



Innovative Storage of SDTM Mappings as JSON

Boosting Efficiency and Data Consistency



Rohan Parmar, Roche, Mississauga, Canada
Edgar Manukyan, Roche, Mississauga, Canada
Rammprasad Ganapathy, Genentech, San Francisco, USA

Oak garden development team

rohan.parmar@roche.com

Table of contents

Table of contents

- | | |
|----------|-----------------------------|
| 1 | CDISC SDTM |
| 2 | Current Challenges |
| 3 | Case for JSON |
| 4 | Implementation & Challenges |
| 5 | Efficiency & Consistency |

What is SDTM Model?



Class - Subject-level observations are represented according to one of the three SDTM general observation classes. Class is a collection of Domains.

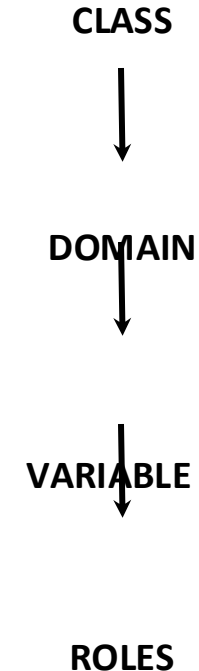
- Interventions
- Events
- Findings

Domain - Collection of logically related observations with a common topic. Eg - Demographics (DM), Concomitant Medications (CM), Substance Use (SU), Adverse Events (AE)

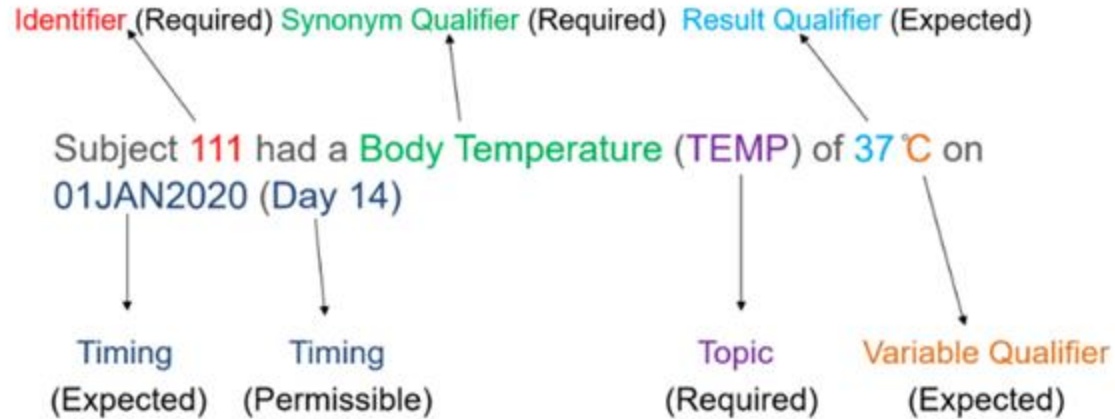
Variables - Each observation can be described by a series of variables, which are classified into the following roles:

- Identifier
- Topic
- Timing
- Qualifier

Hierarchy



SDTM Mapping



DOMAIN	USUBJID	VSTESTCD	VSTEST	VSORRES	VSORRESU	VSDTC	VSDY
VS	111	TEMP	TEMPERATURE	37	C	2020-01-01	14

Current Challenges



Traditional Approach To SDTM Mappings:

- SDTM mappings have been documented in formats like spreadsheets and PDF documents. They are not machine-readable, making it difficult to automate processes.
- Often to meet the Clinical Protocol Requirements, Clinical studies need to collect data that is not defined in MDR. This non-standard data is completely lost between studies.

Magnitude of Non-Standards:

- On an average 20% Non-Standard data is collected per study.
- SDTM mapping of such Non-standard data is done manually by Data standards team, study teams/SMEs across these studies.

Leading to Inconsistency

- Often similar SDTM mappings are interpreted in different ways by study teams and that lead to mapping similar data in different ways.
- Because of the inconsistent mapping, the SDTM domains are not aligned within a molecule, TA. It becomes difficult to pool the data across studies.

Time consuming Process

- Between clinical trial data collection, analysis, and submission to health authorities, a considerable amount of time is spent in mapping the Non-standard clinical trial data from its raw collected format to these standard data models (SDTM).

Case For JSON SDTM Mappings



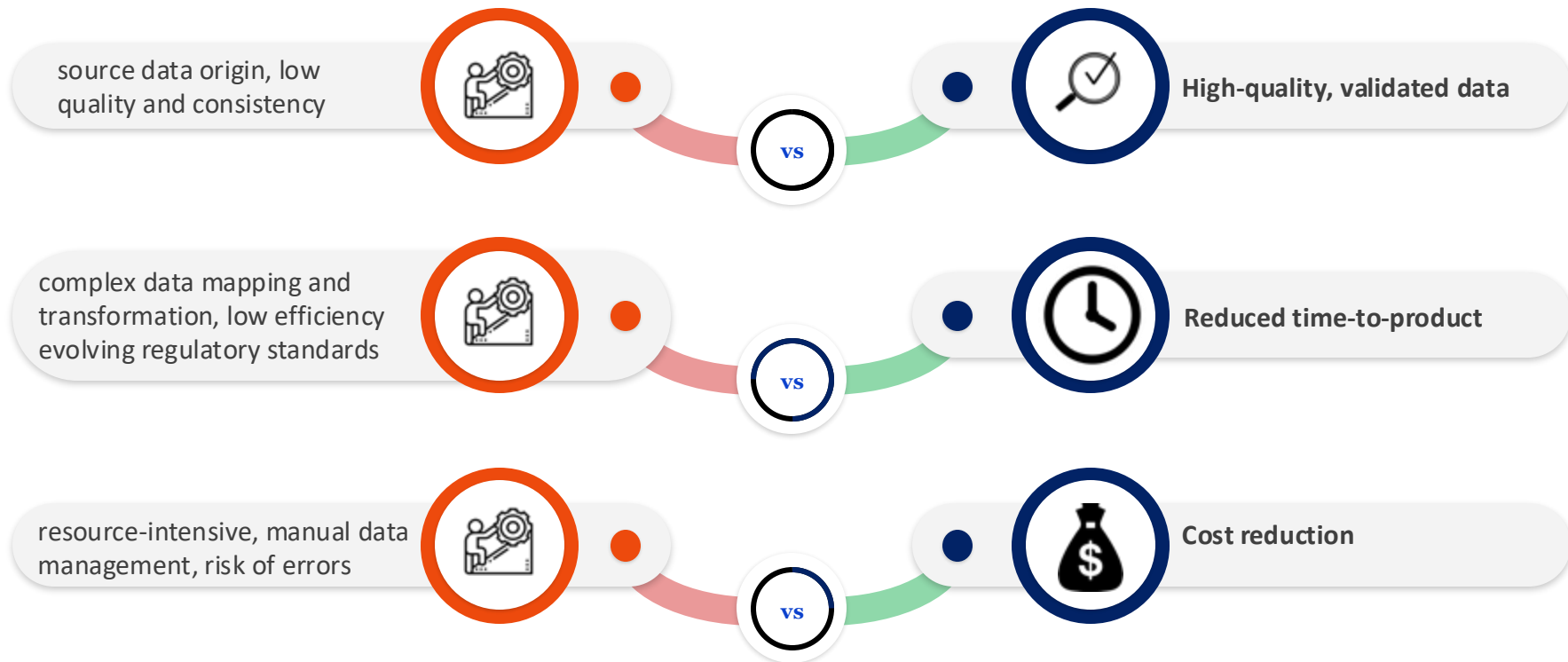
Simplicity and Flexibility

- Syntax is easy to read and write for humans while remaining machine-readable. This dual compatibility allows for seamless integration with a variety of programming languages and tools.
- JSON's machine-readable nature facilitates automation, allowing organizations to streamline workflows and reduce manual effort.

Advantages of JSON mappings

- Hierarchical structure supports the organization of mappings in a logical and intuitive way, improving accessibility and reusability. JSON mappings can be easily queried and updated, further enhancing their utility in dynamic data environments

Machine-readable SDTM mappings



Implementation and Challenges



Reusable SDTM Mappings and JSON Schema



CORE CONCEPT: REUSABLE ALGORITHMS

- SDTM Mappings are defined as algorithms that transforms the collected (CRF, non CRF) source data into the target Tabulation data model. Mapping Algorithms are the syntax of SDTM mappings and drive consistency and reusability of SDTM mappings.

We have designed 22 unique mapping Algorithms to accommodate most (80%) of the TA standards & Non-CRF data models

Key Points:

- Mappings are stored and modified as JSON
- Algorithms can be re-used across Domains
- Algorithms can be pre-specified for Standards
- Users can reuse/add algorithms for non-standards or new data types
 - **Both Standards and Non-Standards can therefore be supported**

22 Unique Algorithms

An abbreviated JSON SDTM mapping example



```
{  
  raw_dataset: 'ACTRAGM',  
  raw_dataset_label: 'Actual Cohort and Treatment Arm',  
  raw_variable: 'COHRT1',  
  raw_variable_label: 'Previous Cohort',  
  raw_variable_type: 'CheckBox',  
  annotation_text: 'SC.SCORRES when SC.SCTESTCD = \'PCOHORT\'',  
  annotation_ordinal: 1,  
  target_sdtm_variable_domain: 'SC',  
  target_sdtm_variable: 'SCORRES',  
  target_sdtm_variable_role: 'Result Qualifier',  
  mapping_algorithm: '01_ASSIGN_NO_CT',  
  @id: 'e4acbfe6-a3c3-4bd2-a857-024990c58faa'  
}
```

Automating code with JSON



```
sc_actragm <-  
  # Map topic variable  
  assign_ct(  
    raw_dat = "ACTRAGM",  
    raw_var = "COHRT",  
    tgt_var = "SCTESTCD",  
    ct_spec = ct_spec,  
    ct_clst = "C74559"  
  ) |>  
  # Map Synonym Qualifier  
  assign_ct(  
    raw_dat = "ACTRAGM",  
    raw_var = "COHRT",  
    tgt_var = "SCTEST",  
    ct_spec = ct_spec,  
    ct_clst = "C103330"  
  ) |>  
  assign_no_ct(  
    tgt_var = "SCORRES",  
    raw_dat = actragm,  
    raw_var = "COHRT1"  
  )
```

JSON SDTM mapping drives automation



SDTM Spec creator automates creation of :

- ❖ Study SDTMv spec
- ❖ Study Controlled Terminology
- ❖ Study VLM
- ❖ SDTM Define.xml



Automates creation of:

- ❖ SDTM(v) datasets
- ❖ SDTM(v) to SDTM
- ❖ XPT file creation



Move towards complete automation



Considering both standards and non-standards,
OAK is already automating 99% of the mappings.

In subsequent releases, we are close to eliminating any programming effort for SDTM(v) creation.

Our Future State
Could Be



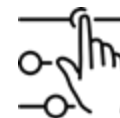
Complete the
SDTM(v) specs



QC the SDTM(v)
output
(99% automated)



Review log
files



Address any automation failures,
program domains (TDDs, RELREC),
variables (ex. ARM) that cannot be
automated



Challenges to using JSON SDTM mappings



Flexibility

- The biggest impact of the flexibility afforded by JSON is that applications need to account for the additional properties that may have been introduced for some mappings.
- Mappings can grow large and complex and may lead to search functionality growing in complexity.

Validation

- In order to ensure that all mappings will work when creating SDTMs there is frequent testing involved. Strict type checking for schema is necessary to ensure mappings are functional.

```
{
  raw_dataset: 'ACTRAGM',
  raw_dataset_label: 'Actual Cohort and
Treatment Arm',
  raw_variable: 'COHRT1',
  raw_variable_label: 'Previous Cohort',
  raw_variable_type: 'CheckBox',
  annotation_text: 'SC.SCORRES when
SC.SCTESTCD = \'PCOHORT\'',
  annotation_ordinal: 1,
  target_sdtm_variable_domain: 'SC',
  target_sdtm_variable: 'SCORRES',
  target_sdtm_variable_role: 'Result
Qualifier',
  mapping_algorithm: '01_ASSIGN_NO_CT',
  @id: 'e4acbfe6-a3c3-4bd2-a857-
024990c58faa'
}
```

Efficiency and Consistency



OAK Automation - Metrics from 25 studies



Average values from 25 studies which have completed SDTMv specs.

Data Tabulation Metrics

Standards SDTMv mappings automated	82% (48375 std mappings)
Non-standard SDTMv mappings added by the user	18% (10420 non-std mappings)

Automated with standard Algorithms



97%

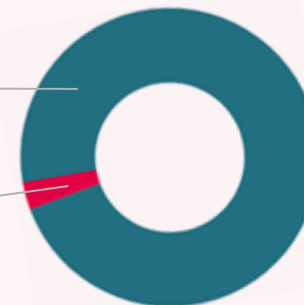
(10149 mappings)

Manually programmed in R



3%

(271 mappings)



Across Standards & Non-standards



97% Automated mappings

(excluding any automation failures)



3% Manually programmed mappings

Annotations Copy Demo



MINT+ [Help / Support](#)

Study ID: BP44315 Analysis label: demo Status: DRAFT RP

[Home](#) [Analysis Configuration](#) [F.A.I.R. Dashboard](#) [SDTM Spec View](#) [Non Standards](#) [Discrepancies](#) [Data Tabulation](#) [Controlled Terminology](#)

Dataset
CM

[FINISH EDITING / UNLOCK STUDY](#) [IMPORT ANNOTATIONS](#)

☐ Show only nonstandards without annotations

	Raw Source Model	Raw Dataset	Raw Dataset Label	Raw Variable	Raw Variable Label	Study Spec	Copy Source	Annotation Texts	Mapping Algorithm	Entity Sub-algorithm	Target SDTM Domain	Target SDTM Variable
	e-CRF DC	CM	Previous and Concomitant			TRUE						
	e-CRF DC	CM	Previous and Concomitant Medication	MDREC	Were there any medications taken?	TRUE						
	e-CRF DC	CM	Previous and Concomitant	MDNUM	MD number	TRUE						
	e-CRF DC	CM	Previous and Concomitant	MDRAW	Medication	TRUE						
	e-CRF DC	CM	Previous and Concomitant	MDIND	Indication	TRUE						
	e-CRF DC	CM	Previous and Concomitant	MD8_DC	Start date	TRUE						
	e-CRF DC	CM	Previous and Concomitant	MD8DTU	Start date unknown	TRUE						
	e-CRF DC	CM	Previous and Concomitant Medication	MDPRIOR	Select if taken prior to study	TRUE						
	e-CRF DC	CM	Previous and Concomitant	MDE_DC	End date	TRUE						
	e-CRF DC	CM	Previous and Concomitant	MDONG	Select if ongoing	TRUE						
	e-CRF DC	CM	Previous and Concomitant Medication	DOS	Dose	TRUE	BP44315	If AE.AESTRTPT is not null then AE.AESTRTPT = 'FIRST DOSE OF STUDY DRUG'	09_IF_THEN_ELSE	06_HARDCODE_NO_CT	AE	AESTRTPT
							BP44315	Else if 'No' then AE.AESTRTPT = 'AFTER'	09_IF_THEN_ELSE	05_HARDCODE_CT	AE	AESTRTPT
							BP44315	If 'Yes' then AE.AESTRTPT = 'BEFORE'	09_IF_THEN_ELSE	05_HARDCODE_CT	AE	AESTRTPT
	e-CRF DC	CM	Previous and Concomitant	DOSU	Dose unit	TRUE						
	e-CRF DC	CM	Previous and Concomitant	MDFORM	Dose form	TRUE						
	e-CRF DC	CM	Previous and Concomitant	MDRTE	Route	TRUE						
	e-CRF DC	CM	Previous and Concomitant	MDFRQ	Dose Frequency	TRUE						
	e-CRF DC	CM	Previous and Concomitant Medication	MDPRMD	Is this medication given as Premedication?	TRUE						

Click any cell to add/edit annotations

Questions?