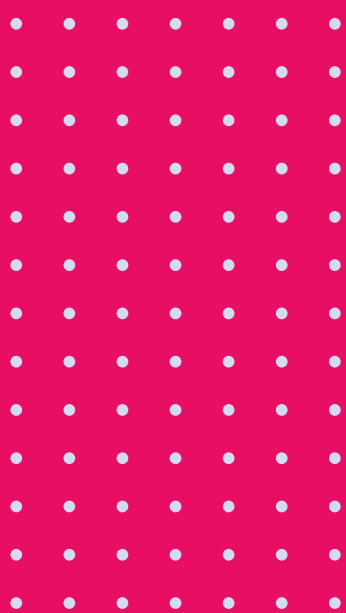




Code Generation

Turning metadata into SAS, R or Python

Stuart Malcolm
Head of Standards, Efficiency and Automation
Veramed.

AD17 PHUSE EU Connect 2023

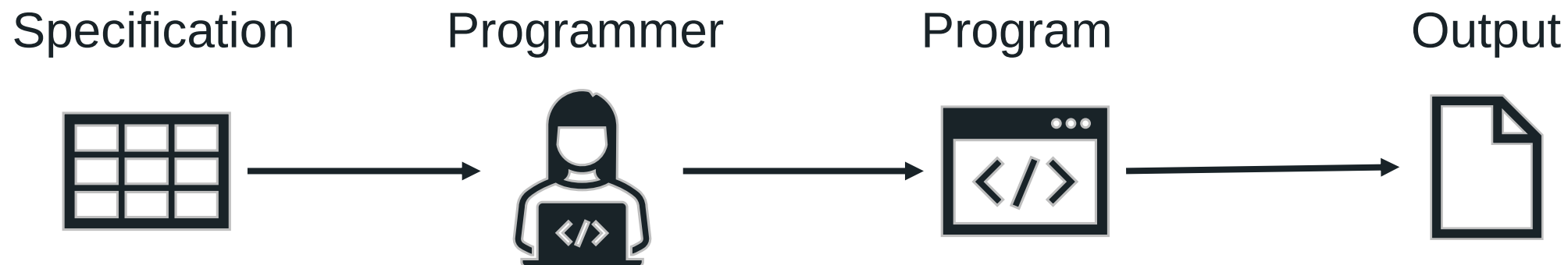


The vision

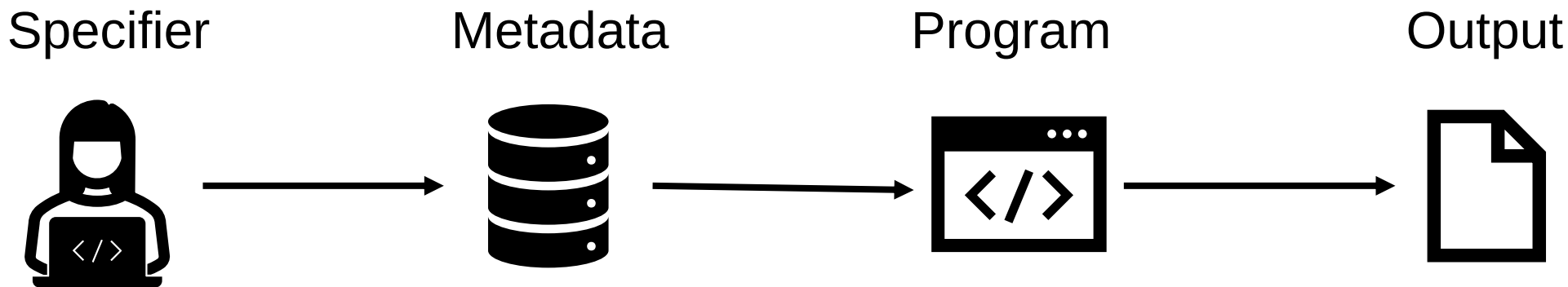
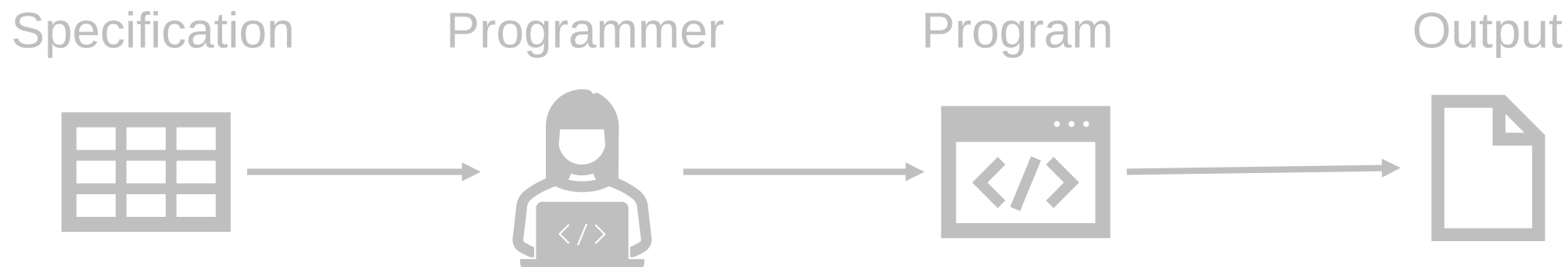
Standards-based metadata-driven automation is a critical component in realizing the primary benefits ... substantially improved efficiency, consistency, and re-usability across the clinical research data lifecycle.

<https://www.cdisc.org/cdisc-360>

From manual programming



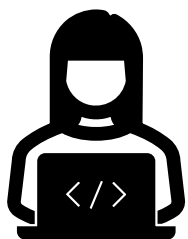
From manual programming to automation



Automation capabilities and requirements



Specifier



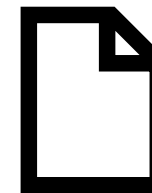
Metadata



Program



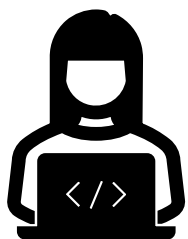
Output



Automation capabilities and requirements



Specifier



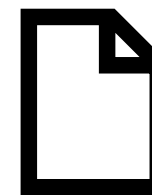
Metadata



Program



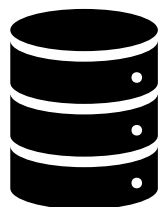
Output



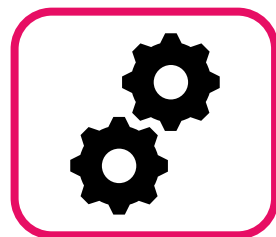
Code Generation



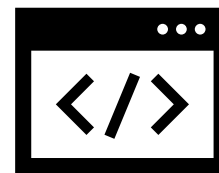
Metadata



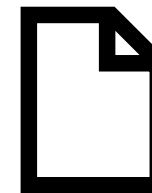
Generator



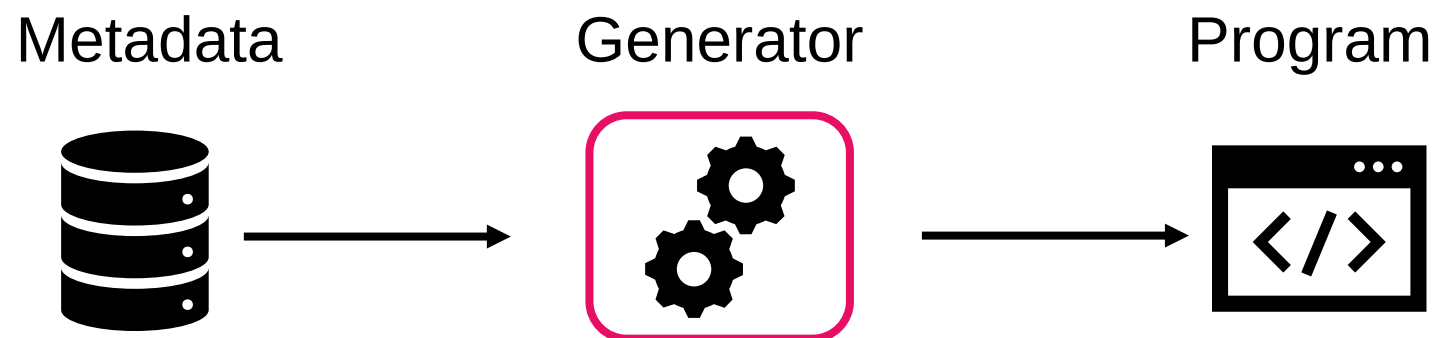
Program



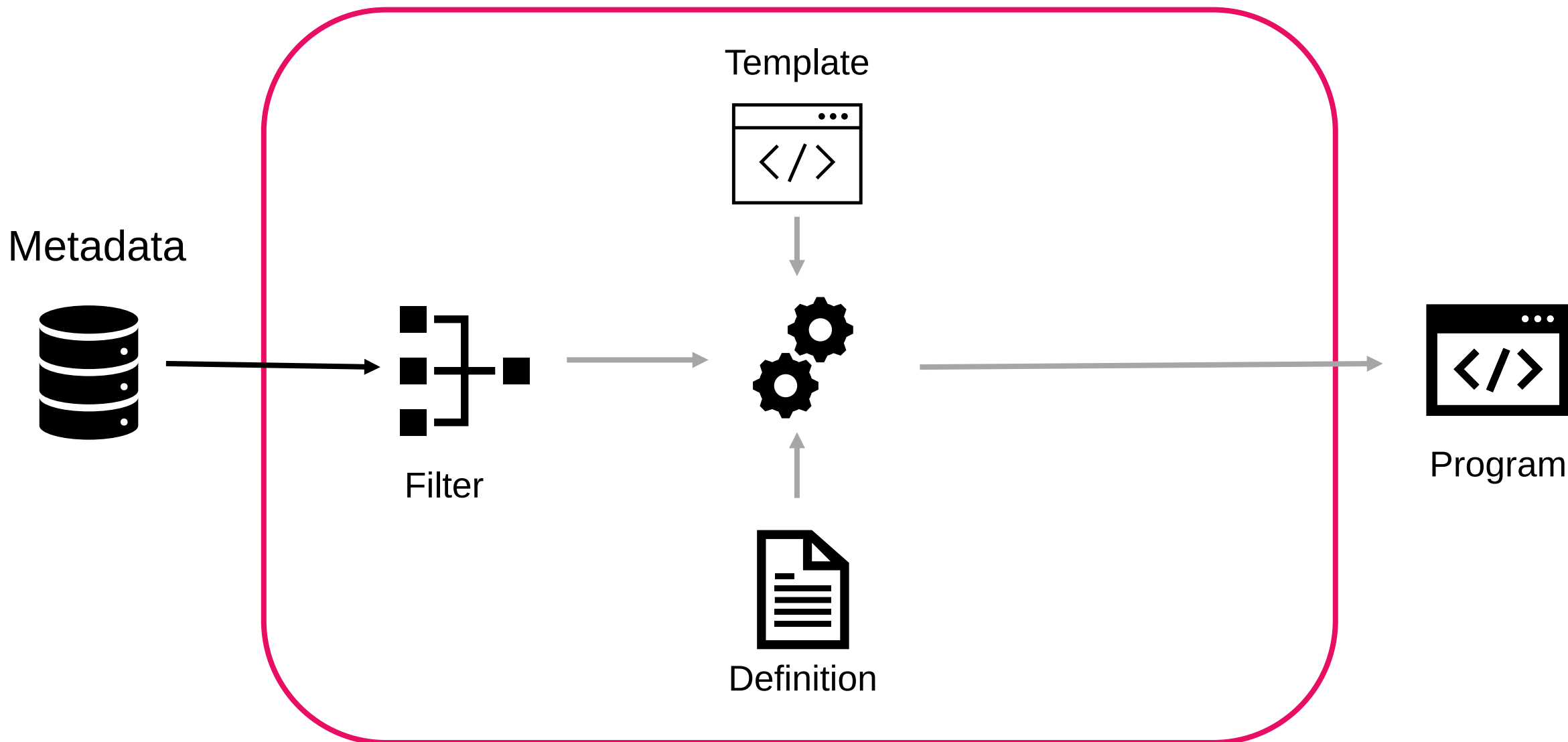
Output



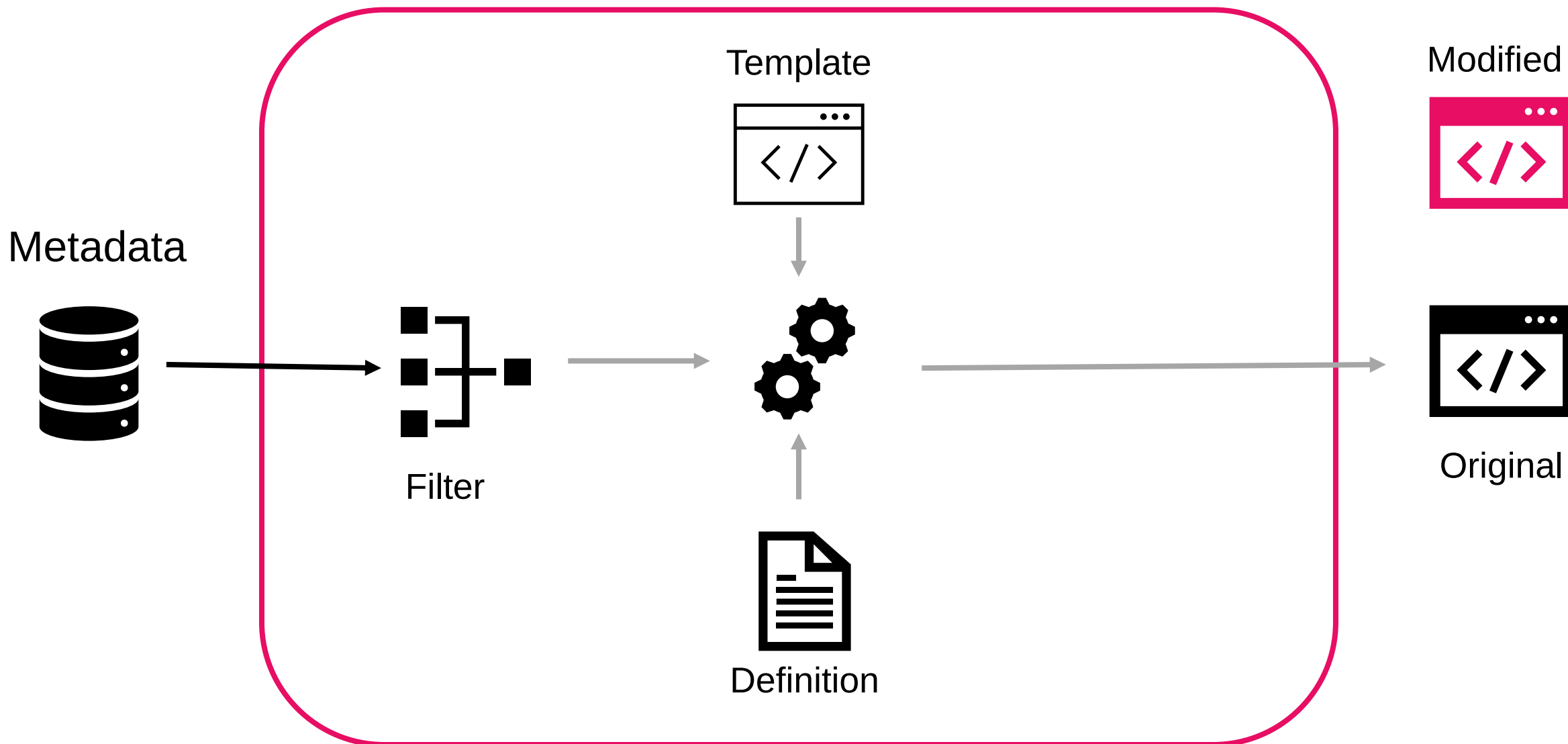
Metadata-driven code generation



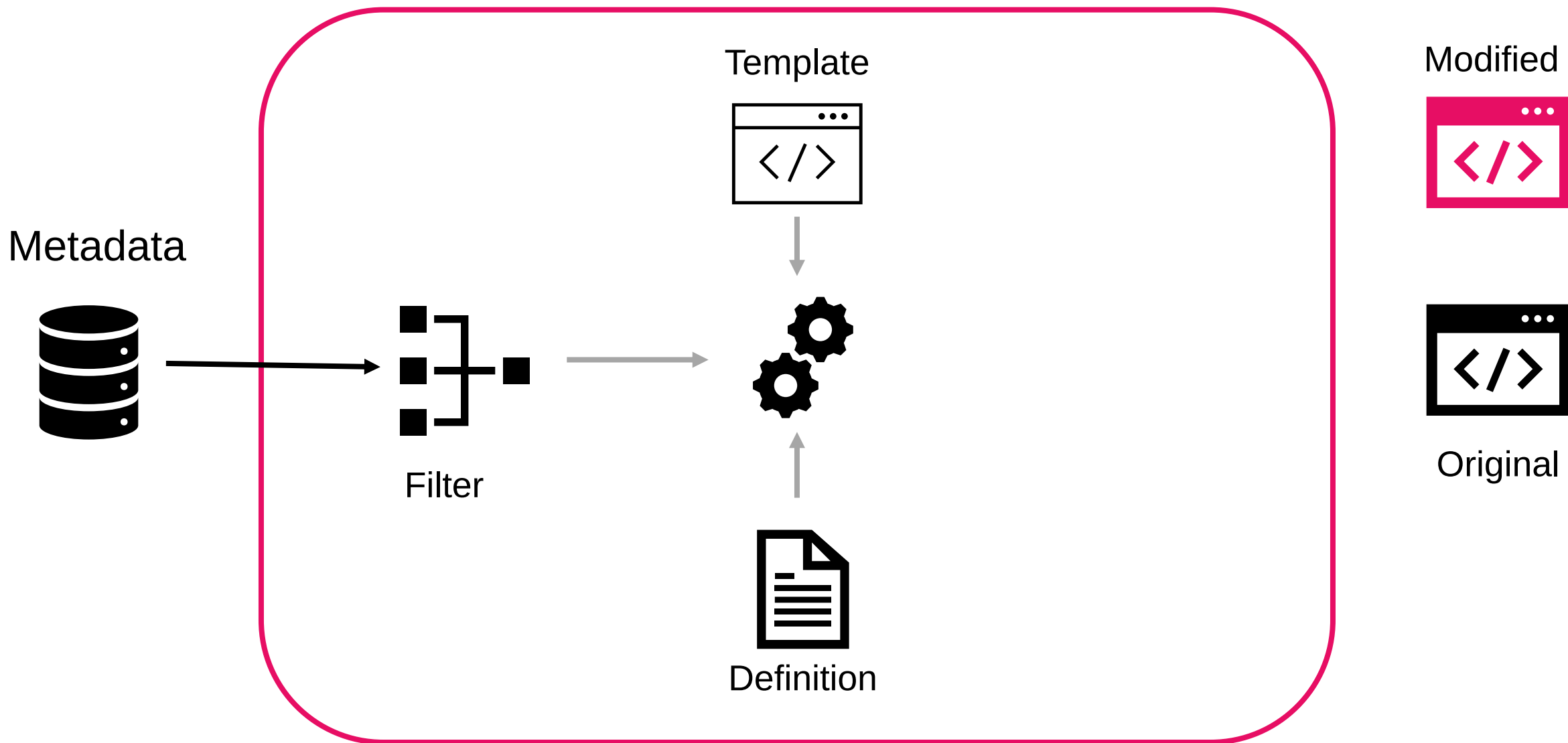
Basic components of a code generator



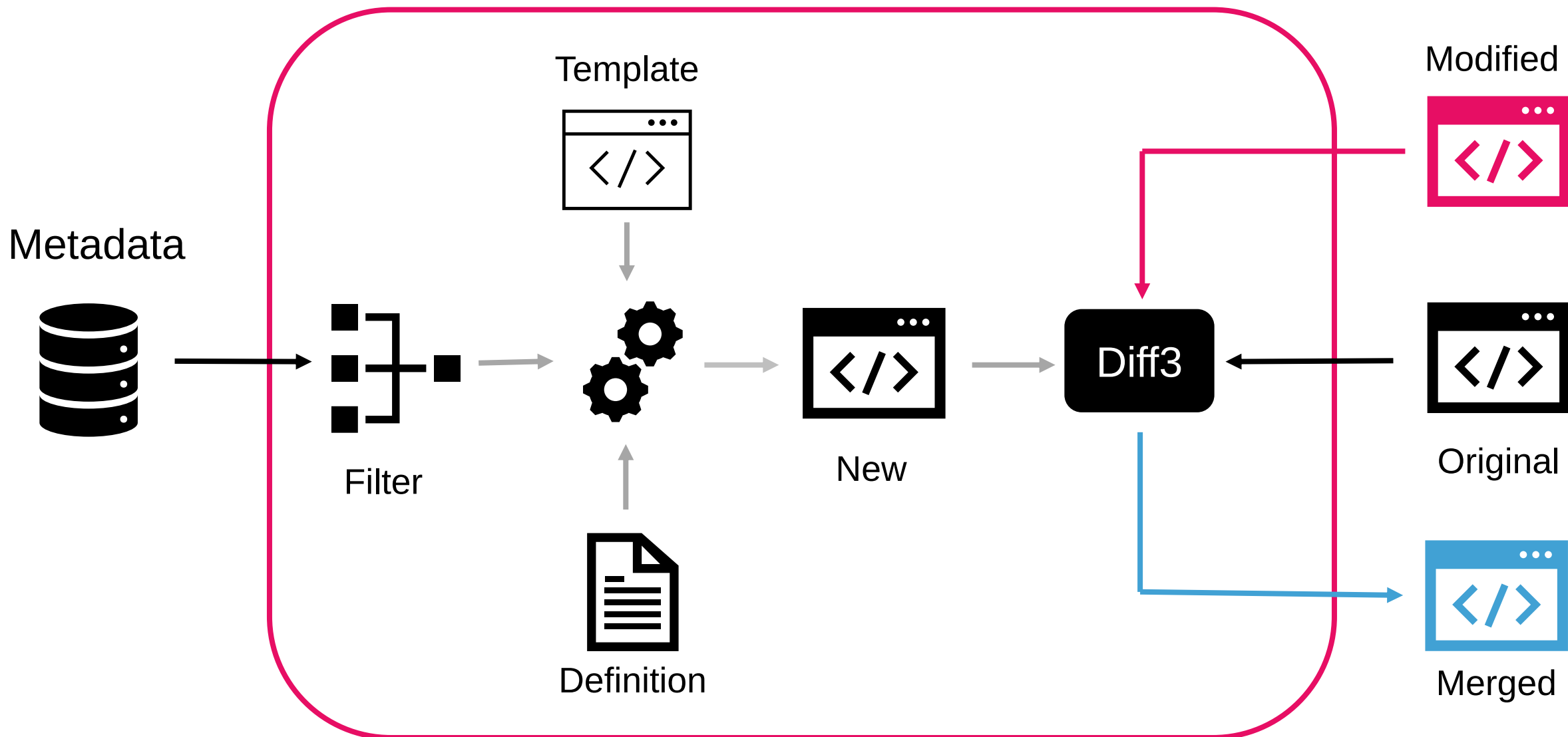
Manual edits of generated programs

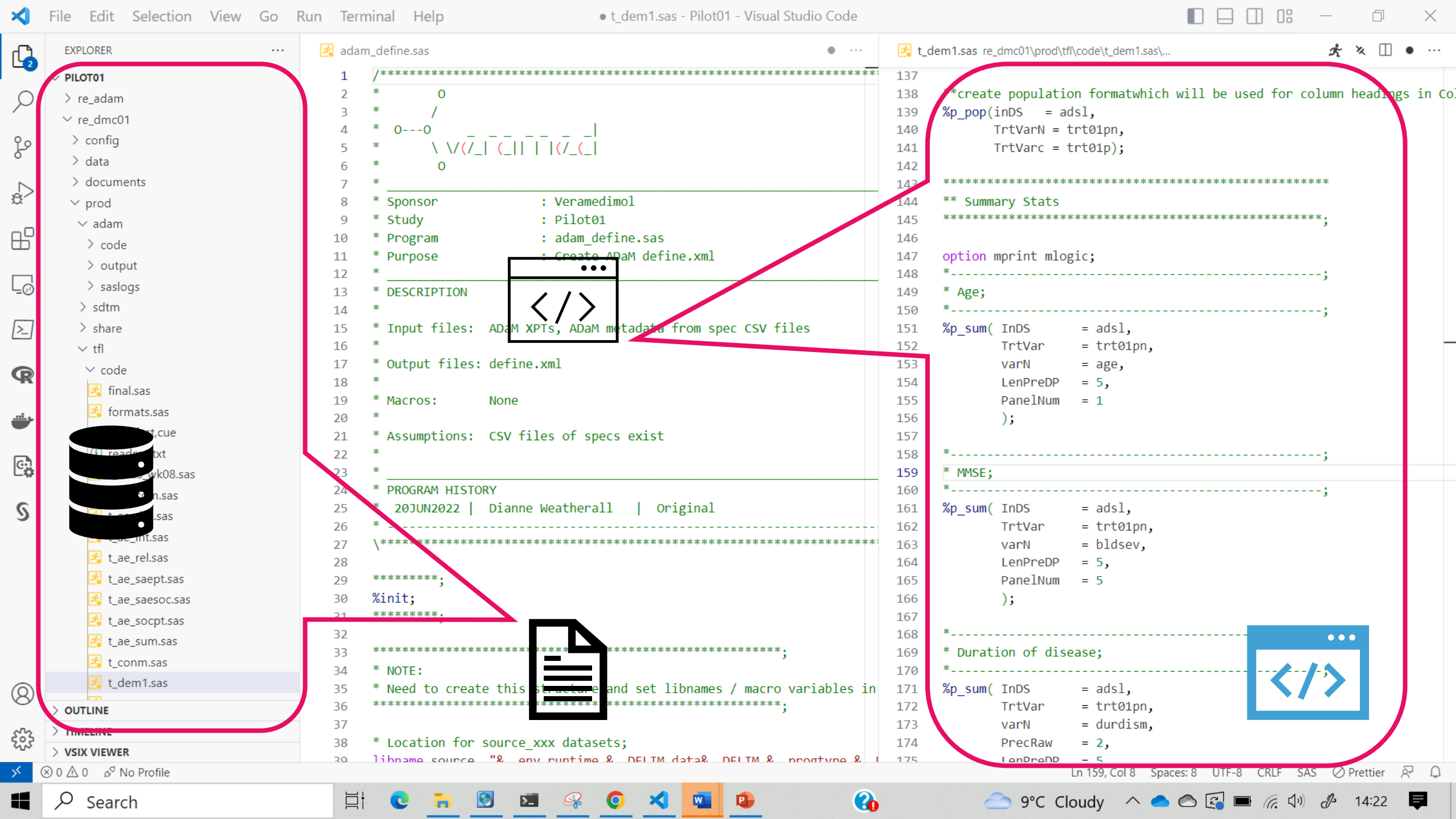


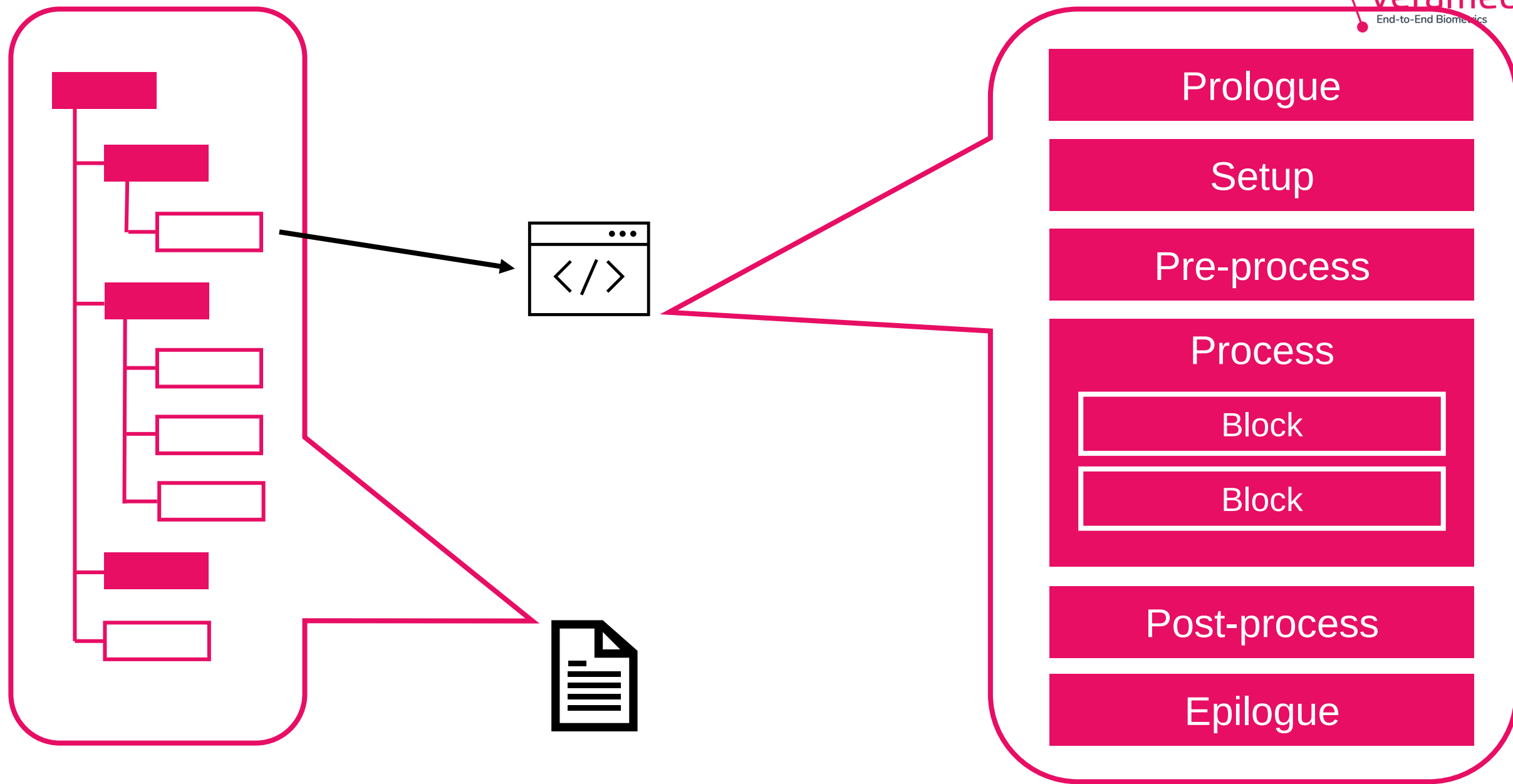
Manual edits of generated programs break e2e automation



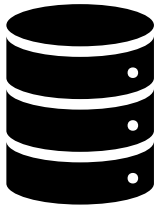
Use Diff3 to merge original, modified and new programs







Metadata



Prologue

Setup

Pre-process

Process

Block

Block

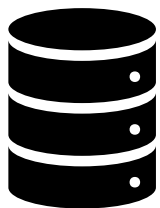
Post-process

Epilogue

Program



Metadata



Program



```
{  
  "name": "Stuart",  
  ds: [  
    'DM', 'AE', 'ADLB'  
  ]  
}
```

```
# Example1.r  
  
{# Variables}  
  
# Name: {{name}}  
  
{# Filters }  
  
name <-  
"{{name|upper}}"
```

```
# Example1.r
```

```
# Name: Stuart  
name <- "STUART"
```

```
{  
  "name": "Stuart",  
  ds: [  
    'DM', 'AE', 'ADLB'  
  ]  
}
```

```
* Example2.sas;  
  
{%for d in ds %}  
data {{d}};  
set  
  {%if d=='ADSL' %}  
  ADAM.{{d}};  
  {% else %}  
  SDTM.{{d}};  
  {% endif %}  
run;  
{% endfor %}
```

```
* Example2.sas;
```

```
data DM;  
  Set SDTM.DM;  
run;  
data AE;  
  Set SDTM.AE;  
run;  
data ADSL;  
  Set ADAM.ADSL;  
run;
```

```
{  
  "name": "Stuart",  
  ds: [  
    'DM', 'AE', 'ADLB'  
  ]  
}
```

```
{# base }  
{%block header}  
# By: {{name}}  
{%endblock%}  
{%block body}  
# Code here  
{%endblock%}  
---  
# {{d|lower}}.py  
{%extends base%}  
{%block body%}  
foo('{{d}}')
```

```
# dm.py  
# By: Stuart  
foo('DM')
```

```
# ae.py  
# By: Stuart  
foo('AE')
```

```
# adsl.py  
# By: Stuart  
foo('ADLB')
```

Template (Jinja2)

Program (SAS)

[illegible]

```

1
2 /*****
3 *
4 * ..- - - - -
5 * ..- - - - -
6 * ..- - - - -
7 * ..- - - - -
8 *
9 * Study ..- - - - - : CDISCPILOT01
10 * Analysis ..- - - - - : PHUSE_EU_CONNECT_2022
11 * Program ..- - - - - : q_t_hema.sas
12 *
13 * DESCRIPTION
14 *
15 * This program creates Display T14030401_SAF_HEMA
16 * -- Table 14.3.4.1 Summary of Hematology Parameters --
17 *
18 * Input files:
19 * -- ADAM.ADLB
20 *
21 * Output file:
22 * -- t14_03_04_01_hema.rtf
23 *
24 * Macros:
25 * -- [TODO]
26 *
27 * Assumptions:
28 * -- none.
29 *

```

Metadata (JSON)

```
{
  "Study": "CDISCILOT01",
  "Analysis": "PHUSE_EU_CONNECT_2022",
  "Group": "Safety",
  "Program": "q_t_hema",
  "Filename": "t14_03_04_01_hema",
  "Type": "rtf",
  "Order": 1,
  "DisplayID": "T14030401_SAF_HEMA",
  "DisplayVersion": 1,
  "StyleID": "table_rtf",
  "Display": {
    "Active": "Y",
    "ResultDisplayOID": "T14030401_SAF_H",
    "ResultDisplayParentOID": null,
    "Version": 1,
    "GroupingDataset": "ADAM.ADSL",
    "GroupingWhereVar": "SAFFL",
    "GroupingWhereComparator": "EQ",
    "GroupingWhereValue": "\\Y\\",
    "GroupingAnalysisVar": "TRTP_DOSEPC",
    "GroupingAnalysisVarOrdFmt": "TRTPN",
    "GroupingByVar": null,
    "GroupingByOrdFmt": null,
    "DisplayTemplate": "template1",
    "DisplayName": "Table 14.3.4.1",
    "DisplayTitle": "Summary of Hematolc",
    "Title1": "Dose Escalation",
    "Title2": "#byline",
    "Title3": "not-used",
    "Title4": "not-used",
    "RowLabelHeader": "Parameter (Unit)",
    "Header1": "Post-Baseline Results",
    "Header2": null,
    "Header3": null,
    "Header4": null,
    "Footer1": "Note: Baseline is taken",
    "Footer2": "Reference: Listing 16.2.",
    "Footer3": null,
    "Footer4": null,
    "Document": null,
    "Results": [
      {
        "ResultDisplayOID": "T14030401_SAF_H",
        "AnalysisResultOID": "T14030401_SAF_H",
        "Version": 1,
        "ResultDescription": "Summary of Hematolc"
      }
    ]
  }
}
```

Template (Jinja2)

```
49
50 # =====
51 # Tasks to build individual TFL
52 # =====
53
54 {% for Output in Outputs %}
55 # -----
56 # Output: {{ Output.DisplayID }}
57 # {{ Output.Display.DisplayName }}-{{ Output.Display.DisplayTitle }}
58 # -----
59
60 Task-{{ Output.Filename }} @{
61   Inputs = @(
62     "$DataPath\{{ Output.Display.GroupingDataset|replace(' ','_') }}.sas7bdat",
63     {% for dataset in Output.Display.Results|map(attribute='Dataset') %}
64       "$DataPath\{{ dataset|replace('ADAM.','') }}.sas7bdat",
65     {% endfor %}
66     "{{ Output.Program }}.sas"
67   )
68   Outputs = "..\output\q_{{ Output.Filename }}.{{ Output.Type }}"
69   Jobs = {
70     $r = viper-run-sas {{ Output.Program }}
71     $script:results += $r
72     write-host $r
73   }
74 }
75 {% endfor %}
76
77 # =====
78 # Task that can be used to build all the TFL datasets..
79 # ..or specific ones (using the TFLID)
80 # =====
81 Task TFL-{{ for Output in Outputs %}}-{{ Output.Filename }} @{
82   {% if not loop.last %}, {% endif %}
83   {% endfor %}
84 # =====
85 # clean all (TFL) outputs
86 # =====
87 Task Clean-{{ for Output in Outputs %}}-{{ Output.Filename }} @{
88   write-output "Deleting RTF output files"
89   {% for Output in Outputs %}
90     del ..\output\{{ Output.Filename }}.rtf
91   {% endfor %}
92   write-output "[TODO]--delete reporting datasets"
93 }
94
```

Program (PowerShell)

```
49
50 # =====
51 # Tasks to build individual TFL
52 # =====
53
54 # -----
55 # Output: T14030401_SAF_HEMA
56 # Table 14.3.4.1 Summary of Hematology Parameters
57 # -----
58
59 Task t14_03_04_01_hema @{
60   Inputs = @(
61     "$DataPath\ADSL.sas7bdat",
62     "$DataPath\ADLB.sas7bdat",
63     "q_t_hema.sas"
64   )
65   Outputs = "..\output\q_t14_03_04_01_hema.rtf"
66   Jobs = {
67     $r = viper-run-sas q_t_hema
68     $script:results += $r
69     write-host $r
70   }
71 }
72
73 # -----
74 # Output: T14030407_CHEM
75 # Table 14.3.4.7 Summary of Chemistry Parameters
76 # -----
77
78 Task t14_03_04_07_chem @{
79   Inputs = @(
80     "$DataPath\ADSL.sas7bdat",
81     "$DataPath\ADLB.sas7bdat",
82     "q_t_chem.sas"
83   )
84   Outputs = "..\output\q_t14_03_04_07_chem.rtf"
85   Jobs = {
86     $r = viper-run-sas q_t_chem
87     $script:results += $r
88     write-host $r
89   }
90 }
91
92 # -----
93 # Output: T14030413_COAG
94
```

Thanks!

Questions?

Stuart.Malcolm@Veramed.co.uk

<https://www.linkedin.com/in/stuart-malcolm/>

