Cytel

phuse·connect
connect·share·advance
5-8 November
ICC Birmingham

EU 2023
The Clinical Data Science Conference

# Leveraging the Analysis Results Standard (ARS): The Cytel PRISM Experience

**Sebastià Barceló, Cytel Inc.**

**PHUSE EU Connect 2023**

5th - 8th November 2023

Paper AD16 – Application Development (AD)

# Agenda

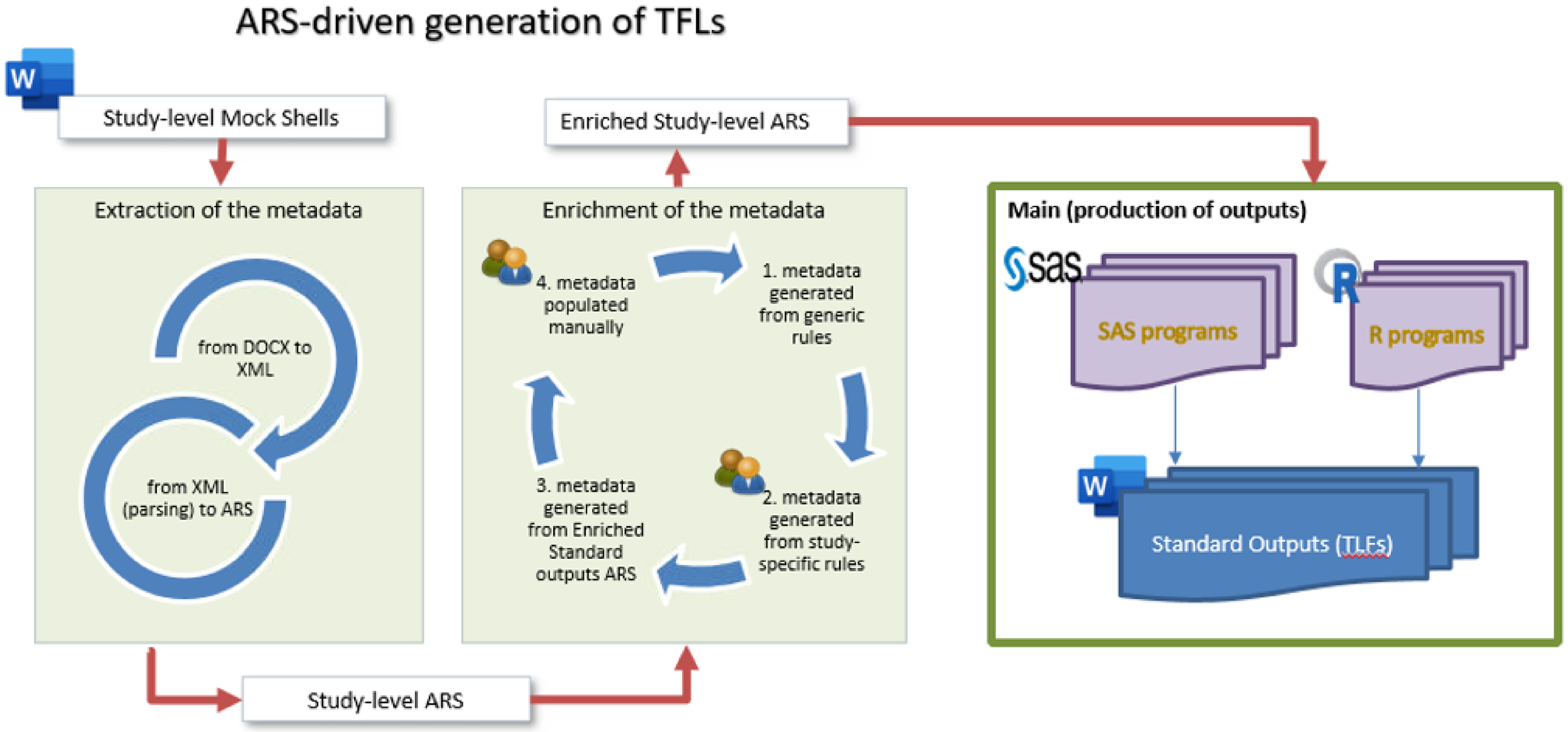
- PRISM Concept
- DOCX → Metadata
- PRISM ARS
- Metadata to Code
- Conclusions

# PRISM Concept

# PRISM Concept

- We wanted a tool to generate code from Metadata.
- From where should we start?
- How much information can we extract from Mock Shells?
- If we cannot extract all the needed metadata, how can we enrich the Metadata to increase the number of outputs we can generate with the tool?
- How can we generate code from our Enriched Metadata?

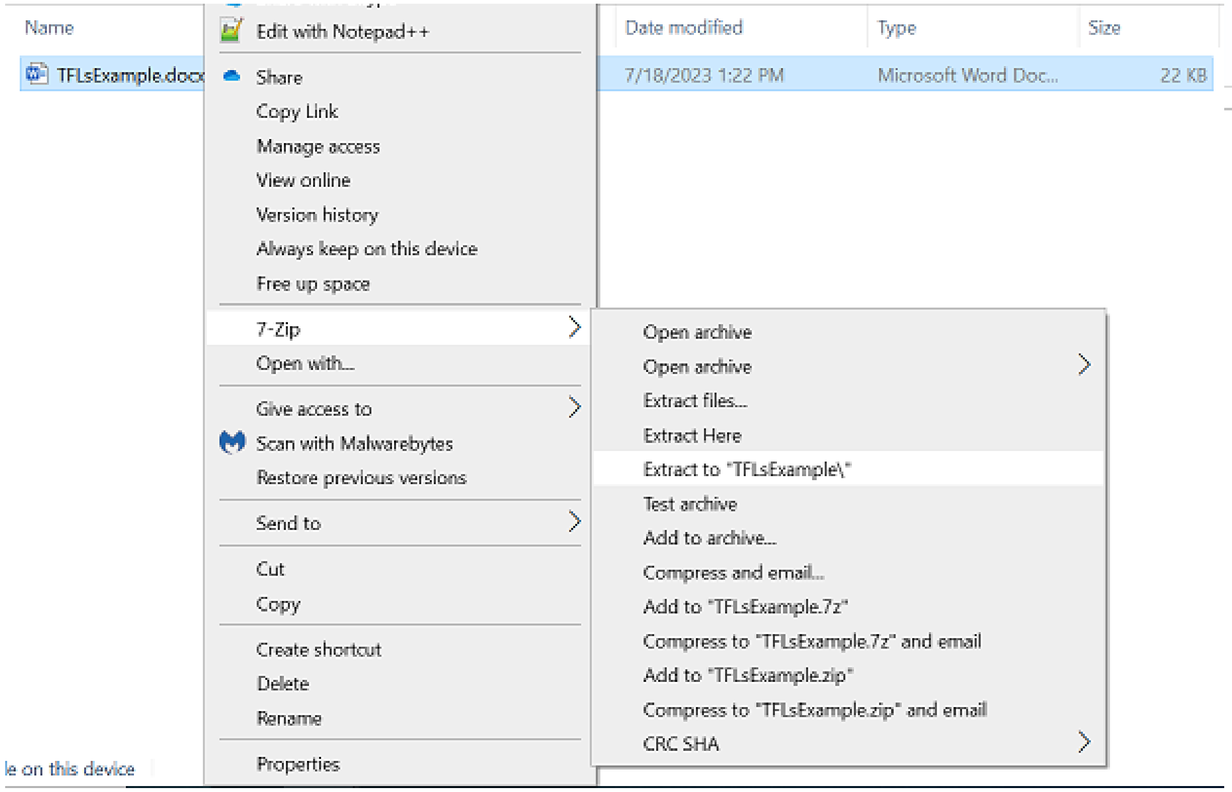
# PRISM Concept

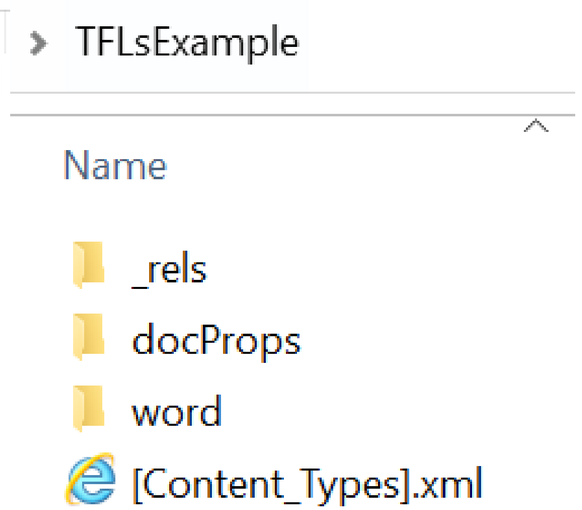


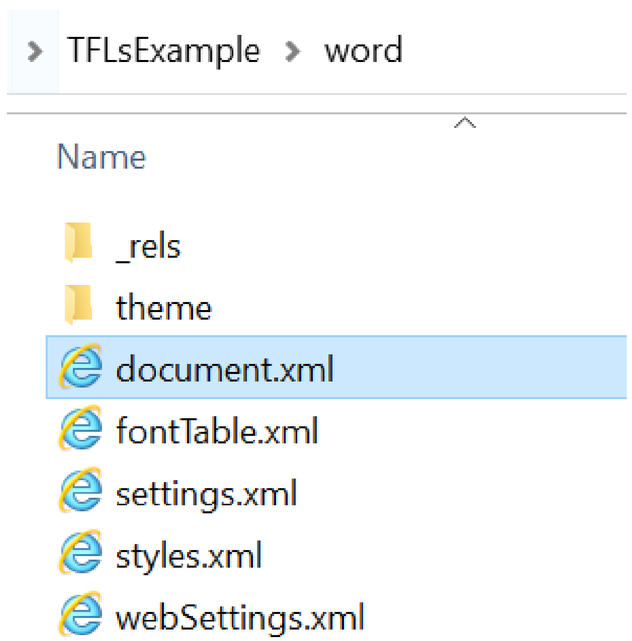
# DOCX→ Metadata

# Mock Shells → XML → Metadata

- **XML is employed in Microsoft Word tools, and you can easily access to those XML files “Unzipping” the DOCX files.**



# Mock Shells → XML → Metadata

**Table 14.1.2: Demographics - Safety Analysis Set**

{Treatment Group: Pbo, Act, Total}

| | Placebo (N=XXXX) | Active Drug (N=XXXX) | All Subjects (N=XXXX) |
|---|---|---|---|
| Age (Years) | | | |
| n (missing) | xxx (x) | xxx (x) | xxx (x) |
| Mean (SD) | xx.x (xx.x) | xx.x (xx.x) | xx.x (xx.x) |
| 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 Categories, n(%) | | | |
| n (missing) | xxx (x) | xxx (x) | xxx (x) |
| <25 | xxx (xx.x) | xxx (xx.x) | xxx (xx.x) |
| 25-39 | xxx (xx.x) | xxx (xx.x) | xxx (xx.x) |
| 40-49 | xxx (xx.x) | xxx (xx.x) | xxx (xx.x) |
| 50-59 | xxx (xx.x) | xxx (xx.x) | xxx (xx.x) |
| >=60 | xxx (xx.x) | xxx (xx.x) | xxx (xx.x) |
| Sex, n(%) | | | |
| n (missing) | xxx (x) | xxx (x) | xxx (x) |
| Male | xxx (xx.x) | xxx (xx.x) | xxx (xx.x) |
| Female | xxx (xx.x) | xxx (xx.x) | xxx (xx.x) |

Percentages are based on the number of non-missing observations.
Output ID: t_demo DDMMMYY HH:MM <PROJECT/TASK LOCATION>..\t_demo.sas
Page 1 of x

{PROGRAMMING NOTE: Example of a note for the programmer}

```
- <w:p w:rsidP="00401510" w:rsidRDefault="00401510" w:rsidRPr="00401510" w:rsidR="00401510" w14:textId="77777777" w14:paraId="094CB5CC">
  + <w:pPr>
  - <w:r w:rsidRPr="00401510">
      + <w:rPr>
        <w:t>Table 14.1.2: Demographics - Safety Analysis Set</w:t>
    </w:r>
</w:p>
```

```
- <w:p w:rsidP="007A1A3C" w:rsidRDefault="007A1A3C" w:rsidR="007A1A3C" w14:textId="77777777" w14:paraId="34DB55B7">
  - <w:pPr>
      <w:pStyle w:val="Token"/>
    </w:pPr>
  - <w:r>
      <w:t xml:space="preserve">{Treatment Group: </w:t>
    </w:r>
    <w:proofErr w:type="spellStart"/>
  - <w:r>
      <w:t>Pbo</w:t>
    </w:r>
    <w:proofErr w:type="spellEnd"/>
  - <w:r>
      <w:t>, Act, Total}</w:t>
    </w:r>
</w:p>
```

```
                <w:t>Sex, n(%)</w:t>
              </w:r>
            </w:p>
          </w:tc>
        + <w:tc>
        + <w:tc>
        + <w:tc>
      </w:tr>
- <w:tr w:rsidRPr="00401510" w:rsidR="00401510" w14:textId=
  - <w:tc>
      + <w:tcPr>
      - <w:p w:rsidP="00401510" w:rsidRDefault="00401510"
          + <w:pPr>
          - <w:r w:rsidRPr="00401510">
              + <w:rPr>
                <w:t xml:space="preserve"> n (missing)</w:t>
            </w:r>
          </w:p>
        </w:tc>
      + <w:tc>
      + <w:tc>
      + <w:tc>
    </w:tr>
- <w:tr w:rsidRPr="00401510" w:rsidR="00401510" w14:textId=
  - <w:tc>
      + <w:tcPr>
      - <w:p w:rsidP="00401510" w:rsidRDefault="00401510"
          + <w:pPr>
          - <w:r w:rsidRPr="00401510">
              + <w:rPr>
                <w:t xml:space="preserve"> Male</w:t>
            </w:r>
          </w:p>
```

# PRISM ARS

# Initial PRISM ARS

**Initial PRISM Metadata Structure (Gray vars not extracted from XML)**

| Titles and Footnotes |
|---|
| Index |
| Prog |
| Suffix |
| OutputID |
| ID |
| Template ID |
| Unique Vs. Repeated |
| NbCol |
| NbLeftCol |
| OverallCol |
| Trt Group Token Label |
| Title1-Title8 |
| Footnote1-Footnote8 |
| Programming Notes |
| Page Header |
| Cell Width |
| Page break |
| Total column name |
| Standard Vs Non Standard |

| Computation |
|---|
| ID |
| TLF |
| Source Data |
| MergeByVars |
| Analysis Population |
| Input dataset filter |
| Analysis Variable Filter |
| BlockID |
| Byvars |
| ByCols |
| Avars |
| Label |
| Summary Type |
| Codelist ID |
| Precision |
| Comments |

| Codelist |
|---|
| ID |
| Block |
| Codelist ID |
| Order |
| Decode |
| Format |

| Page Headers |
|---|
| ID |
| Page Header |
| Page Header Value |
| Avars |
| Seperator (Y/N) |

| Treatment |
|---|
| Trt Group Token Label |
| Codelist Index |
| Codelist Name |
| Code |
| Decode |

**Initial PRISM ARS Metadata structure (Gray vars not extracted from XML)**

| Output |
|---|
| Index |
| Prog |
| Suffix |
| OutputID |
| ID |

| Display |
|---|
| ID |
| Unique vs Repeated |
| Title1 |
| Footnote1 |
| Footnote2 |
| TemplateID |
| Trt Group Token Label |
| Page Header |
| Programming Notes |

| Result |
|---|
| ID |
| ByCols |
| Label |
| BlockID |
| Analysis Variable Filter |
| ByVars |
| Avars |
| Summary Type |
| Codelist ID |
| Precision |
| Comments |

| Codelist |
|---|
| ID |
| Block |
| Codelist ID |
| Order |
| Decode |
| Format |

| Page Headers |
|---|
| ID |
| Page Header |
| Page Header Value |
| Avars |
| Seperator (Y/N) |

| Analysis Set |
|---|
| ID |
| TFL |
| Source Data |
| MergeByVars |
| Analysis Population |
| Input dataset filter |

| Analysis Group |
|---|
| Trt Group Token Label |
| OverallCol |
| Total column name |
| Codelist Index |
| Codelist Name |
| Code |
| Decode |

| Template |
|---|
| ID |
| TemplateID |
| NbCol |
| NbLeftCol |
| Cell Width |
| Standard vs Non Standard |

# PRISM ARS (Codelist)

Initial ARS

| ID | Block | Codelist ID | Order | Decode | Format |
|---|---|---|---|---|---|
| t_01_demo | 2 | Age2 | 1 | <25 | Age2f. |
| t_01_demo | 2 | Age2 | 2 | 25-39 | Age2f. |
| t_01_demo | 2 | Age2 | 3 | 40-49 | Age2f. |
| t_01_demo | 2 | Age2 | 4 | 50-59 | Age2f. |
| t_01_demo | 2 | Age2 | 5 | >=60 | Age2f. |
| t_01_demo | 3 | Sex3 | 1 | Male | Sex3f. |
| t_01_demo | 3 | Sex3 | 2 | Female | Sex3f. |

# PRISM ARS (Analysis Group)

Initial ARS

| Trt Group Token Label | OverallCol | Total column name | Codelist Index | Codelist Name | Code | Decode |
|---|---|---|---|---|---|---|
| Pbo, Act, Total | Y | All Subjects | | | | |
| Merged Header | Y | Total | | | | |

Enriched ARS

| Trt Group Token Label | OverallCol | Total column name | Codelist Index | Codelist Name | Code | Decode |
|---|---|---|---|---|---|---|
| Pbo, Act, Total | Y | All Subjects | 1 | trt1f | 1 | Placebo |
| Pbo, Act, Total | Y | All Subjects | 1 | trt1f | 2 | Active |
| Pbo, Act, Total | Y | All Subjects | 1 | trt1f | 99 | All Subjects |
| Merged Header | Y | Total | 2 | trt2f | 1 | Treatment~Placebo |
| Merged Header | Y | Total | 2 | trt2f | 2 | Treatment~Active |
| Merged Header | Y | Total | 2 | trt2f | 99 | Total |

# PRISM ARS (Analysis Set)

Initial ARS

| ID | TFL | Source Data | MergeByVars | Analysis Population | Input dataset filter |
|---|---|---|---|---|---|
| t_01_demo | Table | | | | |
| t_02_ae1 | Table | | | | |
| t_03_ae1_serious | Table | | | | |

Enriched ARS

| ID | TFL | Source Data | MergeByVars | Analysis Population | Input dataset filter |
|---|---|---|---|---|---|
| t_01_demo | Table | adsl | usubjid | saffl eq "Y" | |
| t_02_ae1 | Table | adsl, adae | usubjid | saffl eq "Y" | trtemfl eq "Y" |
| t_03_ae1_serious | Table | adsl, adae | usubjid | saffl eq "Y" | trtemfl eq "Y" and aeser eq "Y" |

# PRISM ARS (Results)

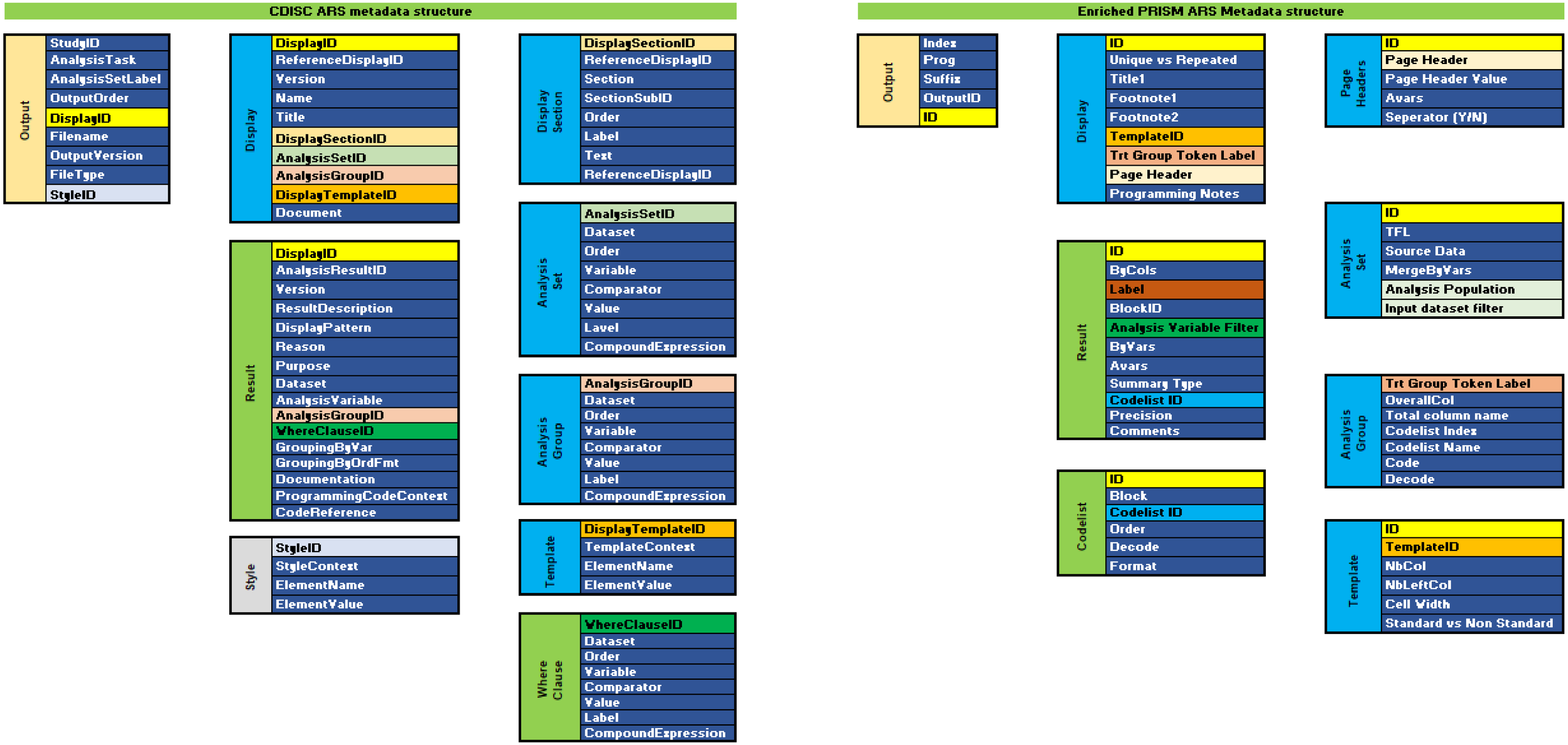
Initial ARS

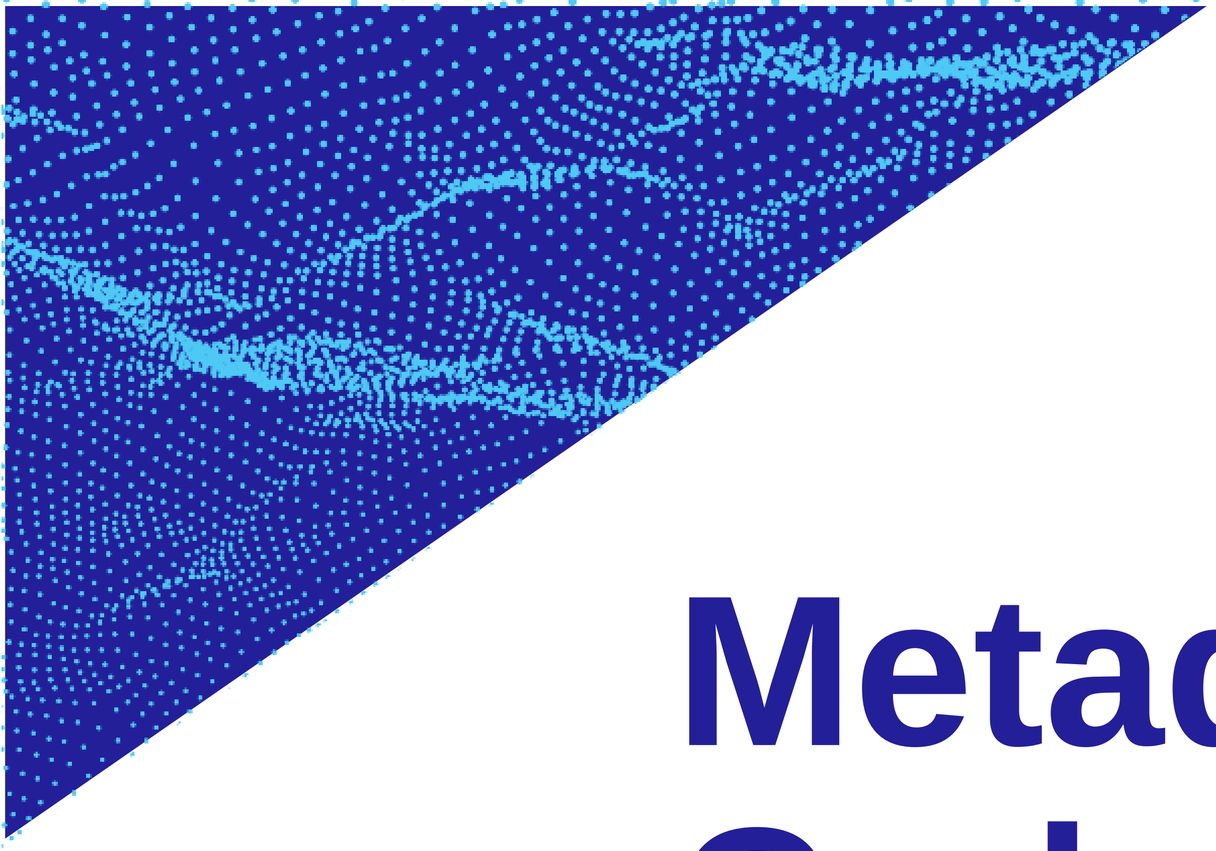
| ID | ByCols | Label | BlockID | Analysis Variable Filter | ByVars | Avars | Summary Type | Codelist ID | Precision | Comments |
|---|---|---|---|---|---|---|---|---|---|---|
| t_01_demo | | Age (Years) | 1 | | | | SUMMARY | | 0 | |
| t_01_demo | | Age Categories, n(%) | 2 | | | | FREQUENCY | Age2 | | |
| t_01_demo | | Sex, n(%) | 3 | | | | FREQUENCY | Sex3 | | |
| t_02_ae1 | | System Organ Class  Preferred Term | 1 | | | | OCCURRENCE | | | |
| t_02_ae1 | | Subjects with at least one TEAE | 2 | | | | FREQUENCY | | | |
| t_04_ex_cycle | | Number of Treatment Cycles (categorized), n (%) | 1 | | | | FREQUENCY | Num1 | | |
| t_04_ex_cycle | | Number of Treatment Cycles | 2 | | | | SUMMARY | | 0 | |
| t_04_ex_cycle | | Duration of Treatment (weeks) | 3 | | | | SUMMARY | | 1 | |

Enriched ARS

| ID | ByCols | Label | BlockID | Analysis Variable Filter | ByVars | Avars | Summary Type | Codelist ID | Precision | Comments |
|---|---|---|---|---|---|---|---|---|---|---|
| t_01_demo | adsl.trt01pn | Age (Years) | 1 | | | age | SUMMARY | | 0 | |
| t_01_demo | adsl.trt01pn | Age Categories, n(%) | 2 | | | agegr1n | FREQUENCY | Age2 | | |
| t_01_demo | adsl.trt01pn | Sex, n(%) | 3 | | | sexn | FREQUENCY | Sex3 | | |
| t_02_ae1 | adsl.trt01pn | System Organ Class  Preferred Term | 1 | | aebodsys, aedecod | | OCCURRENCE | | | |
| t_02_ae1 | adsl.trt01pn | Subjects with at least one TEAE | 2 | | | | FREQUENCY | | | |
| t_04_ex_cycle | adsl.trt01an | Number of Treatment Cycles (categorized), n (%) | 1 | upcase(paramcd) eq 'NUMCYC' | | avalca1n | FREQUENCY | Num1 | | |
| t_04_ex_cycle | adsl.trt01an | Number of Treatment Cycles | 2 | upcase(paramcd) eq 'NUMCYC' | | aval | SUMMARY | | 0 | |
| t_04_ex_cycle | adsl.trt01an | Duration of Treatment (weeks) | 3 | upcase(paramcd) eq 'TRTDURW' | | aval | SUMMARY | | 1 | |

# CDISC ARS vs Enriched PRISM ARS





# Metadata to Code

# Metadata Tag in SAS Template Code

```
%*--- source data;
proc sql noprint;
  create table dsin as select distinct <MERGEBYVAR>
    , &bytrt. format=<BYTRT_FMT><AVAR_REPEAT:START>
    , <AVAR_N> <AVAR_LABEL_N> format=<AVAR_FORMAT_N><AVAR_REPEAT:END>
    from <SOURCE_DATA>(where=(&selpop.));
  %*- list of trt arms (when total column has not to be displayed);
  select distinct(&bytrt.) format=8. into :ltrt separated by ',' from dsin;
quit;

%*--- summary statistics;
%tab( d       = dsin
    , t       = &bytrt. %sysfunc(ifc(%upcase(&total.) ne Y,(&ltrt.),%str()))
    , pn      = <MERGEBYVAR>
    , l_var   = <AVAR_REPEAT:START><AVAR_N> <AVAR_REPEAT:END>
    , v_cont  = [<ACONTVAR_REPEAT:START><ACONTVAR_N> <ACONTVAR_REPEAT:END>]
    , v_cat   = [<ACATVAR_REPEAT:START><ACATVAR_N>(-10002<ACATVAR_CODELIST_REPEAT:START>,
                 <ACATVAR_CODELIST_N><ACATVAR_CODELIST_REPEAT:END>) <ACATVAR_REPEAT:END>]
    , lab_overall=<TOTAL_COLUMN_NAME>
    , label_nmiss=n (missing)
    );
```

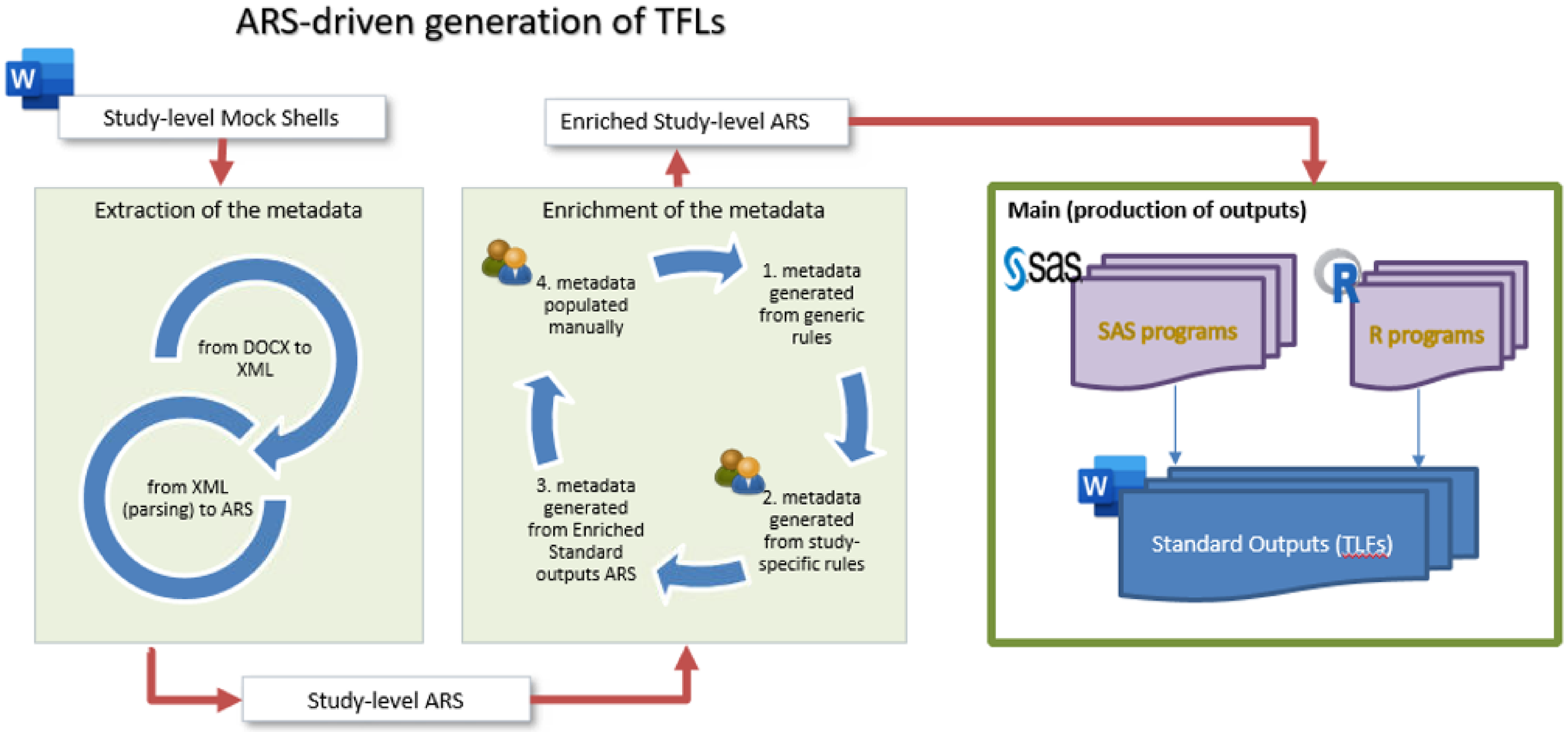
# Metadata Tag in R Template Code

```
#--- summary statistics;
t_demo <-   tab(din           = dsin
                ,L_VAR        = c(<AVAR_REPEAT:START><R_AVAR_N> <AVAR_REPEAT:END>)
                ,LABEL        = c(<AVAR_REPEAT:START><R_AVAR_LABEL_N> <AVAR_REPEAT:END>)
                ,V_CONT       = c(<ACONTVAR_REPEAT:START><R_ACONTVAR_N> <ACONTVAR_REPEAT:END>)
                ,V_CONT_DP    = c(<ACONTVAR_REPEAT:START><R_DP_PRECISION_N> <ACONTVAR_REPEAT:END>)
                ,V_CAT        = c(<ACATVAR_REPEAT:START><R_ACATVAR_N><ACATVAR_REPEAT:END>)
                ,V_CAT_FMT    = c(<ACATVAR_REPEAT:START><R_ACATVAR_FORMAT_N><ACATVAR_REPEAT:END>)
                ,TRT          = c('<R_TREATMENT_VAR>')
                ,TRT_FMT      = c("<R_BYTRT_FMT>")
                ,tableyout    = 2  # Layout of the output data frame 1,2
                )

#--- summary statistics;
t_demo <-   tab(din           = dsin
                ,L_VAR        = c('AGE','AGEGR1N','ETHNICN','WEIGHTBL')
                ,LABEL        = c('Age (years)','Age Categories, n(%)','Sex, n(%)')
                ,V_CONT       = c('AGE')
                ,V_CONT_DP    = c(0)
                ,V_CAT        = c('AGEGR1N','SEXN')
                ,V_CAT_FMT    = c('Age2f','Sex3f')
                ,TRT          = c('TRT01AN')
                ,TRT_FMT      = c("trt1f")
                ,tableyout    = 2  # Layout of the output data frame 1,2
                )
```

# Conclusion

# Conclusion


ARS-driven generation of TFLs
Study-level Mock Shells
Extraction of the metadata
from DOCX to XML
from XML (parsing) to ARS
Study-level ARS
Enriched Study-level ARS
Enrichment of the metadata
4. metadata populated manually
1. metadata generated from generic rules
2. metadata generated from study-specific rules
3. metadata generated from Enriched Standard outputs ARS
Main (production of outputs)
SAS programs
R programs
Standard Outputs (TLFs)

# Thank you.

**Sebastià Barceló, Principal Statistical Programmer**
**Cytel Inc.**
**Route de Pré-Bois 20**
**C.P 1839, 1215 Geneva, SWITZERLAND**
**email: sebastia.barcelobauza@cytel.com**

# Backup Slides

# PRISM ARS (Output)

Initial ARS

| Index | Prog | Suffix | OutputID | ID |
|---|---|---|---|---|
| 1 | t_demo | | t_demo | t_01_demo |
| 2 | t_ae1 | | t_ae1 | t_02_ae1 |
| 3 | t_ae1 | _serious | t_ae1_serious | t_03_ae1_serious |

# PRISM ARS (Display)

Initial ARS

| ID | Unique vs Repeated | Title1 | Footnote1 | Footnote2 | TemplateID | Trt Group Token Label | Page Header | Programming Notes |
|---|---|---|---|---|---|---|---|---|
| t_01_demo | Unique | Table 14.1.2: Demographics - Safety Analysis Set | Percentages are based on the number of non-missing observations. | | | Pbo, Act, Total | | Example of a note for the programmer |
| t_02_ae1 | Unique | Table 14.3.1.2: Treatment Emergent Adverse Events (TEAEs) by System Organ Class and Preferred Term - Safety Analysis Set | System Organ Class and Preferred Terms are sorted in decreasing frequency. | MedDRA Dictionary (Version X.X) is used for coding adverse events. | | Merged Header | | |
| t_03_ae1_serious | Repeated | Table 14.3.1.3: Serious Treatment Emergent Adverse Events (TEAEs) by System Organ Class and Preferred Term - Safety Analysis Set | System Organ Class and Preferred Terms are sorted in decreasing frequency. | MedDRA Dictionary (Version X.X) is used for coding adverse events. | | Merged Header | | |

Enriched ARS

| ID | Unique vs Repeated | Title1 | Footnote1 | Footnote2 | TemplateID | Trt Group Token Label | Page Header | Programming Notes |
|---|---|---|---|---|---|---|---|---|
| t_01_demo | Unique | Table 14.1.2: Demographics - Safety Analysis Set | Percentages are based on the number of non-missing observations. | | Template_DM | Pbo, Act, Total | | Example of a note for the programmer |
| t_02_ae1 | Unique | Table 14.3.1.2: Treatment Emergent Adverse Events (TEAEs) by System Organ Class and Preferred Term - Safety Analysis Set | System Organ Class and Preferred Terms are sorted in decreasing frequency. | MedDRA Dictionary (Version X.X) is used for coding adverse events. | Template_OCC1 | Merged Header | | |
| t_03_ae1_serious | Repeated | Table 14.3.1.3: Serious Treatment Emergent Adverse Events (TEAEs) by System Organ Class and Preferred Term - Safety Analysis Set | System Organ Class and Preferred Terms are sorted in decreasing frequency. | MedDRA Dictionary (Version X.X) is used for coding adverse events. | Template_OCC1 | Merged Header | | |

# PRISM ARS (Template)

Initial ARS

| ID | TemplateID | NbCol | NbLeftCol | Cell Width | Standard vs Non Standard |
|---|---|---|---|---|---|
| t_01_demo | | 4 | 1 | %cellwidth | Standard |
| t_02_ae1 | | 4 | 1 | %cellwidth | Standard |
| t_03_ae1_serious | | 4 | 1 | %cellwidth | Standard |

Enriched ARS

| ID | TemplateID | NbCol | NbLeftCol | Cell Width | Standard vs Non Standard |
|---|---|---|---|---|---|
| t_01_demo | Template_DM | 4 | 1 | %cellwidth | Standard |
| t_02_ae1 | Template_OCC1 | 4 | 1 | %cellwidth | Standard |
| t_03_ae1_serious | Template_OCC1 | 4 | 1 | %cellwidth | Standard |

# PRISM ARS (Page Headers)

Initial ARS

| ID | Page Header | Page Header Value | Avars | Seperator (Y/N) |
|---|---|---|---|---|
| t_12_vs_chg1 | Vital Signs Parameter (unit): Parameter1 (Unit1) | Vital Signs Parameter (unit) | | 0 |
| l_17_ae | Subject: xxxxxxxx / Country: xxxxx / Race: xxxxxx / Sex: xxx / Age (years): xx / Weight (kg): xxx.x | Subject | | 1 |
| l_17_ae | Subject: xxxxxxxx / Country: xxxxx / Race: xxxxxx / Sex: xxx / Age (years): xx / Weight (kg): xxx.x | / Country | | 1 |
| l_17_ae | Subject: xxxxxxxx / Country: xxxxx / Race: xxxxxx / Sex: xxx / Age (years): xx / Weight (kg): xxx.x | / Race | | 1 |
| l_17_ae | Subject: xxxxxxxx / Country: xxxxx / Race: xxxxxx / Sex: xxx / Age (years): xx / Weight (kg): xxx.x | / Sex | | 1 |
| l_17_ae | Subject: xxxxxxxx / Country: xxxxx / Race: xxxxxx / Sex: xxx / Age (years): xx / Weight (kg): xxx.x | / Age (years) | | 1 |
| l_17_ae | Subject: xxxxxxxx / Country: xxxxx / Race: xxxxxx / Sex: xxx / Age (years): xx / Weight (kg): xxx.x | / Weight (kg) | | 1 |

Enriched ARS

| ID | Page Header | Page Header Value | Avars | Seperator (Y/N) |
|---|---|---|---|---|
| t_12_vs_chg1 | Vital Signs Parameter (unit): Parameter1 (Unit1) | Vital Signs Parameter (unit) | strip(VSPARAM) \|\| " (" \|\| strip(VSPARAMU) \|\| ")" | 0 |
| l_17_ae | Subject: xxxxxxxx / Country: xxxxx / Race: xxxxxx / Sex: xxx / Age (years): xx / Weight (kg): xxx.x | Subject | usubjid | 1 |
| l_17_ae | Subject: xxxxxxxx / Country: xxxxx / Race: xxxxxx / Sex: xxx / Age (years): xx / Weight (kg): xxx.x | / Country | strip(country) | 1 |
| l_17_ae | Subject: xxxxxxxx / Country: xxxxx / Race: xxxxxx / Sex: xxx / Age (years): xx / Weight (kg): xxx.x | / Race | race | 1 |
| l_17_ae | Subject: xxxxxxxx / Country: xxxxx / Race: xxxxxx / Sex: xxx / Age (years): xx / Weight (kg): xxx.x | / Sex | sex | 1 |
| l_17_ae | Subject: xxxxxxxx / Country: xxxxx / Race: xxxxxx / Sex: xxx / Age (years): xx / Weight (kg): xxx.x | / Age (years) | compress(put(age,3.)) | 1 |
| l_17_ae | Subject: xxxxxxxx / Country: xxxxx / Race: xxxxxx / Sex: xxx / Age (years): xx / Weight (kg): xxx.x | / Weight (kg) | compress(put(weightbl,best8.)) | 1 |