



2023
EUROPE
INTERCHANGE
COPENHAGEN | 26-27 APRIL



A Language-Orientated Approach to CDISC Metadata-Driven Automation

Stuart Malcolm,
Head of Standards, Efficiency and Automation,
Veramed



Meet the Speaker

Stuart Malcolm

Title: Head of Standards, Efficiency and Automation

Organization: Veramed

Stuart Malcolm is Head of Standards, Efficiency and Automation at Veramed where he is responsible for the delivery of software platforms, tools, and techniques to optimise the delivery of clinical trial analysis projects

Stuart has over 25 years experience delivering software-based solutions in clinical trials, and previously telecoms, finance and media.

Stuart blogs intermittently at www.metadatadriven.com

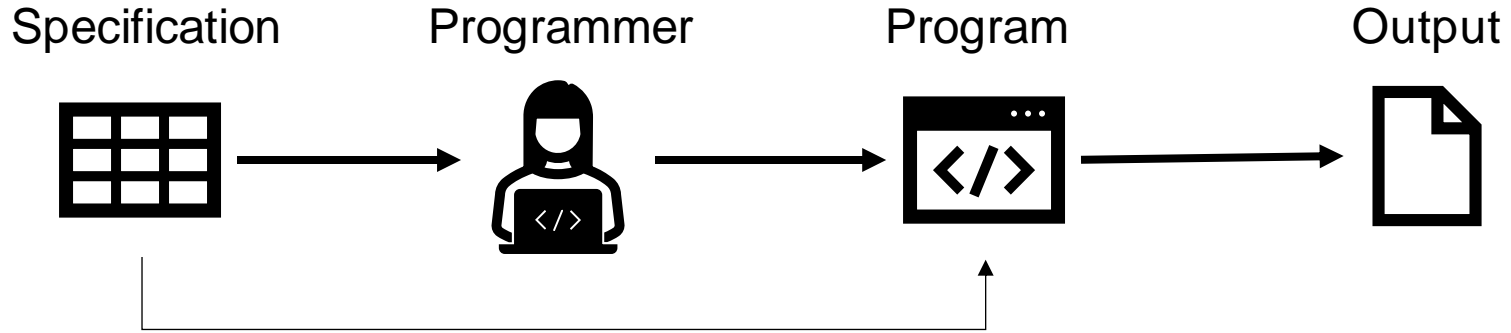
Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.*
- *{Please disclose any financial relationship or conflict of interest relevant to this presentation here OR}*
- *The author(s) have no real or apparent conflicts of interest to report.*

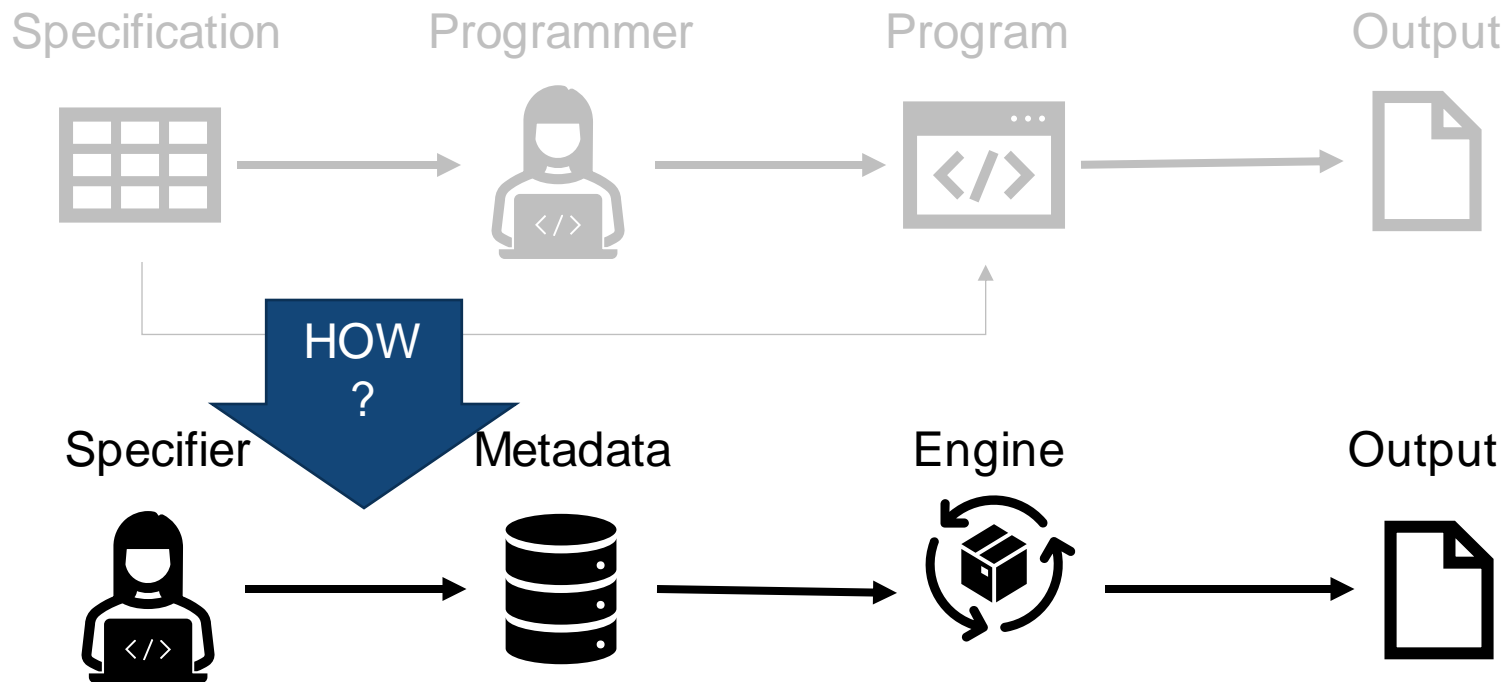


What is the problem with metadata-driven automation?

Achieving the vision of end-to-end automation

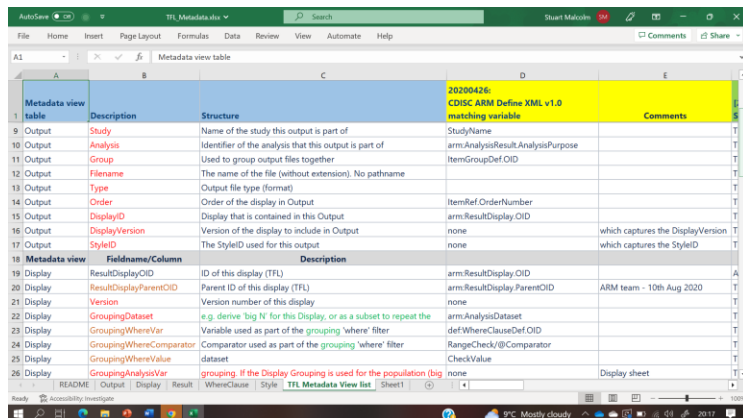


Achieving the vision of end-to-end automation



These are NOT efficient metadata user interfaces...

Excel

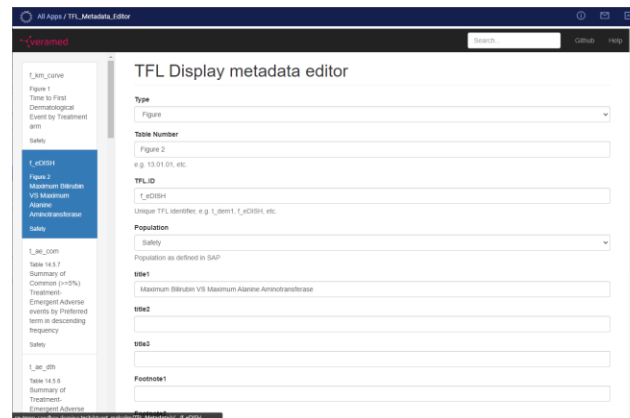


The screenshot shows an Excel spreadsheet titled 'TFL_Metadata.xlsx'. The main table has the following data:

	Description	Structure	CDISC ARM Define XML v1.0 matching variable	Comments
9	Output Study	Name of the study this output is part of	StudyName	
10	Output Analysis	Identifier of the analysis that this output is part of	arm:AnalysisResult:AnalysisPurpose	
11	Output Group	Used to group output files together	ItemGroupDef:OID	
12	Output Filename	The name of the file (without extension). No pathname		
13	Output Type	Output file type (format)		
14	Output Order	Order of the display in Output	ItemRef:OrderNumber	
15	Output DisplayID	Display that is contained in this Output	arm:ResultDisplay:OID	
16	Output DisplayVersion	Version of the display to include in Output	none	which captures the DisplayVersion
17	Output StyleID	The StyleID used for this output	none	which captures the StyleID

Metadata view	Fieldname/Column	Description	
19	Display ResultDisplayOID	ID of this display (TFL)	arm:ResultDisplay:OID
20	Display ResultDisplayParentOID	Parent ID of this display (TFL)	arm:ResultDisplay:ParentOID
21	Display Version	Version number of this display	none
22	Display GroupingDataset	e.g. derive 'big N' for this Display, or as a subset to repeat the	arm:AnalysisDataset
23	Display GroupingWhereVar	Variable used as part of the grouping 'where' filter	def:WhereClauseDef:OID
24	Display GroupingWhereComparator	Comparator used as part of the grouping 'where' filter	RangeCheck/@Comparator
25	Display GroupingWhereValue	dataset	CheckValue
26	Display GroupingAnalysisVar	grouping: If the Display Grouping is used for the population (big N)	none

Database Form

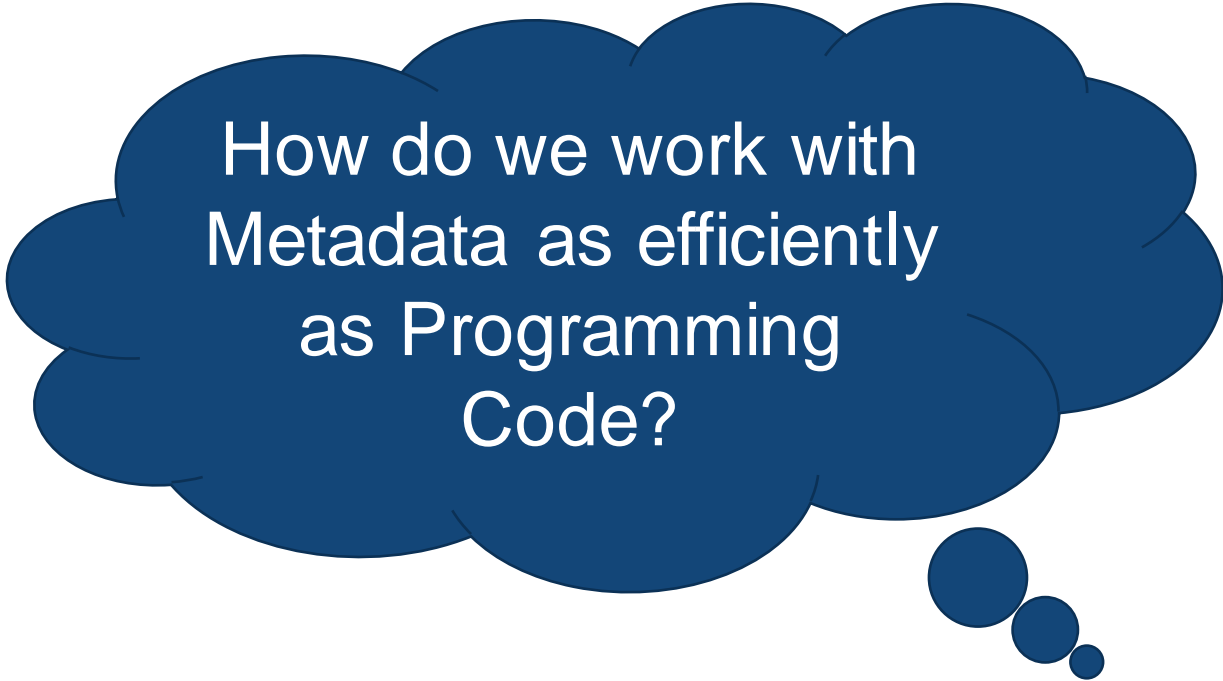


The screenshot shows a web-based form titled 'TFL Display metadata editor'. It contains several input fields and dropdown menus for entering metadata. The form is organized into sections, with a left sidebar showing a tree view of the data model.

Fields include:

- Type: Figure
- Table Number: Figure 2
- TFL_ID: f_0018H
- Population: Safety
- title1: Maximum Blebsin VS Maximum Abiraterone Ammoniochloride
- title2: Population as defined in SAP
- title3: Maximum Blebsin VS Maximum Abiraterone Ammoniochloride
- Footnote1: Summary of Common (≥10%) Treatment-Related Adverse Events by Preferred Term in Descending Frequency

Problem statement: “Metadata as a first-class citizen”

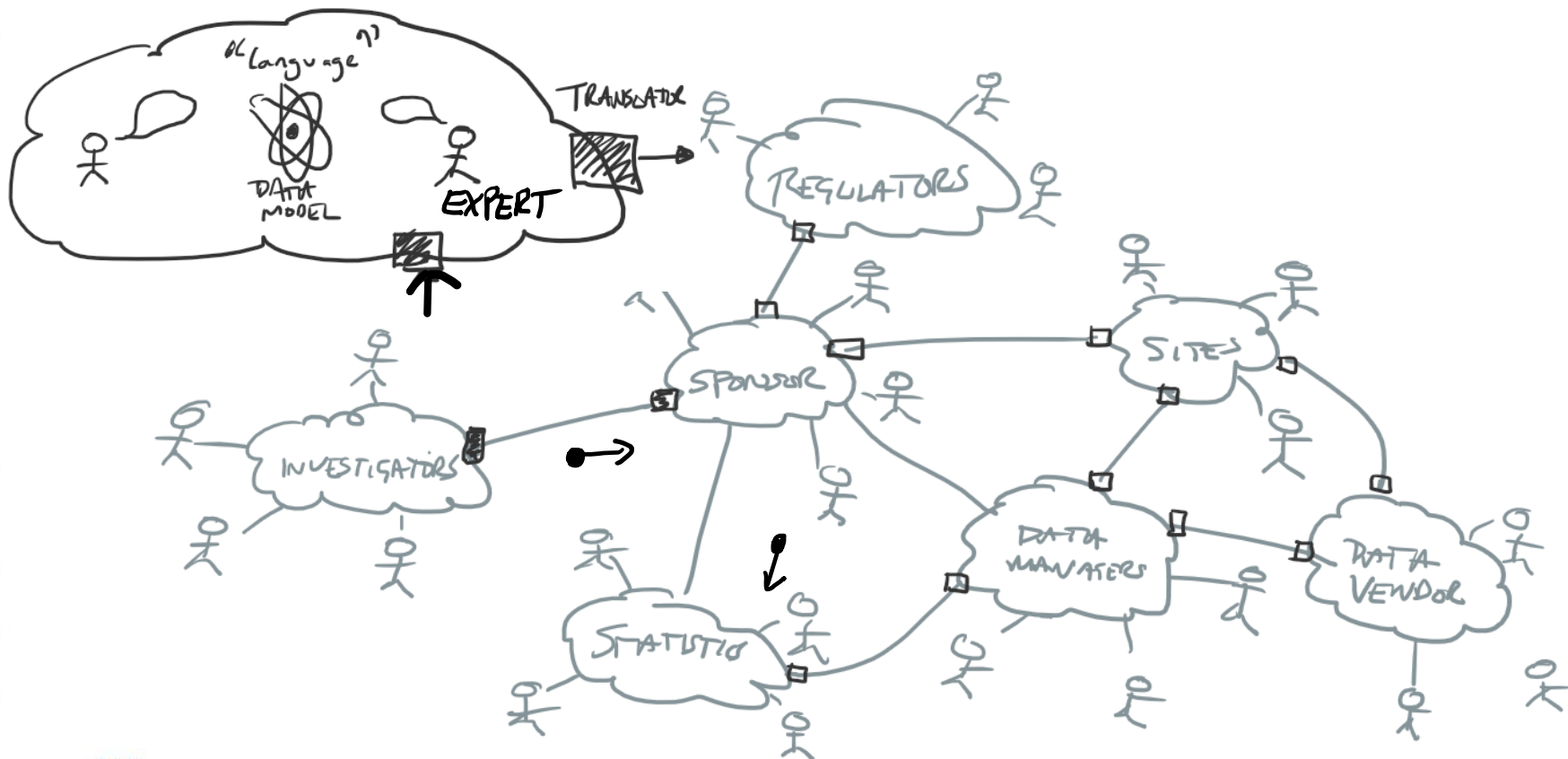


How do we work with
Metadata as efficiently
as Programming
Code?



Language-orientated programming

The language of clinical trials



Eclipse Modelling Framework (EMF)

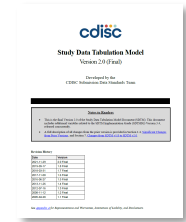
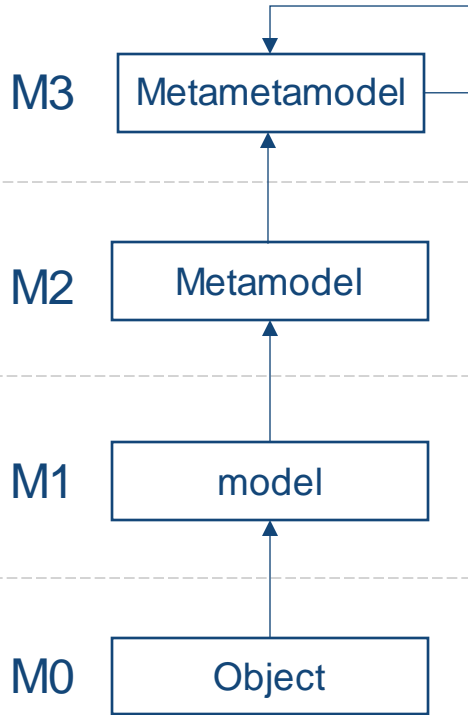
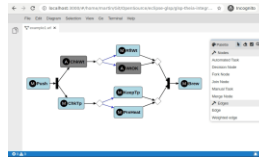


Table with 10 columns: Study ID, Start Date, End Date, Study Name, Study Type, Study Status, Study Location, Study Country, Study Site, Study Site ID.

Study ID	Start Date	End Date	Study Name	Study Type	Study Status	Study Location	Study Country	Study Site	Study Site ID
1001	2010-01-01	2010-12-31	Study 1001	Phase 1	Completed	USA	USA	Site 1001	1001
1002	2011-01-01	2011-12-31	Study 1002	Phase 2	In Progress	USA	USA	Site 1002	1002
1003	2012-01-01	2012-12-31	Study 1003	Phase 3	Completed	USA	USA	Site 1003	1003
1004	2013-01-01	2013-12-31	Study 1004	Phase 4	Completed	USA	USA	Site 1004	1004
1005	2014-01-01	2014-12-31	Study 1005	Phase 5	Completed	USA	USA	Site 1005	1005

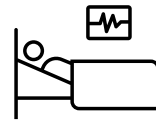


```

classDiagram
    class Person {
        -id: Integer
        -name: String
        -birthdate: Date
        -salary: Integer
        +ref employees: Person
    }
    class Employee {
        -id: Integer
        -name: String
        -birthdate: Date
        -salary: Integer
    }
    class Patient {
        -id: Integer
        -name: String
        -birthdate: Date
        -salary: Integer
    }
    Person "1" -- "*" Employee
    Person "1" -- "*" Patient
  
```

Table with 10 columns: STUDYID, STARTDATE, ENDDATE, STUDYNAME, STUDYTYPE, STUDYSTATUS, STUDYLOCATION, STUDYCOUNTRY, STUDYSITE, STUDYSITEID.

STUDYID	STARTDATE	ENDDATE	STUDYNAME	STUDYTYPE	STUDYSTATUS	STUDYLOCATION	STUDYCOUNTRY	STUDYSITE	STUDYSITEID
1001	2010-01-01	2010-12-31	Study 1001	Phase 1	Completed	USA	USA	Site 1001	1001
1002	2011-01-01	2011-12-31	Study 1002	Phase 2	In Progress	USA	USA	Site 1002	1002
1003	2012-01-01	2012-12-31	Study 1003	Phase 3	Completed	USA	USA	Site 1003	1003
1004	2013-01-01	2013-12-31	Study 1004	Phase 4	Completed	USA	USA	Site 1004	1004
1005	2014-01-01	2014-12-31	Study 1005	Phase 5	Completed	USA	USA	Site 1005	1005



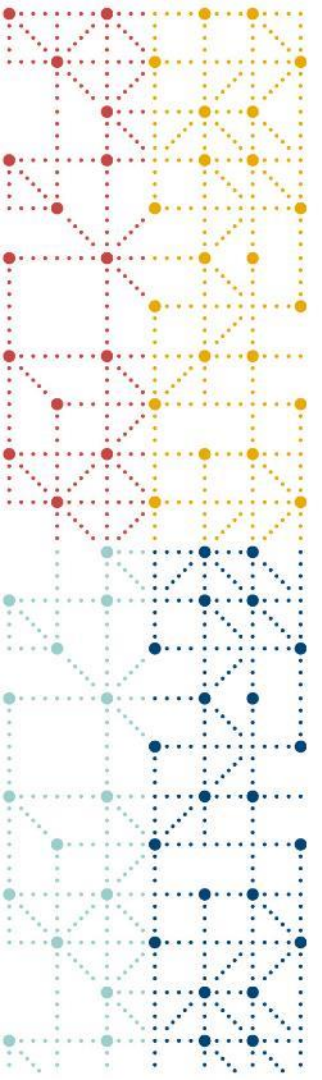
Events



Interventions



Findings



Example



Study Data Tabulation Model

Version 2.0 (Final)

Developed by the
CDISC Submission Data Standards Team

Notes to Readers

- This is the final Version 2.0 of the Study Data Tabulation Model Document (SDTM). This document includes additional variables related to the SDTM Implementation Guide (SDTMIG) Version 3.4, released concurrently.
- A full description of all changes from the prior version is provided in Section 1.4, [Significant Changes from Prior Versions](#), and Section 7, [Changes from SDTM v1.8 to SDTM v2.0](#).

Revision History

Date	Version
2021-11-29	2.0 Final
2019-09-17	1.8 Final
2018-03-31	1.7 Final
2017-11-08	1.6 Final
2016-06-27	1.5 Final
2013-11-26	1.4 Final
2012-07-16	1.3 Final
2008-11-12	1.2 Final
2005-04-28	1.1 Final

See [Appendix A](#) for Representations and Warranties, Limitations of Liability, and Disclaimers.

CONTENTS

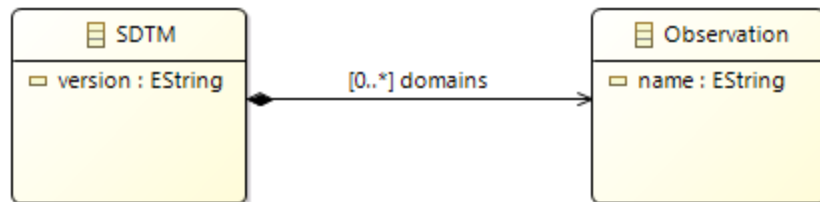
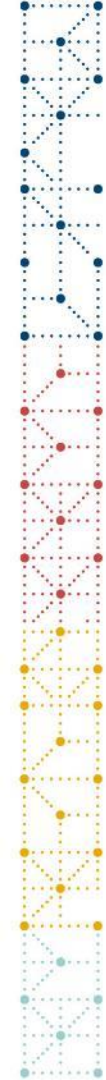
1	INTRODUCTION	4
1.1	PURPOSE	4
1.2	IMPLEMENTATION ADVICE FOR THIS MODEL	4
1.3	RELATIONSHIP TO PRIOR CDISC MODELS	5
1.4	SIGNIFICANT CHANGES FROM PRIOR VERSIONS	5
2	MODEL CONCEPTS AND TERMS – ORGANIZATION OF THE SDTM	8
2.1	MODEL CONCEPTS AND TERMS – VARIABLES	9
2.2	TABLE STRUCTURE	10
3	STUDY SUBJECT DATA	11
3.1	THE GENERAL OBSERVATION CLASSES	11
3.1.1	The Interventions Observation Class	12
3.1.2	The Events Observation Class	15
3.1.3	The Findings Observation Class	20
3.1.3.1	Findings About Events or Interventions	32
3.1.4	Identifiers for All Classes	32
3.1.5	Timing Variables for All Classes	34
3.2	SPECIAL-PURPOSE DOMAINS	40
3.2.1	Demographics	40
3.2.2	Comments	44
3.2.3	Subject Summary Domains	45
3.2.3.1	Subject Elements	45
3.2.3.2	Subject Repro Stages	46
3.2.3.3	Subject Visits	47
3.2.3.4	Subject Disease Milestones	49
4	ASSOCIATED PERSONS DATA	50
5	STUDY-LEVEL DATA	51
5.1	THE TRIAL DESIGN MODEL	51
5.1.1	Trial Arms and Trial Elements	51
5.1.1.1	Trial Elements	51
5.1.1.2	Trial Arms	52
5.1.2	Trial Sets	53
5.1.3	Trial Repro Stages and Trial Repro Paths	54
5.1.3.1	Trial Repro Stages	54
5.1.3.2	Trial Repro Paths	54
5.1.4	Trial Planned Data Collection	55
5.1.4.1	Trial Visits	56
5.1.4.2	Trial Disease Assessments	56
5.1.4.3	Trial Disease Milestones	57
5.1.5	Trial Inclusion/Exclusion Criteria	58
5.1.6	Trial Summary Information	59
5.1.7	Challenge Agent Characterization	60
5.2	STUDY REFERENCES	61
5.2.1	Device Identifiers Dataset	61
5.2.2	Non-host Organism Identifiers Dataset	62



2 Model Concepts and Terms – Organization of the SDTM

The **SDTM** provides a general framework for describing the organization of information collected during human and animal studies. The model is built around the concept of **observations**, which consist of discrete pieces of information collected during a study. Observations normally correspond to rows in a dataset. A **domain** is a collection of observations on a particular topic (see Concept Map, below). For example, "Subject 101 had an adverse event of mild nausea starting on study day 6" is an observation belonging to the Adverse Events domain in a clinical trial.

The primary purpose of the SDTM is to represent data about study subjects—which may be humans or animals—or medical devices. The SDTM includes a general model for representing data in 3 "general observation" classes. Within those classes, data are grouped by topic into domains, represented in separate datasets. The 3 general observation classes (i.e., Interventions, Events, Findings) are described further in Section 3.1, [The General Observation Classes](#). Additional special-purpose datasets about individuals are described in Section 3.2, [Special-purpose Domains](#).



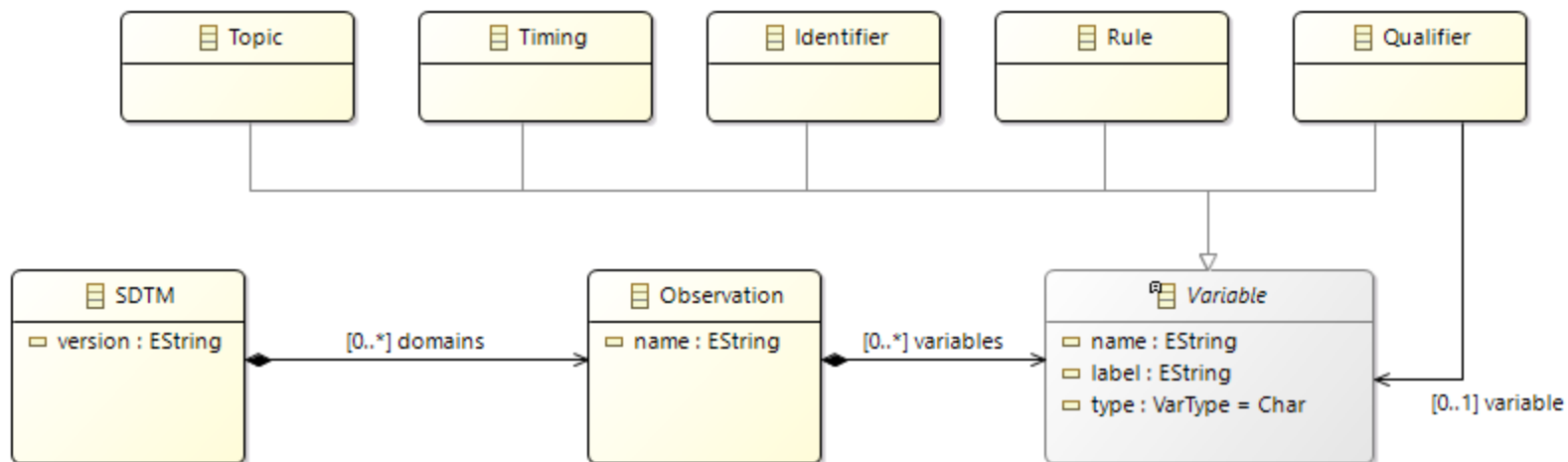
2.1 Model Concepts and Terms – Variables

All datasets are structured as flat files with rows representing observations and columns representing variables; each dataset is described by metadata definitions that provide information about the variables used in the dataset.

Metadata are described in the CDISC Define-XML specification, available at <https://www.cdisc.org/standards/data-exchange/define-xml>.

Each observation consists of a series of named **variables**. Each variable, which normally corresponds to a column in a dataset, can be classified according to its role. A *role* describes the type of information conveyed by the variable about each distinct observation and how it can be used. There are variables which play different roles in different datasets. This is most common for variables which appear in both trial design datasets and general observation class datasets. For example, ARMCD is the topic variable in Trial Arms (TA), but a record qualifier in Demographics (DM) and Trial Visits (TV). Variables which appear in multiple general observation classes have the same role, although the variable qualified by a variable qualifier or synonym qualifier can be different in different general observation classes. For example, --MODIFY qualifies --TRT in interventions, --TERM in events, and --ORRES in findings.

SDTM variables can be classified into **5 major roles**:



eclipse-workspace - platform:/resource/demo2.sdtm.model/model/sdtm.aird/sdtm - Papyrus

File Edit Diagram Navigate Search Project Run Window Help

Model Explorer

type filter text

- > arctica.model.dataset.model
- > arctica.model.dataset.model.edit
- > arctica.model.dataset.model.editor
- > arctica.workbench.pipeline.dsl
- > arctica.workbench.pipeline.dsl.ide
- > arctica.workbench.pipeline.dsl.tests
- > arctica.workbench.pipeline.dsl.ui
- > arctica.workbench.pipeline.dsl.ui.tests
- > arctica.model.dataset.dsl
- > arctica.model.dataset.dsl.ide
- > arctica.model.dataset.dsl.tests
- > arctica.model.dataset.dsl.ui
- > arctica.model.dataset.dsl.ui.tests
- > demo1.sdtm.model
- > demo1.sdtm.model.edit
- > demo1.sdtm.model.editor
- > demo2.sdtm.model
 - Project Dependencies
 - src [in demo2.sdtm.model]
 - JRE System Library [JavaSE-17]
 - Plug-in Dependencies
 - META-INF
 - model
 - sdtm.aird
 - sdtm.ecore
 - sdtm
 - sdtm
 - SDTM
 - Observation
 - Variable
 - VarType
 - Topic -> Variable
 - Qualifier -> Variable
 - Identifier -> Variable
 - Rule -> Variable
 - Timing -> Variable

sdTM.genmodel

Properties

sdtm

Ecore

Documentation

Name: sdtm

Annotation

Ns URI: http://www.example.org/sdtm

Generation

Ns Prefix: sdtm

Rulers & Grid

Diagram

```

classDiagram
    class VarType {
        - Char
        - Num
    }
    class SDTM {
        version : EString
    }
    class Observation {
        name : EString
    }
    class Variable {
        name : EString
        label : EString
        type : VarType = Char
    }
    class Topic
    class Timing
    class Identifier
    class Rule
    class Qualifier

    SDTM --> Observation : [0..*] domains
    Observation --> Variable : [0..*] variables
    Variable --> VarType : [0..1] variable
    Topic --> Variable
    Timing --> Variable
    Identifier --> Variable
    Rule --> Variable
    Qualifier --> Variable
  
```

Palette

- Existing Elements
- Add
- Remove
- Classifier
- Class
- Datatype
- Enumeration
- ETypeParameter
- Feature
- Literal
- Operation
- Attribute
- Relation
- SuperType
- Reference
- Bi-directional Reference
- Composition
- Dynamic
- Dynamic instance
- Package

Navigate

Edit

Refresh F5

Unsynchronized

Export diagram as image

Show/Hide

Select

Layout

Reset Origin Ctrl+Home

Validate diagram

Find Ctrl+Alt+Shift+F

Quick search Ctrl+O

Profiles

Show EClass information

Show References

OCLE

Generate

- All
- Model Code
- Edit Code
- Editor Code

Remove from Context Ctrl+Alt+Shift+Down

Problems

Git Repositories

Git Staging

Console

Templates

No consoles to display at this time.

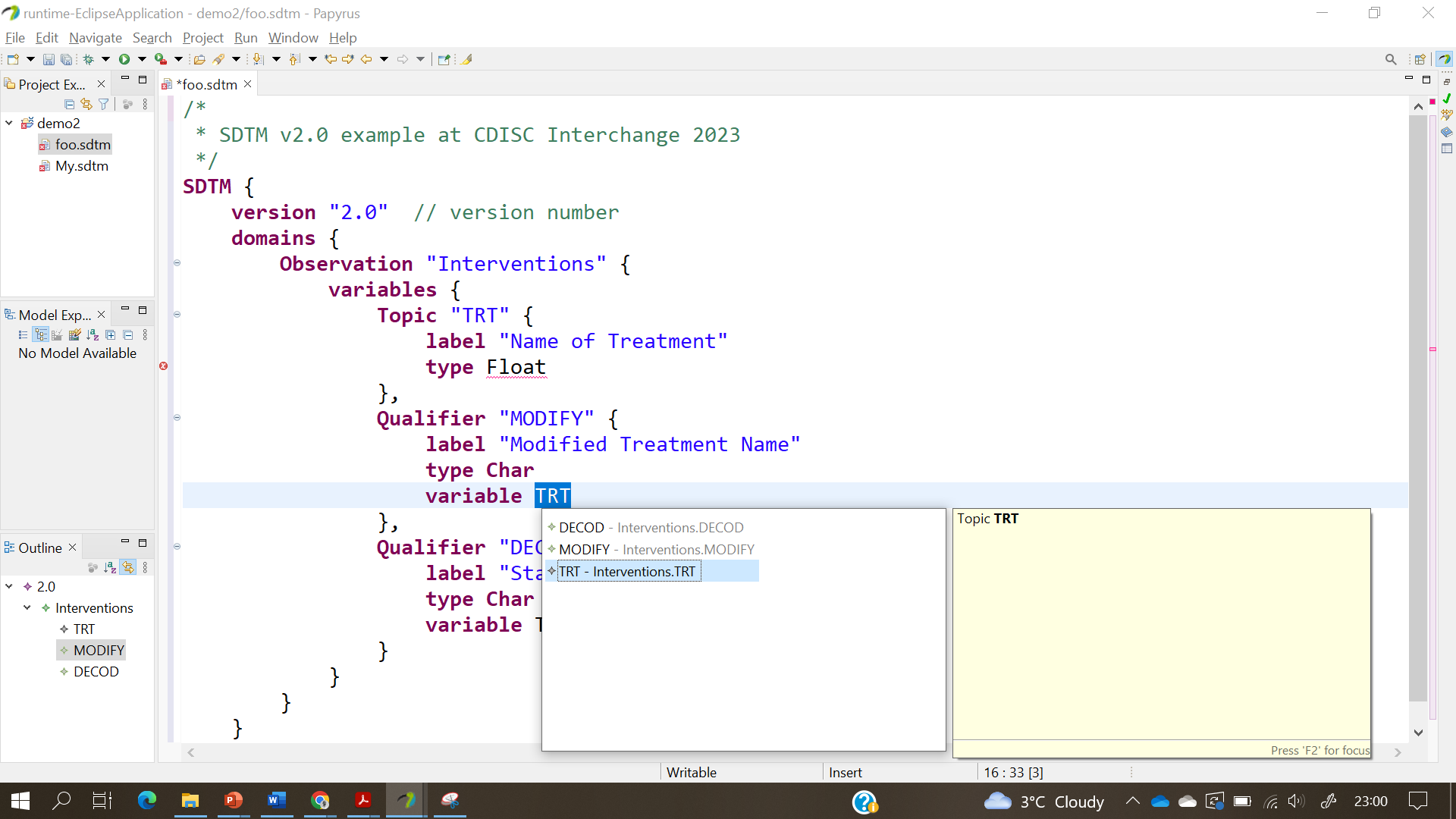
Synchronized diagram

The screenshot displays two windows from a CDISC software interface:

- Resource Set Window:** Shows a hierarchical tree structure. The path is:
 - platform:/resource/demo2/My.sdtm
 - SDTM 2.0
 - Observation Interventions
 - Topic TRT
 - Qualifier MODIFY
 - Qualifier DECOD** (highlighted)
 - Qualifier MOOD
 - Qualifier CAT
 - Qualifier SCAT
 - Observation Events
 - Observation Findings

- Properties Window:** Shows a table of properties for the selected 'Qualifier DECOD' element.

Property	Value
Label	Standardized Treatment Name
Name	DECOD
Type	Char
Variable	Topic TRT (selected in dropdown) <ul style="list-style-type: none"> Qualifier CAT Qualifier DECOD Qualifier MODIFY Qualifier MOOD Qualifier SCAT Topic TRT (highlighted)



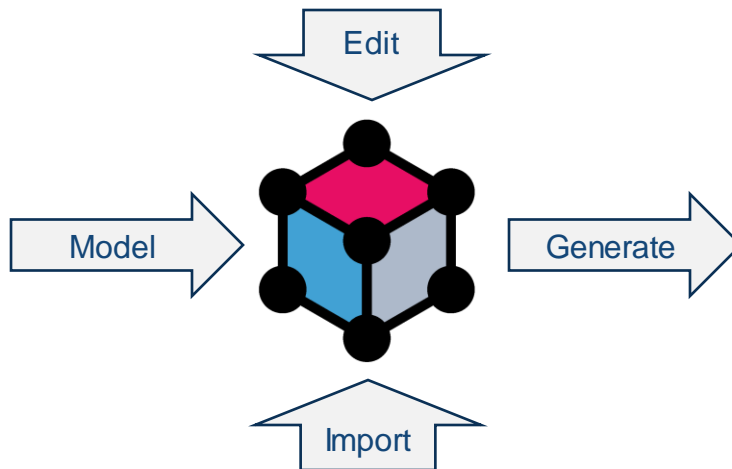
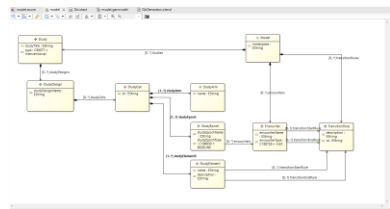
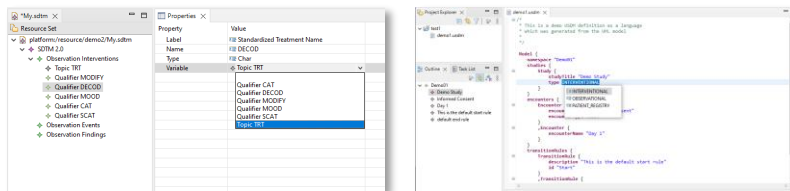
```
/*
 * SDTM v2.0 example at CDISC Interchange 2023
 */
SDTM {
  version "2.0" // version number
  domains {
    Observation "Interventions" {
      variables {
        Topic "TRT" {
          label "Name of Treatment"
          type Float
        },
        Qualifier "MODIFY" {
          label "Modified Treatment Name"
          type Char
          variable TRT
        },
        Qualifier "DECOD" {
          label "Standardized Treatment Name"
          type Char
          variable TRT
        }
      }
    }
  }
}
```

- ◆ DECOD - Interventions.DECOD
- ◆ MODIFY - Interventions.MODIFY
- ◆ TRT - Interventions.TRT

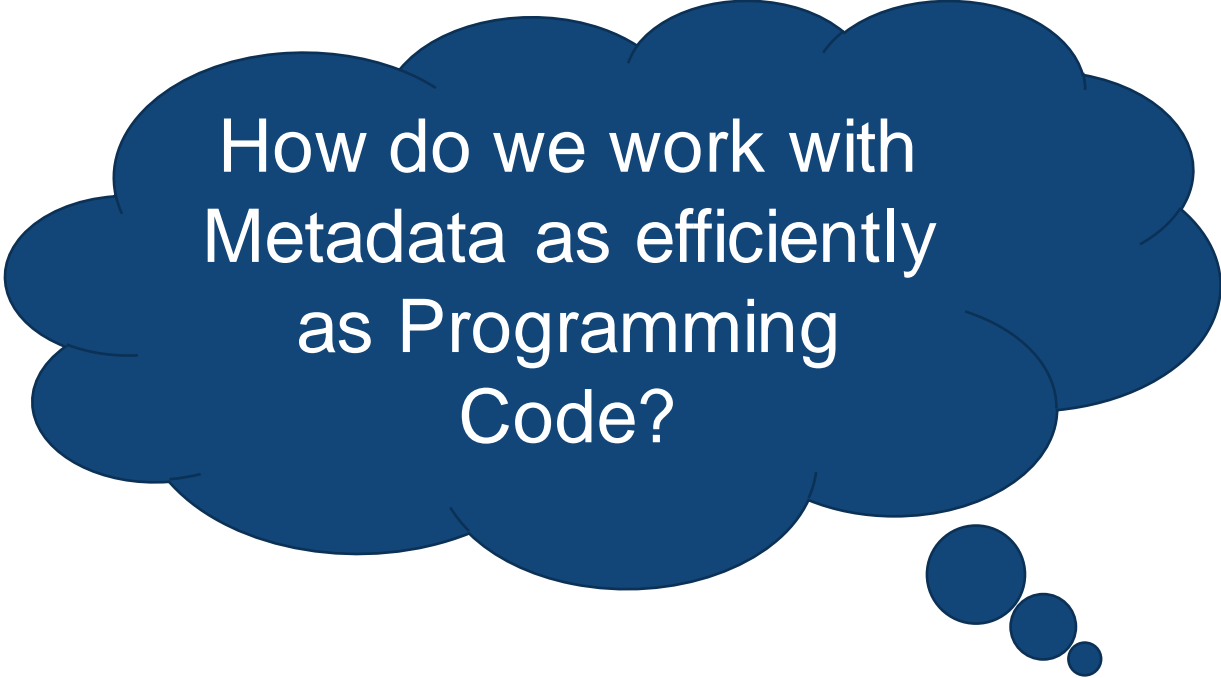
Topic TRT

Press 'F2' for focus

ARCTICA



Problem statement: “Metadata as a first-class citizen”



How do we work with
Metadata as efficiently
as Programming
Code?



Thank You!

