

AD11

Automation of the CDISC Pilot Study

Stuart Malcolm

Head of Standards, Efficiency and Automation

15-NOV-2022



1.Introduction

Context and Architecture

2.Methodology

Creation and Validation

3.Metadata

Specs and Pipelines

4.Code Generation

The DOM and Templates



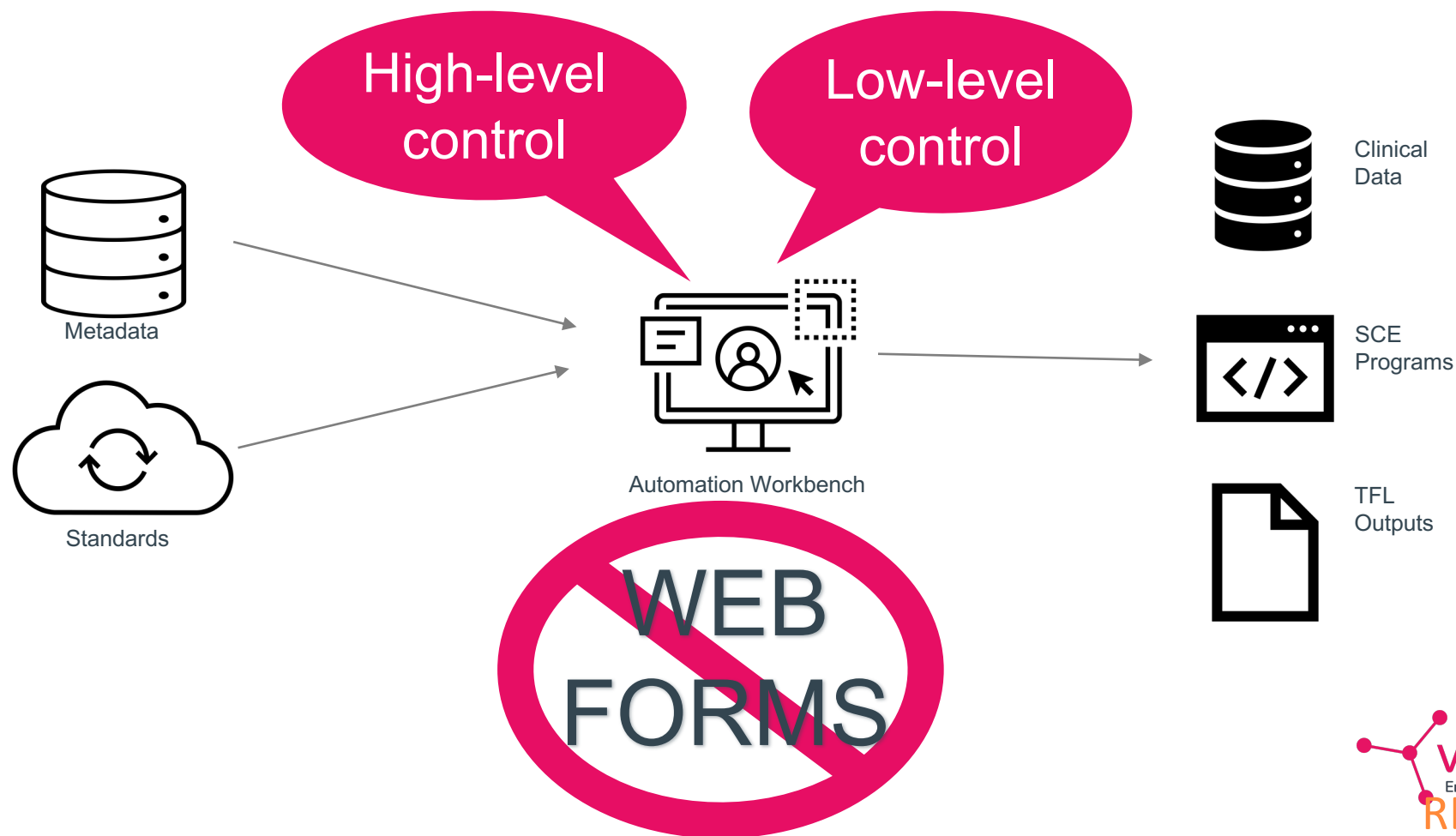
What is 360?

360 is the CDISC initiative aimed at implementing standards as linked metadata with a conceptual foundation providing the additional semantics needed to support metadata driven-automation across the end-to-end clinical research data lifecycle.

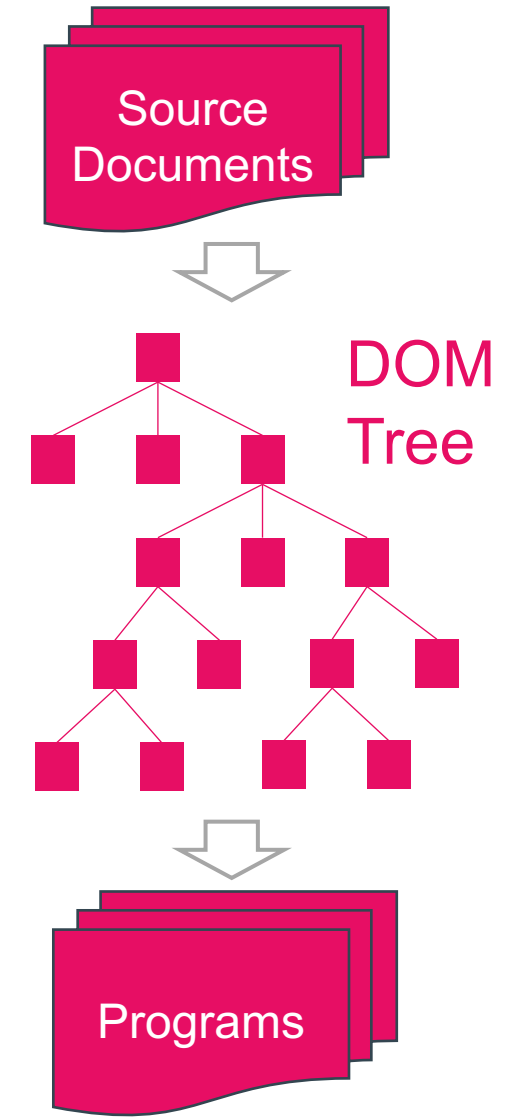
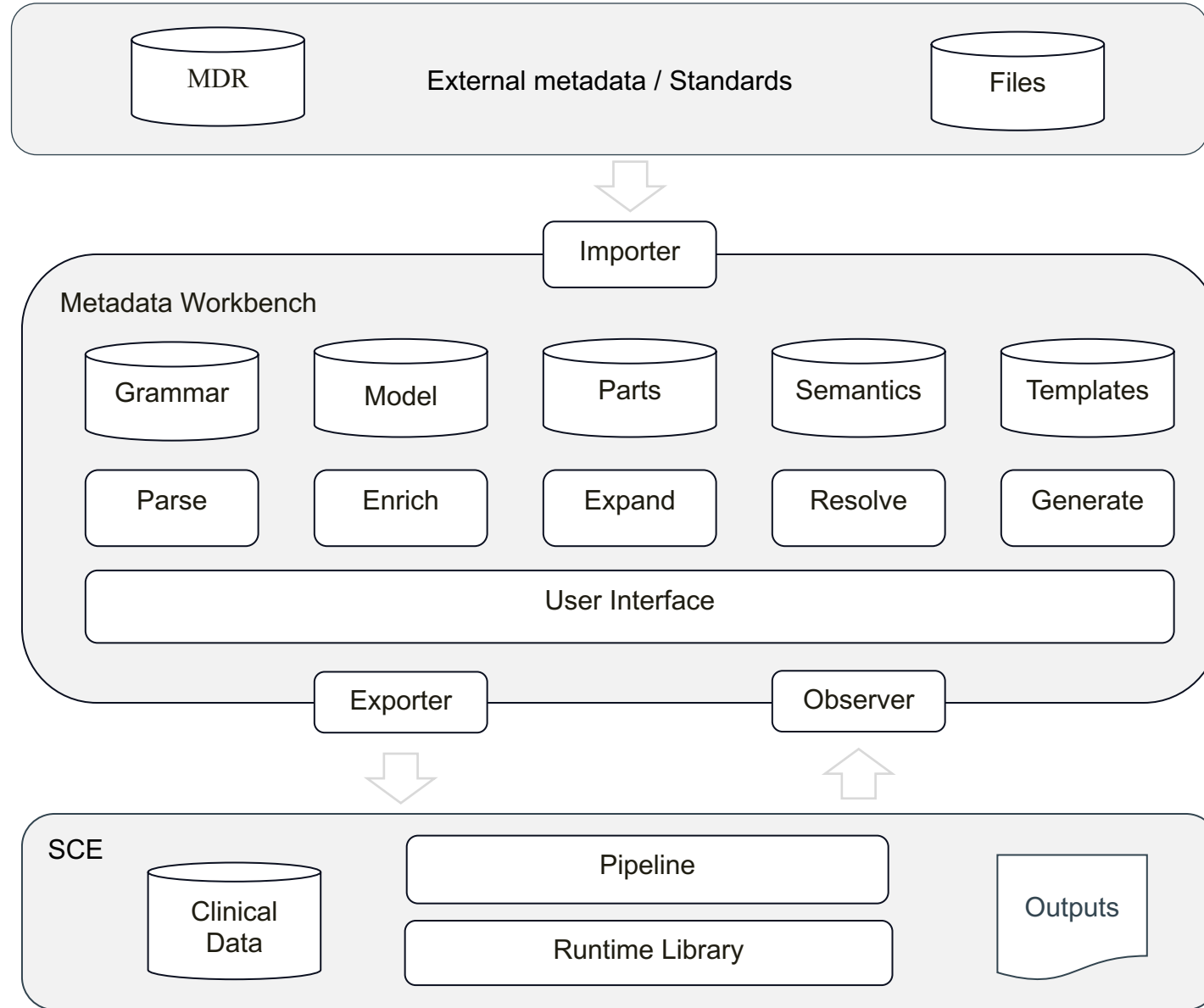
New software tools will be able to consume this new metadata to ease standards implementations while increasing data processing efficiencies.

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

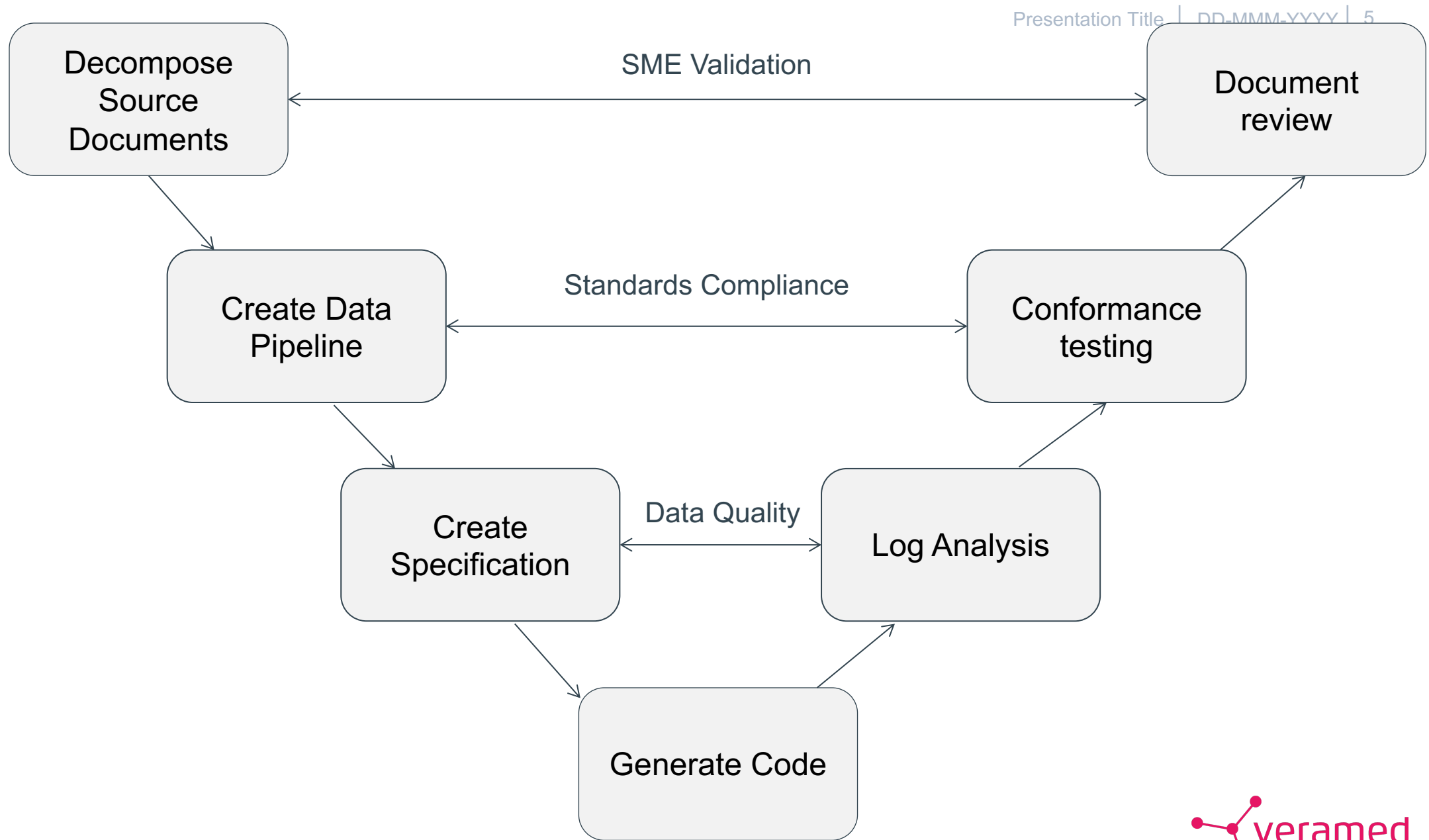
These benefits drive the return on investment in the CDISC standards implementations expected by CDISC stakeholders.

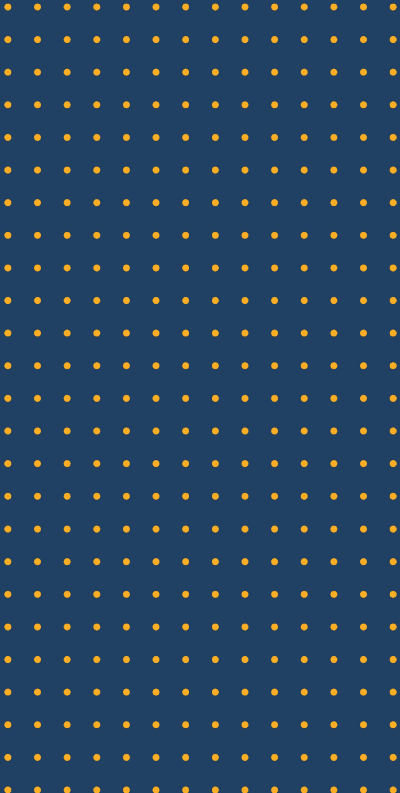


Architecture



Methodology





Metadata

Specifications and Pipelines

ADSL (Subject-Level Analysis Dataset Structure)

5. SAMPLE SIZE CONSIDERATIONS

Approximately 100 patients will be randomized to each of the 3 treatment groups. Previous experience with the oral formulation of xanomeline suggests that this sample size has 90% power to detect a 3.0 mean treatment difference in ADAS-Cog ($p < .05$, two-sided), based on a standard deviation of 6.5. Furthermore, this sample size has 80% power to detect a 0.36 mean treatment difference in CIBIC+ ($p < .05$, two-sided), based on a standard deviation of 0.9.

6. ANALYSIS POPULATIONS

For this study, the following definitions are used:

Screen Failures

Patients entered into the study are those from whom informed consent for the study has been obtained. Patients entered into the study but not assigned to a treatment group are considered to be screen failures. Demographic data for screen failures will be included in the data tabulation datasets, but not in the analysis datasets or in the analyses.

ADSL.RANDFL Randomized

Patients who are enrolled in the study are those who have been assigned to a treatment group. Patients who are entered into the study but fail to meet criteria specified in the protocol for treatment assignment will not be enrolled in the study. Patients are randomly assigned to treatment groups at Week 0 (Visit 3).

ADSL.ITTFLL ITT Population

All patients randomized

ADSL.SAFFLL Safety population

All patients randomized and known to have taken at least one dose of randomized drug

Efficacy population

All patients who were randomized and took at least 1 post-baseline measure for both ADAS-Cog and CIBIC+.

ADSL.COMPLFL Completers

All patients in the efficacy population who completed the Week 24 visit (Visit 12).

7. GENERAL CONSIDERATIONS FOR DATA ANALYSES

All statistical tests will be 2-sided with a significance level of 0.05. One-sided p-values will not be reported. Summary statistics for continuous variables will include the number of observations, mean, standard deviation, median, minimum, and maximum. Summary statistics for categorical variables will include the frequency and percentage.

```
{
  "document": {
    "uri": "/path/to/source/docs/cdiscpilot01_sap.pdf",
    "md5": "6fc234ad3752e1267b34fb12bcd6718b",
    "annotationGroupSets": [
      {
        "annotationGroup": {
          "name": "ADSL",
          "description": "Subject-Level Analysis Dataset Structure",
          "annotation": [
            { "title": "ADSL.RANDFL", "page": "8" },
            { "title": "ADSL.ITTFLL", "page": "8" },
            { "title": "ADSL.SAFFLL", "page": "8" },
            { "title": "ADSL.EFFFL",
              "page": "8",
              "body": {
                "label": "Efficacy Population Flag",
                "core": "Cond",
                "simpleDataType": "Char",
                "valueList": [ "Y", "N" ]
              }
            },
            { "title": "ADSL.COMPLFL" }
          ]
        }
      ]
    }
  }
}
```

ADSL.EFFFL:

label: "Efficacy..."

Core: "Cond"

simpleDatatype: "Char"

valueList:

- "Y"

- "N"

RefData (Reference Data)

8.2. Assessment Windows

In general, assessments will be assigned to visits as collected on the CRFs, and will disregard the actual date of the assessment. For example, if an assessment is recorded on the Visit 10 CRF page, the assessment will be assigned to Week 16 (Visit 10).

The ADAS-Cog (11), CIBIC+, and NPI-X assessments will also be assigned to visits based on the actual visit dates, as will laboratory assessments. Actual visit days will be determined relative to the date of randomization, using the algorithm {day = visit date – randomization date}. If multiple assessments fall into the same visit window (windows defined in following table), then the assessment closest to the target day will be selected. [Note that retrieval visits (visit number 201) are included for the purpose of selecting assessments for the week 24 visit window.] If two assessments are equidistant from the target day, then the assessment prior to the target day will be selected. In situ where imputation of missing values is also involved, imputation will use the assessments within the windows.

Variable	Scheduled Visit	Time Interval (label on output)	Time Interval (Day)	Target Time Point (Day)
ADAS-Cog CIBIC+	3	Baseline	≤ 1	1
	8	Week 8	2-84	56
	10	Week 16	85-140	112
	12	Week 24	>140	168
NPI-X	3	Baseline	≤ 1	1
	4	Week 2	2-21	14
	5	Week 4	22-35	28
	7	Week 6	36-49	42
	8	Week 8	50-63	56
	8.1	Week 10 (Tel)	64-77	70
	9	Week 12	78-91	84
	9.1	Week 14 (Tel)	92-105	98
	10	Week 16	106-119	112
	10.1	Week 18 (Tel)	120-133	126
	11	Week 20	134-147	140
	11.1	Week 22 (Tel)	148-161	154
	12	Week 24	162-175	168
	13	Week 26	>175	182

Windows:
type: "Excel"
file: "assessments.xlsx"
sheet: "Windows"

```

{
  "document": {
    "uri": "/path/to/source/docs/cdiscpilot01_sap.pdf",
    "md5": "6fc234ad3752e1267b34fb12bcd6718b",
    "annotationGroupSets": [
      {
        "annotationGroup": {
          "name": "RefData",
          "description": "Reference Data",
          "annotation": [
            {
              "title": "Windows",
              "page": "10",
              "body": {
                "type": "Excel",
                "file": "assessments.xlsx",
                "sheet": "Windows",
                "md5": "70504eaf1db01879cabb6d083bdae179"
              }
            }
          ]
        }
      }
    ]
  }
}

```

In general, assessments will be assigned to visits as collected on the CRFs, and will disregard the actual date of the assessment. For example, if an assessment is recorded on the Visit 10 CRF page, the assessment will be assigned to Week 14 (Visit 10).

The ADAS-Cog (11), CIBIC+, and NPI-X assessments will also be assigned to visits based on the actual visit dates, as well as laboratory assessments. Actual visit dates will be determined relative to the date of randomization, using the algorithm (day = visit date - randomization date). If multiple assessments fall into the same visit window (windows defined in following table), then the assessment falling in the target day will be selected. (Note that several visits (visit number 20) are included for the purpose of selecting assessments for the week 24 visit window. If two assessments are equidistant from the target day, then the assessment prior to the target day will be selected. In situations where imputation of missing values is also involved, imputation will use the assessments within the windows.

Variable	Scheduled Visit	Time Interval (label on output)	Time Interval (Day)	Target Visit Point (Day)
ADAS-Cog	3	Baseline	<2	1
	8	Week 8	2-84	56
	10	Week 16	85-140	112
	12	Week 24	>140	168
NPI-X	3	Baseline	<2	1
	4	Week 2	2-21	14
	5	Week 4	22-35	28
	7	Week 6	36-49	42
	8	Week 8	50-63	56
	8.1	Week 10 (Tet)	64-77	70
	9	Week 12	78-91	84
	9.1	Week 14 (Tet)	92-105	98
	10	Week 16	106-119	112
	10.1	Week 18 (Tet)	120-133	126
	11	Week 20	134-147	140
	11.1	Week 22 (Tet)	148-161	154
	12	Week 24	162-175	168
	13	Week 26	>175	182

FINAL SAP

- 10 -

```
{
  "document": {
    "uri": "/path/to/source/docs/cdiscpilot01_sap.pdf",
    "md5": "6fc234ad3752e1267b34f012bc067180",
    "annotationGroupSets": [
      {
        "name": "RefData",
        "description": "Reference Data",
        "annotation": {
          "title": "Windows",
          "page": "10",
          "body": {
            "type": "Excel",
            "file": "assessments.xlsx",
            "sheet": "Windows",
            "md5": "78584eaf1d0b1879cabb6d083bdae179"
          }
        }
      }
    ]
  }
}
```

```
{
  "Assessment": "ADAS-Cog",
  "EPOCH": "Treatment",
  "VISITNUM": 12,
  "VISIT": "WEEK 24",
  "VISITDY": 168,
  "AVISIT": "End of treatment",
  "AWLO": 141,
  "AWTARGET": 168,
  "AWHI": "INF"
},
{
  "Assessment": "NPI-X",
  "EPOCH": "Screening",
  "VISITNUM": 3,
  "VISIT": "BASELINE",
  "VISITDY": 1,
  "AVISIT": "Baseline",
  "AWLO": 1,
  "AWTARGET": 1,
  "AWHI": 2
},
}
```

1	Assessment	EPOCH	VISITNUM	VISIT	VISITDY	AVISIT	AWLO	AWTARGET	AWHI
2	ADAS-Cog	Screening	3	BASELINE	1	Baseline	1	1	2
3	ADAS-Cog	Treatment	8	WEEK 8	56	Week 8	2	56	84
4	ADAS-Cog	Treatment	10	WEEK 16	112	Week 16	85	112	140
5	ADAS-Cog	Treatment	12	WEEK 24	168	End of treatment	141	168	INF
6	NPI-X	Screening	3	BASELINE	1	Baseline	1	1	2
7	NPI-X	Treatment	3.5	AMBUL ECG PLACEMENT	13				
8	NPI-X	Treatment	4	WEEK 2	14	Week 2	2	14	21
9	NPI-X	Treatment	5	WEEK 4	28	Week 4	22	28	35
10	NPI-X	Treatment	7	WEEK 6	42	Week 6	36	42	49
11	NPI-X	Treatment	8	WEEK 8	56	Week 8	50	56	63
12	NPI-X	Treatment	8.1	WEEK 10 (T)	70	Week 10	64	70	77
13	NPI-X	Treatment	9	WEEK 12	84	Week 12	78	84	91
14	NPI-X	Treatment	9.1	WEEK 14 (T)	98	Week 14	92	98	105
15	NPI-X	Treatment	10	WEEK 16	112	Week 16	106	112	119
16	NPI-X	Treatment	10.1	WEEK 18 (T)	126	Week 18	120	126	133
17	NPI-X	Treatment	11	WEEK 20	140	Week 20	134	140	147
18	NPI-X	Treatment	11.1	WEEK 22 (T)	154	Week 22	148	154	161
19	NPI-X	Treatment	12	WEEK 24	168	End of treatment	162	168	175
20	NPI-X	Treatment	13	WEEK 26	182	End of study	176	182	INF
21									

Datasheet	Field	Core	Role	Reference	Type	Valid Values	Syntax
Windows	Assessment	Required	Key		text	ADAS-Cog, CIBIC+, NPI-X, Labs	
Windows	EPOCH	Expected			text		
Windows	VISITNUM	Required			text		(\d+(\.\d+)?) UNSCHEDULED
Windows	VISIT	Required	Key		text		
Windows	VISITDY	Required			text		(-?\d+) \.
Windows	AVISIT	Expected			text		
Windows	AWLO	Expected			text		(INF \d+)?
Windows	AWTARGET	Expected			integer		
Windows	AWHI	Expected			text		(INF \d+)?

PIPELINE DSL

```
/*
 * Define a pipeline that is used as a local standards
 * repository, for this demo, we import CDISC Library
 */
pipeline standards {
  /*
   * CDISC Library interface schema uses Swagger
   */
  resource CDISC_API_Schema {
    uri: 'CDISC1-cdisc-library-2.0-swagger.json'
    type: file
  }
  /*
   * The REST interface to CDISC Library which
   * conforms to the API interface defined above
   */
  resource CDISC_Library {
    uri: 'https://library.cdisc.org/api'
    conforms_to: CDISC_API_Schema
    type: rest
    credentials: "~/.arctica/keys"
  }
  /*
   * Import the CDISC library contents (all of it) from
   * the REST API into a local folder resource.
   */
  process import_cdisc {
    input: CDISC_Library
    output: CDISC
    template: 'import_rest_all.ps1.jinja2' // import everything!
  }
  /*
   * Local cache of everything in the CDISC Library
   */
  resource CDISC {
    extension: 'json'
    type: folder
  }
}
```

```
/*
 * SDTM pipeline (Data Store)
 */
pipeline sdtm {
  resource define {
    uri: "define.xml"
    type: file
  }
  resource datasets {
    type: folder
    extension: "xpt"
    conforms_to: define
  }
}
/*
 * ADAM pipeline that creates ADSL
 */
pipeline adam {
  /*
   * INPUTS
   */
  import sdtm.*
  import specification.adam
  /*
   * PROCESS
   */
  process adsl {
    input: sdtm.datasets
    output: adam.datasets[adsl]
    template: "adam/adsl.sas.jinja"
    conforms_to: specification.adam[adsl]
  }
  /*
   * OUTPUT
   */
  resource datasets {
    extension: "sas7bdat"
    type: folder
    conforms_to: standards.CDISC[adamig_1_3]
  }
}
```

Code Generation

The DOM and Templates

CODE GENERATION

DOM Tree (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_HEMA",
...   "ResultDisplayParentOID": null,
...   "Version": 1,
...   "GroupingDataset": "ADAM.ADSL",
...   "GroupingWhereVar": "SAFFL",
...   "GroupingWhereComparator": "EQ",
...   "GroupingWhereValue": "\"Y\"",
...   "GroupingAnalysisVar": "TRTP_DOSEPC",
...   "GroupingAnalysisVarOrdFmt": "TRTPN_DOSEP",
...   "GroupingByVar": null,
...   "GroupingByOrdFmt": null,
...   "DisplayTemplate": "template1",
...   "DisplayName": "Table 14.3.4.1",
...   "DisplayTitle": "Summary of Hematology Pa",
...   "Title1": "Dose Escalation",
...   "Title2": "#bvlne",

```

Template (Jinja2)

```

1  {#- Top-level SAS output program Jinja2 template -#}
2  /*****
3  *
4  * .. / \ .. | .. / \ .. | .. / \ .. | .. / \ .. | .. / \ ..
5  * .. / \ .. | .. / \ .. | .. / \ .. | .. / \ .. | .. / \ ..
6  * .. / \ .. | .. / \ .. | .. / \ .. | .. / \ .. | .. / \ ..
7  * .. / \ .. | .. / \ .. | .. / \ .. | .. / \ .. | .. / \ ..
8  *
9  * Study ..... : {{ Output.Study }}
10 * Analysis ..... : {{ Output.Analysis }}
11 * Program ..... : {{ Output.Filename }}.sas
12 *
13 * DESCRIPTION
14 *
15 * This program creates Display {{ Output.DisplayID }}
16 * .. {{ Output.Display.DisplayName }} .. {{ Output.Display
17 *
18 * Input files:
19 {#- comma separated list of result datasets #}
20 * .. {{ Output.Display.Results|map(attribute="Dataset") }}
21 *
22 * Output file:
23 * .. {{ Output.Filename }}.{{ Output.Type }}
24 *
25 * Macros:
26 {%- block comment_macros %}
27 * .. u_autoexec
28 {%- endblock %}
29 *

```

Program (SAS)

```

1
2 /*****
3 *
4 * ..- - - - -
5 * ..- - - - -
6 * ..- - - - -
7 * ..- - - - -
8 *
9 * Study ..- - - - - : CDISCPIL0T01
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 *

```

DOM Tree (JSON)

```
{
  "Study": "CDISCPilot01",
  "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"
      }
    ]
  }
}
```

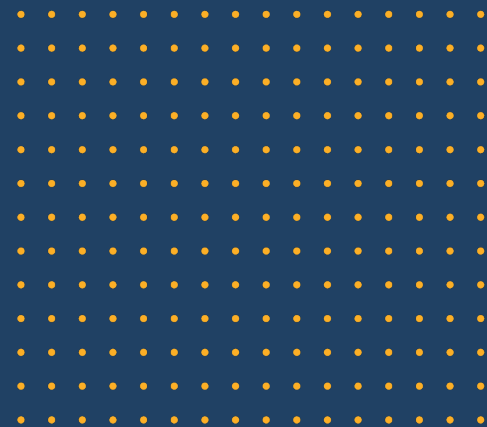
Template (Jinja2)

```
49 # =====
50 # Tasks to build individual TFL
51 # =====
52
53 {% for Output in Outputs %}
54 # Output: {{ Output.DisplayID }}
55 # {{ Output.Display.DisplayName }}-{{ Output.Display.DisplayTitle }}
56 # =====
57
58 Task-{{ Output.Filename }} @{{
59   Inputs = @(
60     "$DataPath\{{ Output.Display.GroupingDataset|replace('ADAM.', '') }}.sas7bdat",
61     {{%- for dataset in Output.Display.Results|map(attribute='Dataset') %}}
62     "$DataPath\{{ dataset|replace('ADAM.', '') }}.sas7bdat",
63     {{%- endfor %}}
64   )
65   {{ Output.Program }}.sas"
66   Outputs = "..\output\q_{{ Output.Filename }}.{{ Output.Type }}"
67   Jobs = {
68     $r = viper-run-sas {{ Output.Program }}
69     $script:results += $r
70     write-host $r
71   }
72 }
73 {% endfor %}
74
75 # =====
76 # Task that can be used to build all the TFL datasets..
77 # ..or specific ones (using the TFLID)
78 # =====
79 Task TFL-{{%- for Output in Outputs %}}-{{ Output.Filename }}-{{
80   {{%- if not loop.last %}},{{%- endif %}}
81   {{%- endfor %}}
82 }
83
84 # =====
85 # clean all (TFL) outputs
86 # =====
87 Task Clean-{{
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 # Tasks to build individual TFL
51 # =====
52
53
54 # Output: T14030401_SAF_HEMA
55 # Table 14.3.4.1 Summary of Hematology Parameters
56 # =====
57
58 Task t14_03_04_01_hema @{{
59   Inputs = @(
60     "$DataPath\ADSL.sas7bdat",
61     "$DataPath\ADLB.sas7bdat",
62     ..
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     ..
83     "q_t_chem.sas"
84   )
85   Outputs = "..\output\q_t14_03_04_07_chem.rtf"
86   Jobs = {
87     $r = viper-run-sas q_t_chem
88     $script:results += $r
89     write-host $r
90   }
91 }
92
93 # =====
94 # Output: T14030413_COAG
```

Summary



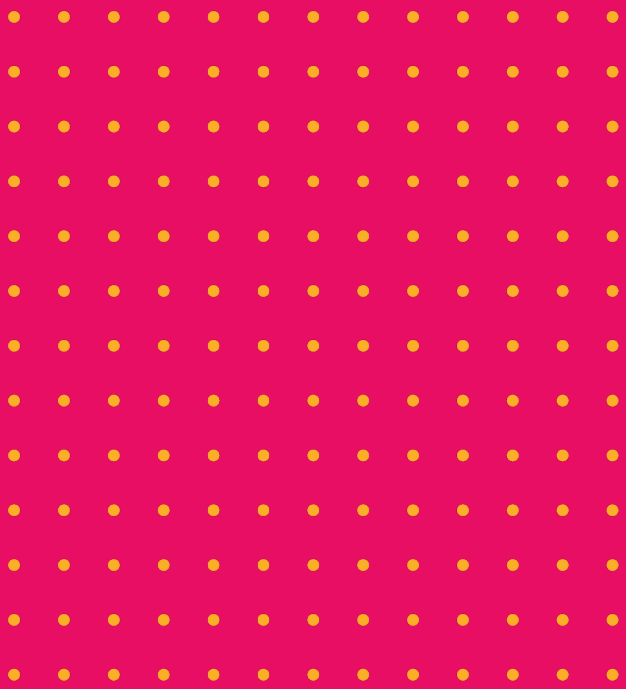
Metadata as a first-class citizen

Automation component architecture

Methodology based on traceability and validation

Build internal metadata as DOM Tree

Code generation using templates



Thank you!

Find us at **BOOTH 26**

Demos:

- Data Visualisation
 - Automation
 - Cloud SCE
 - RWE

Contact me:

Stuart.malcolm@veramed.co.uk

www.linkedin.com/in/stuart-malcolm

