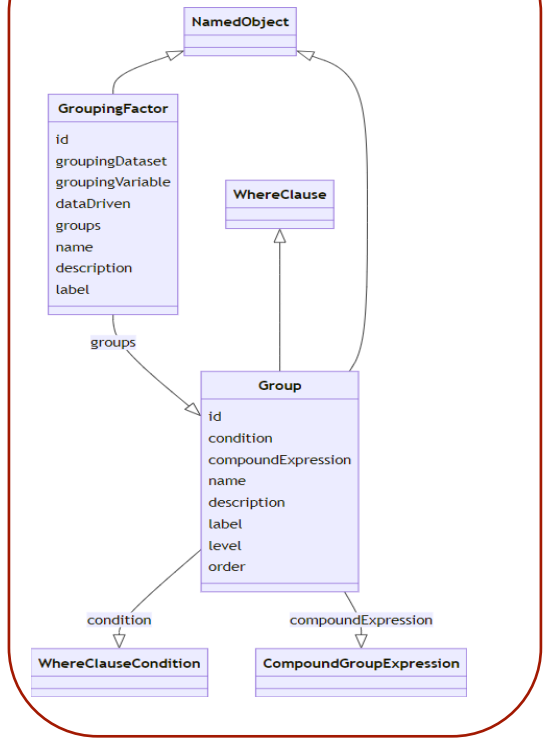
Analysis Set



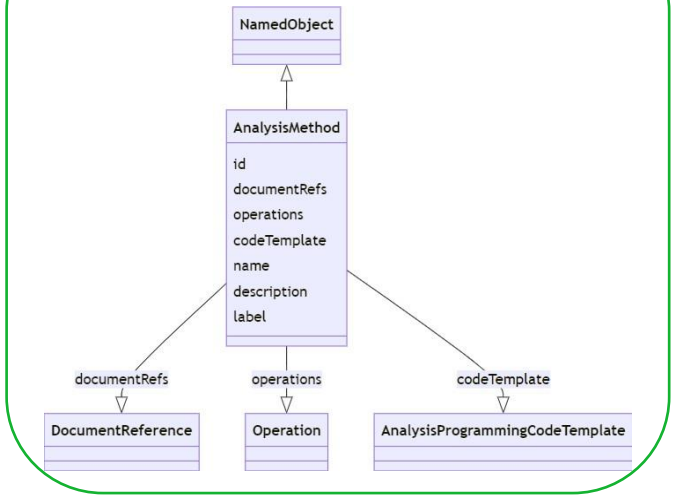
Data Subset



Analysis Grouping



Analysis Method



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 Page x of y

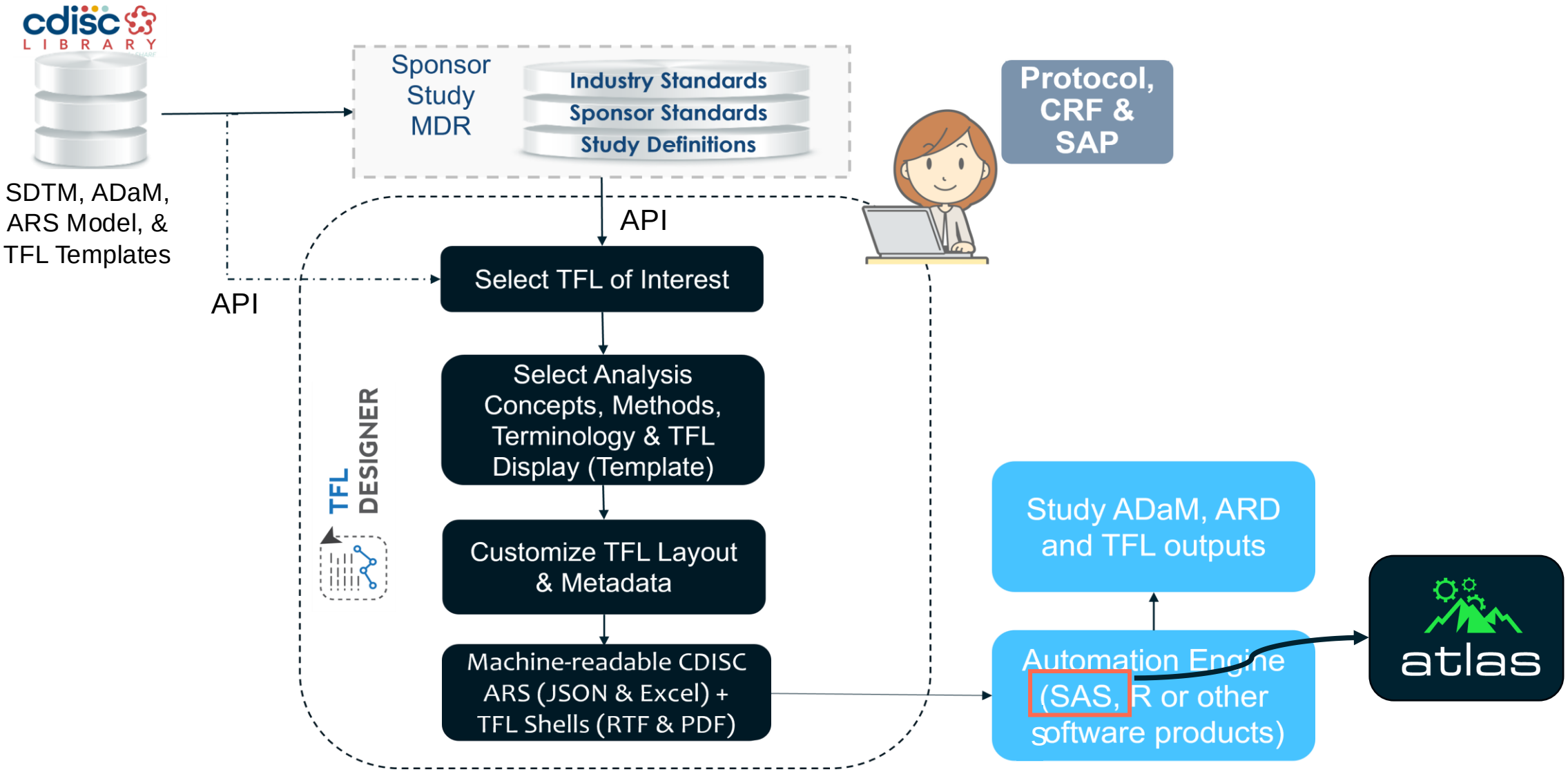
Table 14.3.1.1
Summary of TEAE by System Organ Class and Preferred Term
Safety Population

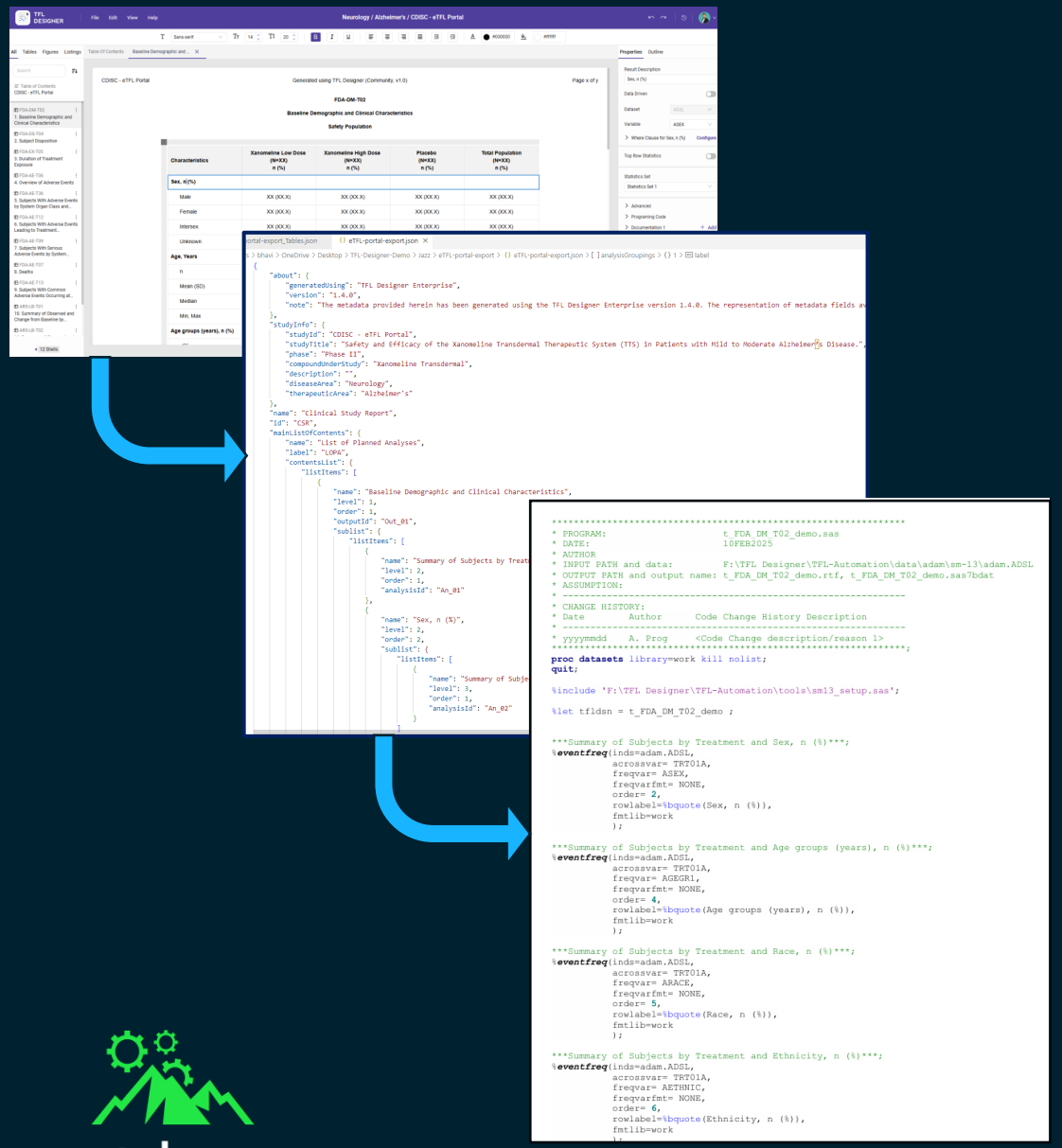
System Organ Class Preferred Term [a], n (%)	Placebo (N=XX)	Xanomeline Low Dose (N=XX)	Xanomeline High Dose (N=XX)
Number of subjects with at least one event	XX (XX.X)	XX (XX.X)	XX (XX.X)
<SOC 1>	XX (XX.X)	XX (XX.X)	XX (XX.X)
<Preferred Term 1>	XX (XX.X)	XX (XX.X)	XX (XX.X)
...	XX (XX.X)	XX (XX.X)	XX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

Metadata-driven TFL Deliverable Workflow



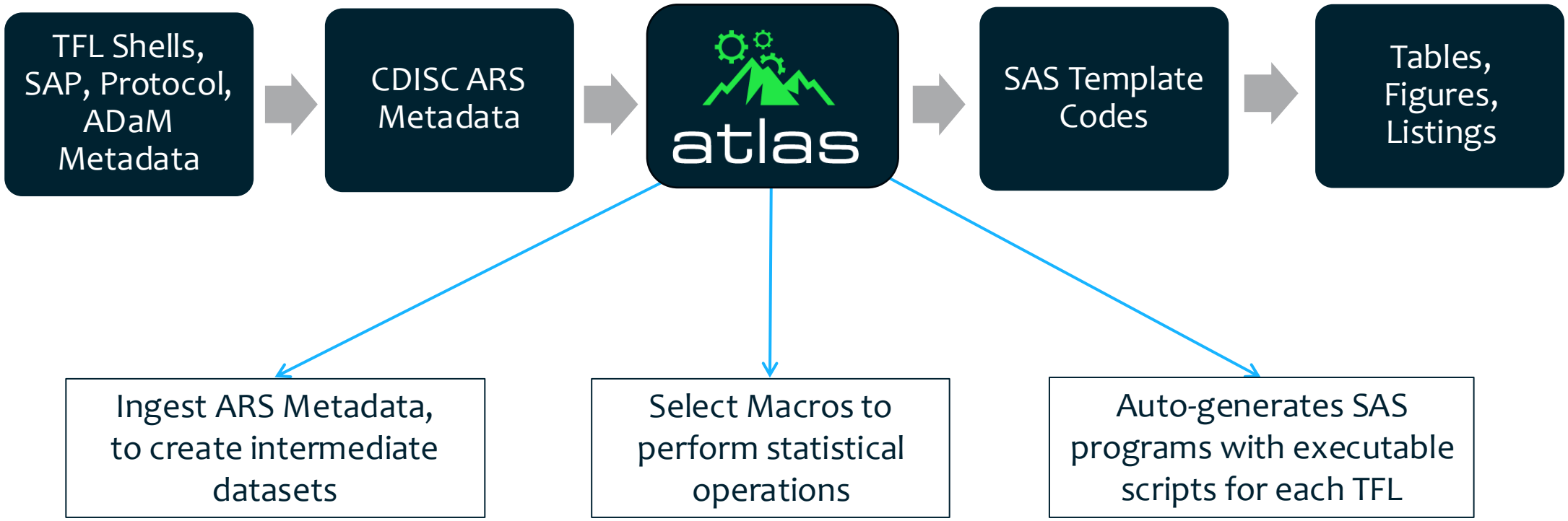


What is atlas?

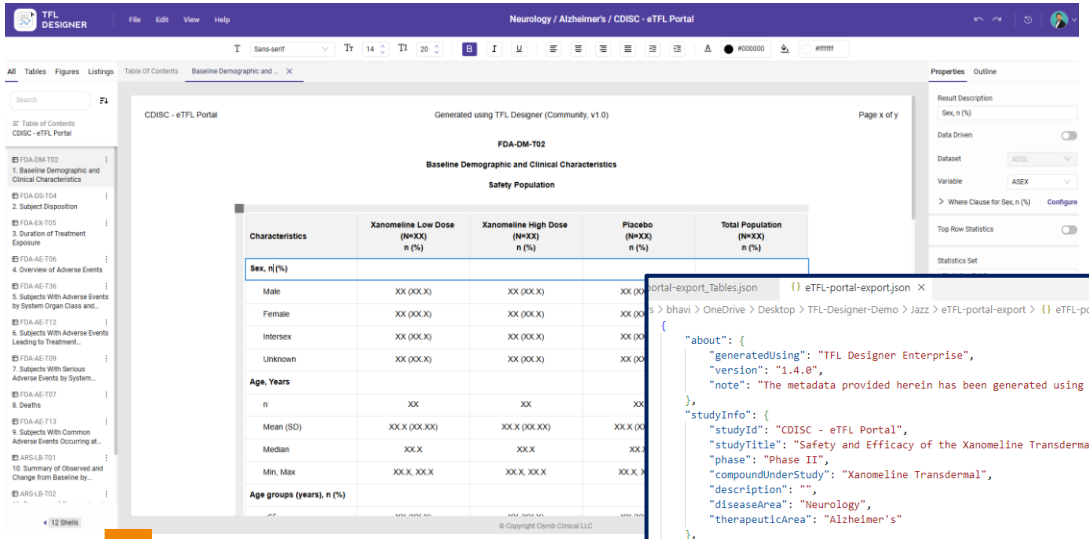
atlas is a **SAS-based** automation engine developed by Clymb that ingests **CDISC ARS** metadata, applies statistical operations, and generates **ready-to-use, executable SAS programs** for each TFL.

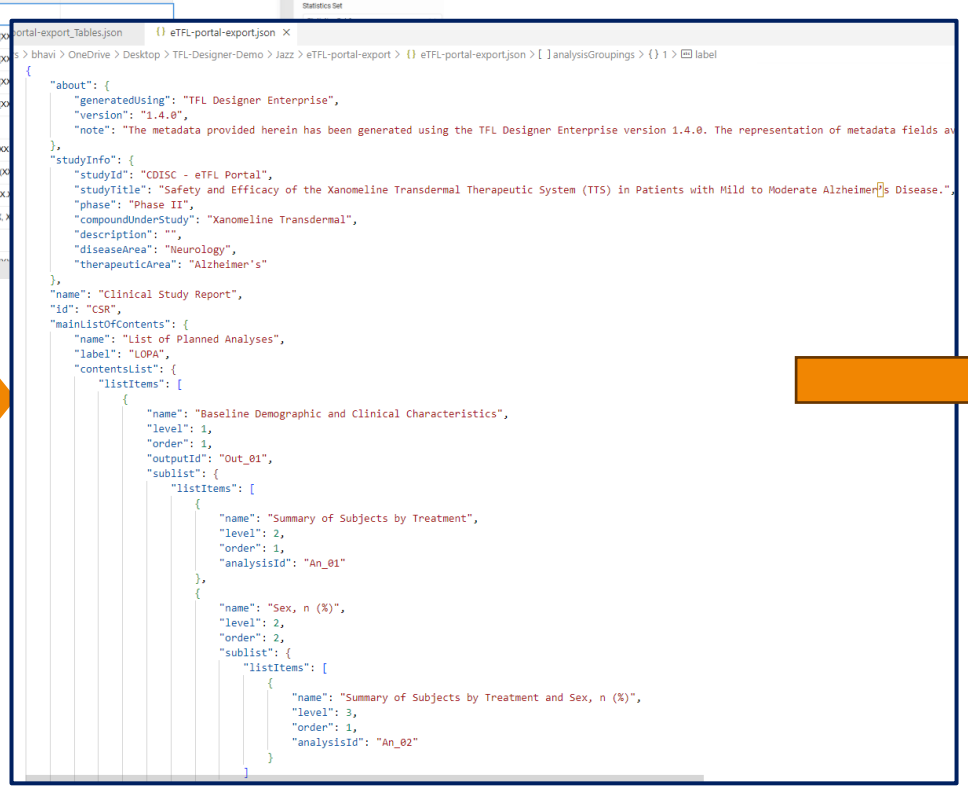
atlas

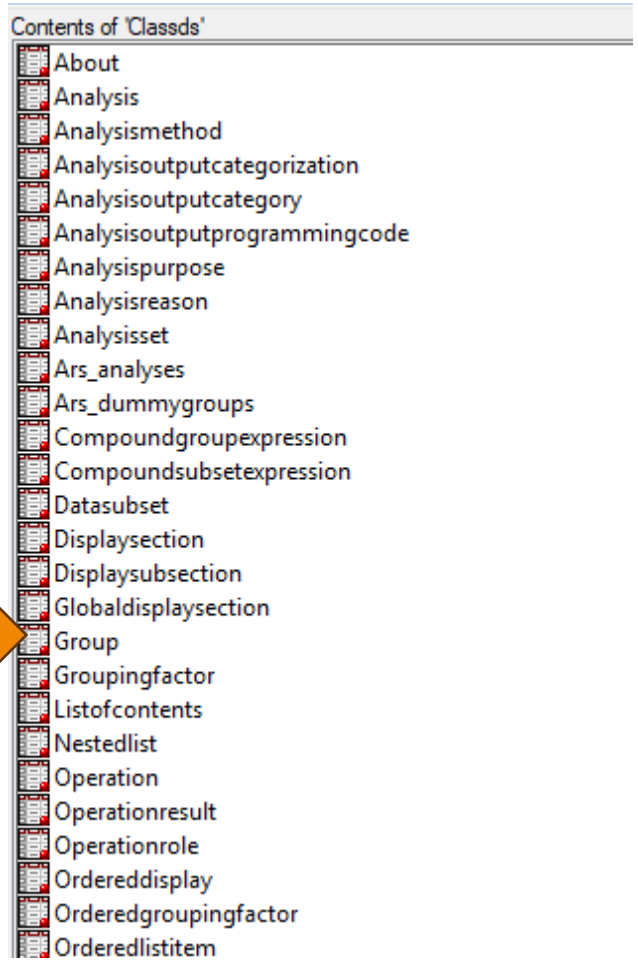
High-level atlas meta-programming process flow: Shells -> TFLs



Export Analysis Results Metadata from TFL Designer: JSON







Select Metadata of Interest from Class Datasets

Contents of 'Classds'

- About
- Analysis
- Analysismethod
- Analysisoutputcategorization
- Analysisoutputcategory
- Analysisoutputprogrammingcode
- Analysispurpose
- Analysisreason
- Analysisset
- Ars_analyses
- Ars_dummygroups
- Compoundgroupexpression
- Compoundsubsetexpression
- Datasubset
- Displaysection
- Displaysubsection
- Globaldisplaysection
- Group
- Groupingfactor
- Listofcontents
- Nestedlist
- Operation
- Operationresult
- Operationrole
- Ordereddisplay
- Orderedgroupingfactor
- Orderedlistitem

VIEWTABLE: Classds.Analysisoutputcategory

	tablepath	datapath	id	label
1	/root/analysisOutputCategorizations/cate	/analysisOutputCategorizations/0/catego	ANSET_01	Safety
2	/root/analysisOutputCategorizations/cate	/analysisOutputCategorizations/0/catego	ANSET_01_ANINT_01	Vital Signs
3	/root/analysisOutputCategorizations/cate	/analysisOutputCategorizations/0/catego	ANSET_01_ANINT_02	Laboratory Test Results
4	/root/analysisOutputCategorizations/cate	/analysisOutputCategorizations/0/catego	ANSET_01_ANINT_03	Adverse Events
5	/root/analysisOutputCategorizations/cate	/analysisOutputCategorizations/0/catego	ANSET_01_ANINT_04	Demographics
6	/root/analysisOutputCategorizations/cate	/analysisOutputCategorizations/0/catego	ANSET_01_ANINT_05	Disposition
7	/root/analysisOutputCategorizations/cate	/analysisOutputCategorizations/0/catego	ANSET_01_ANINT_06	Exposure

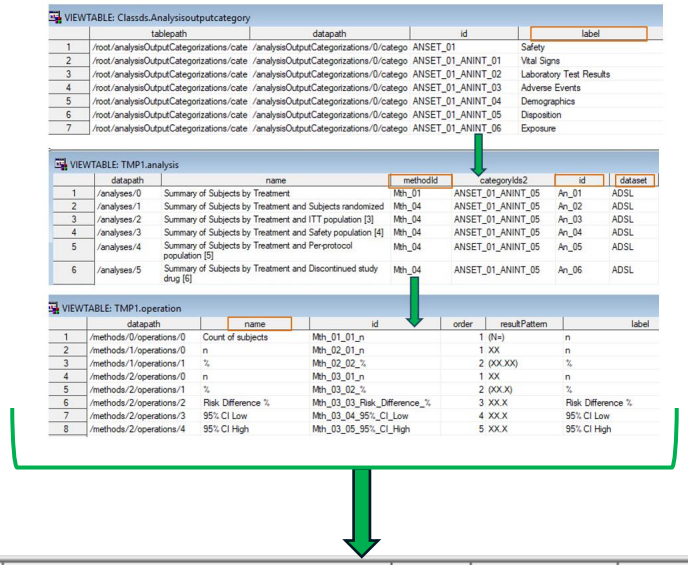
VIEWTABLE: TMP1.analysis

	datapath	name	methodId	categoryIds2	id	dataset
1	/analyses/0	Summary of Subjects by Treatment	Mth_01	ANSET_01_ANINT_05	An_01	ADSL
2	/analyses/1	Summary of Subjects by Treatment and Subjects randomized	Mth_04	ANSET_01_ANINT_05	An_02	ADSL
3	/analyses/2	Summary of Subjects by Treatment and ITT population [3]	Mth_04	ANSET_01_ANINT_05	An_03	ADSL
4	/analyses/3	Summary of Subjects by Treatment and Safety population [4]	Mth_04	ANSET_01_ANINT_05	An_04	ADSL
5	/analyses/4	Summary of Subjects by Treatment and Per-protocol population [5]	Mth_04	ANSET_01_ANINT_05	An_05	ADSL
6	/analyses/5	Summary of Subjects by Treatment and Discontinued study drug [6]	Mth_04	ANSET_01_ANINT_05	An_06	ADSL

VIEWTABLE: TMP1.operation

	datapath	name	id	order	resultPattern	label
1	/methods/0/operations/0	Count of subjects	Mth_01_01_n	1	(N=)	n
2	/methods/1/operations/0	n	Mth_02_01_n	1	XX	n
3	/methods/1/operations/1	%	Mth_02_02_%	2	(XX.XX)	%
4	/methods/2/operations/0	n	Mth_03_01_n	1	XX	n
5	/methods/2/operations/1	%	Mth_03_02_%	2	(XX.X)	%
6	/methods/2/operations/2	Risk Difference %	Mth_03_03_Risk_Difference_%	3	XX.X	Risk Difference %
7	/methods/2/operations/3	95% CI Low	Mth_03_04_95%_CI_Low	4	XX.X	95% CI Low
8	/methods/2/operations/4	95% CI High	Mth_03_05_95%_CI_High	5	XX.X	95% CI High

Create Intermediate Metadata Dataset



tfl_id	tfltitle	catlabel	anlsdisplabel	dataset	groupingid1var	groupingid2var	analysisid	methodid	grouped_results
Out_02	FDA-DM-T02	Demographics		ADSL	TRT01A	USUBJID	An_08	Mth_01	n
Out_02	FDA-DM-T02	Demographics	Sex, n (%)	ADSL	TRT01A	ASEX	An_09	Mth_04	n %
Out_02	FDA-DM-T02	Demographics	Age, Years	ADSL	TRT01A	AGE	An_10	Mth_07	n MeanSD Median MinMax Q1 Q3
Out_02	FDA-DM-T02	Demographics	Age groups (years), n (%)	ADSL	TRT01A	AGEGR1	An_11	Mth_04	n %
Out_02	FDA-DM-T02	Demographics	Race, n (%)	ADSL	TRT01A	ARACE	An_12	Mth_04	n %
Out_02	FDA-DM-T02	Demographics	Ethnicity, n (%)	ADSL	TRT01A	AETHNIC	An_13	Mth_04	n %
Out_02	FDA-DM-T02	Demographics	Country of participation, n (%)	ADSL	TRT01A	ACOUNTRY	An_14	Mth_04	n %
Out_02	FDA-DM-T02	Demographics	Baseline Weight (kg)	ADSL	TRT01A	WEIGHTBL	An_15	Mth_07	n MeanSD Median MinMax Q1 Q3
Out_02	FDA-DM-T02	Demographics	Baseline Height (cm)	ADSL	TRT01A	HEIGHTBL	An_16	Mth_07	n MeanSD Median MinMax Q1 Q3
Out_02	FDA-DM-T02	Demographics	Baseline BMI (kg/m^2)	ADSL	TRT01A	BMIBL	An_17	Mth_07	n MeanSD Median MinMax Q1 Q3
Out_02	FDA-DM-T02	Demographics	Duration of Disease (Months)	ADSL	TRT01A	DURDIS	An_18	Mth_07	n MeanSD Median MinMax Q1 Q3

Assign SAS Macros by Statistic Operation

tfl_id	tfltitle	catlabel	anlsdisplabel	dataset	groupingid1var	groupingid2var	analysisid	methodid	grouped_results
Out_02	FDA-DM-T02	Demographics		ADSL	TRT01A	USUBJID	An_08	Mth_01	n
Out_02	FDA-DM-T02	Demographics	Sex, n (%)	ADSL	TRT01A	ASEX	An_09	Mth_04	n %
Out_02	FDA-DM-T02	Demographics	Age, Years	ADSL	TRT01A	AGE	An_10	Mth_07	n MeanSD Median MinMax Q1 Q3
Out_02	FDA-DM-T02	Demographics	Age groups (years), n (%)	ADSL	TRT01A	AGEGR1	An_11	Mth_04	n %
Out_02	FDA-DM-T02	Demographics	Race, n (%)	ADSL	TRT01A	ARACE	An_12	Mth_04	n %
Out_02	FDA-DM-T02	Demographics	Ethnicity, n (%)	ADSL	TRT01A	AETHNIC	An_13	Mth_04	n %
Out_02	FDA-DM-T02	Demographics	Country of participation, n (%)	ADSL	TRT01A	ACOUNTRY	An_14	Mth_04	n %
Out_02	FDA-DM-T02	Demographics	Baseline Weight (kg)	ADSL	TRT01A	WEIGHTBL	An_15	Mth_07	n MeanSD Median MinMax Q1 Q3
Out_02	FDA-DM-T02	Demographics	Baseline Height (cm)	ADSL	TRT01A	HEIGHTBL	An_16	Mth_07	n MeanSD Median MinMax Q1 Q3
Out_02	FDA-DM-T02	Demographics	Baseline BMI (kg/m^2)	ADSL	TRT01A	BMIBL	An_17	Mth_07	n MeanSD Median MinMax Q1 Q3
Out_02	FDA-DM-T02	Demographics	Duration of Disease (Months)	ADSL	TRT01A	DURDIS	An_18	Mth_07	n MeanSD Median MinMax Q1 Q3

```
proc sql ;
  select count (mthdord) into :mthdord from analyses&j;
quit;

    %do k=1 %to &mthdord.;
       data mthdord&k.;
            set analyses&j;
            where mthdord = &k.;
            if catlabel = "Demographics" then do;
                if methodId = "Mth_04" then do;
                    call symputx(cats('anlsnam', &k.), analysisname);
                    call symputx(cats('dsn', &k.), dsn );
                    call symputx(cats('grpvar1', &k.),grpvar1);
                    call symputx(cats('grpvar2', &k.), grpvar2);
                    call symputx(cats('resfmt', &k.), resfmt);
                    call symputx(cats('anlsord', &k.), anlsord);
                    call symputx(cats('library', &k.), library);
                    call symputx(cats('anlslab', &k.), anlsdisplabel);
                end;
                else if methodId = "Mth_07" then do;
                    call symputx(cats('anlsnam', &k.), analysisname);
                    call symputx(cats('dsn', &k.), dsn );
                    call symputx(cats('grpvar1', &k.),grpvar1);
                    call symputx(cats('grpvar2', &k.), grpvar2);
                    call symputx(cats('grpreslt', &k.), grouped_results);
                    call symputx(cats('nmaxdigt', &k.), nmaxdigt);
                    call symputx(cats('vmaxdigt', &k.), vmaxdigt);
                    call symputx(cats('n_dec', &k.), n_dec);
                    call symputx(cats('anlsord', &k.), anlsord);
                    call symputx(cats('anlslab', &k.), anlsdisplabel);
                end;
            end;
       run;

    %end;
```

Generate Individual TFL Programs with TFL Metadata and SAS Template Codes

```
proc sql ;
  select count (mthdord) into :mthdord from analyses&j;
quit;

    %do k=1 %to &mthdord.;
       data mthdord&k.;
            set analyses&j;
            where mthdord = &k.;
            if catlabel = "Demographics" then do;

               if methodId = "Mth_04" then do;
                    call symputx(cats('anlsnam', &k.), analysisname);
                    call symputx(cats('dsn', &k.), dsn );
                    call symputx(cats('grpvar1', &k.),grpvar1);
                    call symputx(cats('grpvar2', &k.), grpvar2);
                    call symputx(cats('resfmt', &k.), resfmt);
                    call symputx(cats('anlsord', &k.), anlsord);
                    call symputx(cats('library', &k.), library);
                    call symputx(cats('anlslab', &k.), anlsdisplab
               end;

               else if methodId = "Mth_07" then do;
                    call symputx(cats('anlsnam', &k.), analysisname);
                    call symputx(cats('dsn', &k.), dsn );
                    call symputx(cats('grpvar1', &k.),grpvar1);
                    call symputx(cats('grpvar2', &k.), grpvar2);
                    call symputx(cats('grpreslt', &k.), grouped_results);
                    call symputx(cats('nmaxdigt', &k.), nmaxdigt);
                    call symputx(cats('vmaxdigt', &k.), vmaxdigt);
                    call symputx(cats('n_dec', &k.), n_dec);
                    call symputx(cats('anlsord', &k.), anlsord);
                    call symputx(cats('anlslab', &k.), anlsdisplabel);
               end;
            end;
        run;

    %end;
```

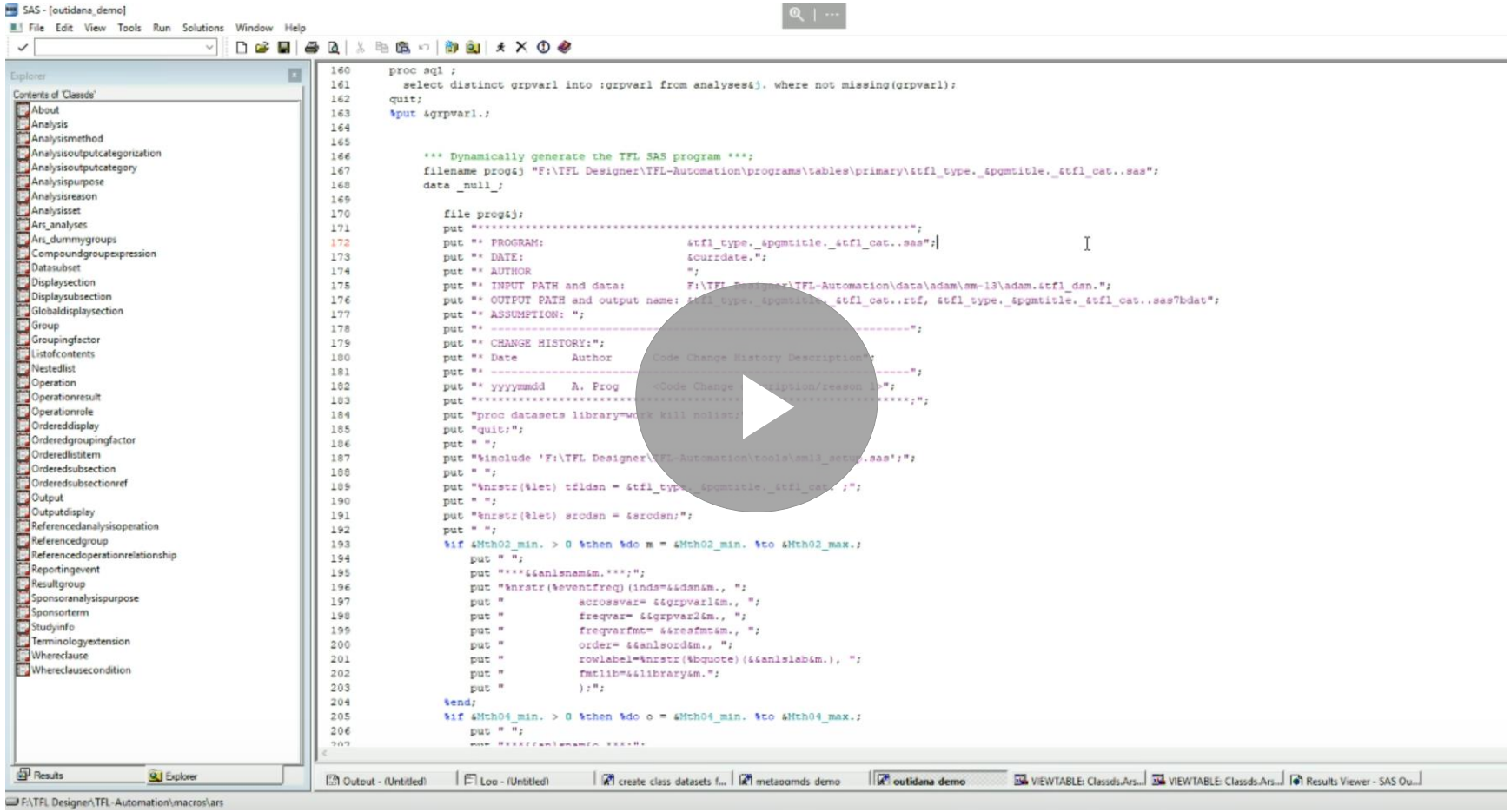
```
/* Step 3: Dynamically generate the TFL SAS program with executable scripts */
 filename prog&j "File location\&tfl_type._&pgmtitle._&tfl_cat..sas";
 data _null_;

    file prog&j;
    *Add header properties;
    put "*************************************************************************";
    put "* PROGRAM:                      &tfl_type._&pgmtitle._&tfl_cat..sas";
    put "* DATE:                         &currdate.";
    put "* AUTHOR                        ";
    put "* INPUT PATH and data:          F:\TFL Designer\TFL-Automation\data\adam\sm-13\adam.&tfl_dsn.";
    put "* INPUT PATH and data:          File location\adam.&tfl_dsn.";
    put "* OUTPUT PATH and output name: F:\TFL Designer\TFL-Automation\outputs\tables\primary
                                        &tfl_type._&pgmtitle._&tfl_cat..rtf,";
    put "*                              &tfl_type._&pgmtitle._&tfl_cat..sas7bdat";

    *add other study level maros and libraries as needed;
    put "%include 'F:\TFL Designer\TFL-Automation\tools\sml3_setup.sas';";
    put " ";
    put "%nrstr(%let) tfldsn = &tfl_type._&pgmtitle._&tfl_cat. ;";
    put " ";

    *Assign macro based on methodid and analysis category;
          %if &Mth04_min. > 0 %then %do m = &Mth04_min. %to &Mth04_max.;
              put " ";
              put "***&&anlsnam&m.***;";
              put "%nrstr(%eventfreq)(inds=&&dsn&m., ";
              put "           acrossvar= &&grpvar1&m., ";
              put "           freqvar= &&grpvar2&m., ";
              put "           freqvarfmt= &&resfmt&m., ";
              put "           order= &&anlsord&m., ";
              put "           rowlabel=%nrstr(%bquote)(&&anlslab&m.), ";
              put "           fmtlib=&&library&m.";
              put "           );";
          %end;

          %if &Mth04_min. > 1 or &Mth07_min. > 1 %then %do;
              put "%nrstr(%stack)(outds=%nrstr(&tfldsn.),";
              put "       varlist=&&anlsvar&j.);";
          %end;
```

Ready-to-use Executable SAS Program

```
/* Step 3: Dynamically generate the TFL SAS program with executable scripts */
 filename prog&j "File location\&tfl_type._&pgmtitle._&tfl_cat..sas";
 data _null_;

    file prog&j;
    *Add header properties;
    put "************************************************************************";
    put "* PROGRAM:                     &tfl_type._&pgmtitle._&tfl_cat..sas";
    put "* DATE:                        &currdate.";
    put "* AUTHOR                       ";
    put "* INPUT PATH and data:         F:\TFL Designer\TFL-Automation\data\adam\sm-13\adam.&tfl_dsn.";
    put "* INPUT PATH and data:         File location\adam.&tfl_dsn.";
    put "* OUTPUT PATH and output name: F:\TFL Designer\TFL-Automation\outputs\tables\primary
                                       &tfl_type._&pgmtitle._&tfl_cat..rtf,";
    put "*                             &tfl_type._&pgmtitle._&tfl_cat..sas7bdat";

    *add other study level maros and libraries as needed;
    put "%include 'F:\TFL Designer\TFL-Automation\tools\sml3_setup.sas';";
    put " ";
    put "%nrstr(%let) tfldsn = &tfl_type._&pgmtitle._&tfl_cat. ;";
    put " ";

    *Assign macro based on methodid and analysis category;
            %if &Mth04_min. > 0 %then %do m = &Mth04_min. %to &Mth04_max.;
                put " ";
                put "***&&anlsnam&m.***;";
                put "%nrstr(%eventfreq)(inds=&&dsn&m., ";
                put "           acrossvar= &&grpvar1&m., ";
                put "           freqvar= &&grpvar2&m., ";
                put "           freqvarfmt= &&resfmt&m., ";
                put "           order= &&anlsord&m., ";
                put "           rowlabel=%nrstr(%bquote)(&&anlslab&m.), ";
                put "           fmtlib=&&library&m.";
                put "           );";
            %end;

            %if &Mth04_min. > 1 or &Mth07_min. > 1 %then %do;
                put "%nrstr(%stack)(outds=%nrstr(&tfldsn.),";
                put "       varlist=&&anlsvar&j.);";
            %end;
```

```
***Summary of Subjects by Treatment and Country of participation,
%eventfreq(inds=adam.ADSL,
           acrossvar= TRT01A,
           freqvar= ACOUNTRY,
           freqvarfmt= NONE,
           order= 7,
           rowlabel=%bquote(Country of participation, n (%)),
           fmtlib=work
           );

***Summary of Subjects by Treatment and Age, Years***;
%univars(inds=adam.ADSL,
           acrossvar= TRT01A,
           univars= AGE,
           stats= n MeanSD Median MinMax Q1 Q3,
           ndig= 2,
           ndvar= 3,
           ndecm=0,
           order= 3,
           rowlabel=%bquote(Age, Years),
           );
```

Let's see atlas in operation!



Video Presentation

video link: https://www.loom.com/share/0a46a34356014dbea99de144ffc706ed

atlas Impact and Learnings

Impact


Metadata setup


Manual programming


Time spent on complex tasks

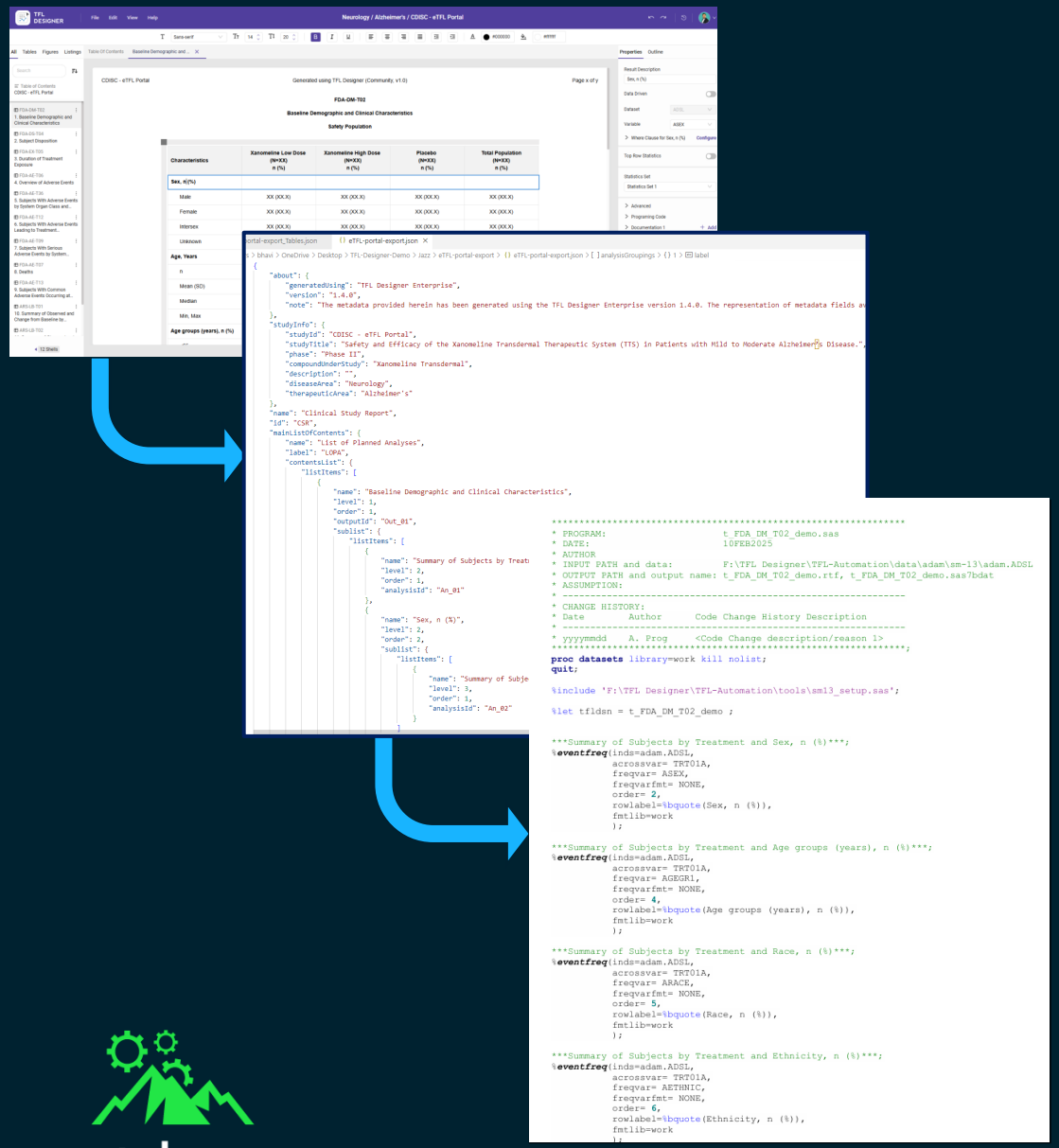

Time spent on repetitive tasks


Quality and efficiency


Overall Cost

Learnings

- Need of a design tool to generate both ARM and ADM metadata
- Training, creation, and implementation of automation process takes time.
- Macro availability and structure can be limiting.
- Multiple aspects of the programming workflow may have to change:
  - ADAM structure
  - Specification development
  - Validation process



atlas

atlas Next Steps

- Build Macro Library for more complex analyses
- Add functionalities to detect changes with Metadata updates
- Integrating metadata with tracker functions

Thank you!

Julie Wachara
Senior Statistical Programmer
jwachara@clymbclinical.com

EMAIL
info@clymbclinical.com

WEBSITE
www.clymbclinical.com

PHONE
781-692-3613