



ACORN

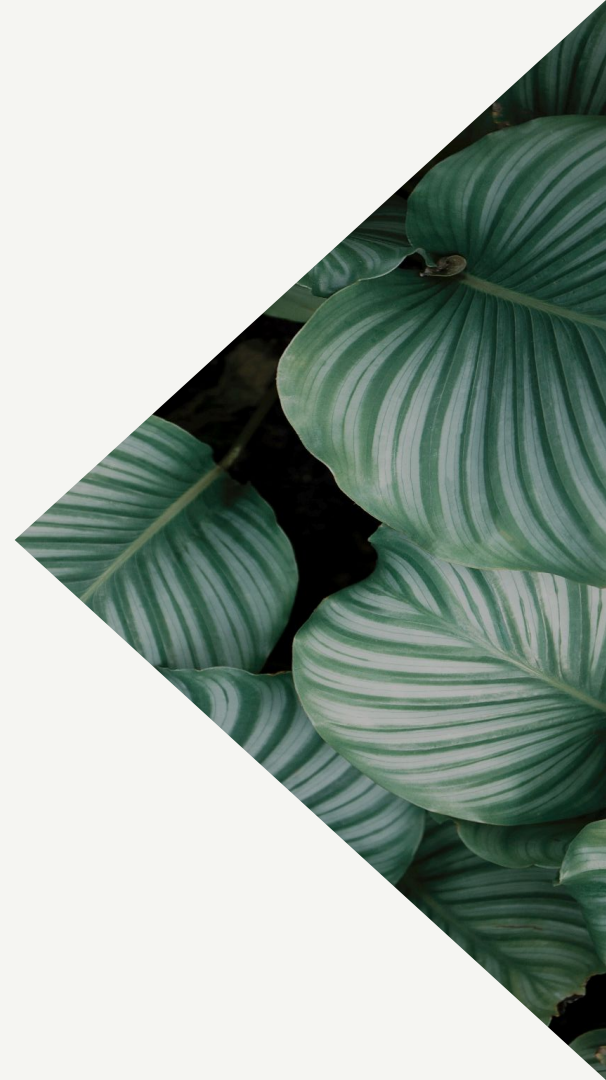
SDTM Mapping
Recommendation Engine

Table of contents

Table of contents

- | | |
|----------|--------------------|
| 1 | CDISC SDTM |
| 2 | Problem Statement |
| 3 | Status Quo |
| 4 | ACORN Introduction |
| 5 | Methodology |
| 6 | Results |
| 7 | Next Steps |

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.

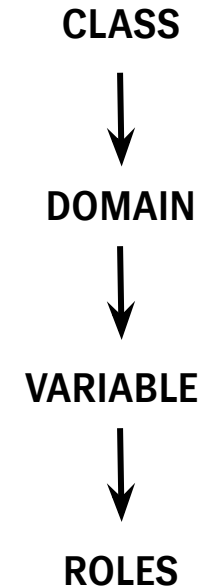
- Interventions
- Events
- Findings

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

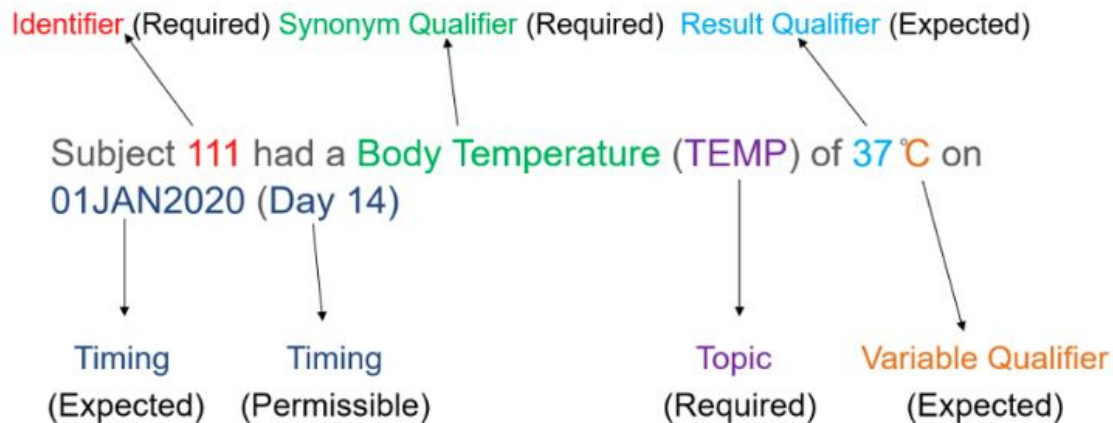
Variables - Each observation can be described by a series of variables, which are classified into the following roles:

- Identifier
- Topic
- Timing
- Qualifier

Hierarchy

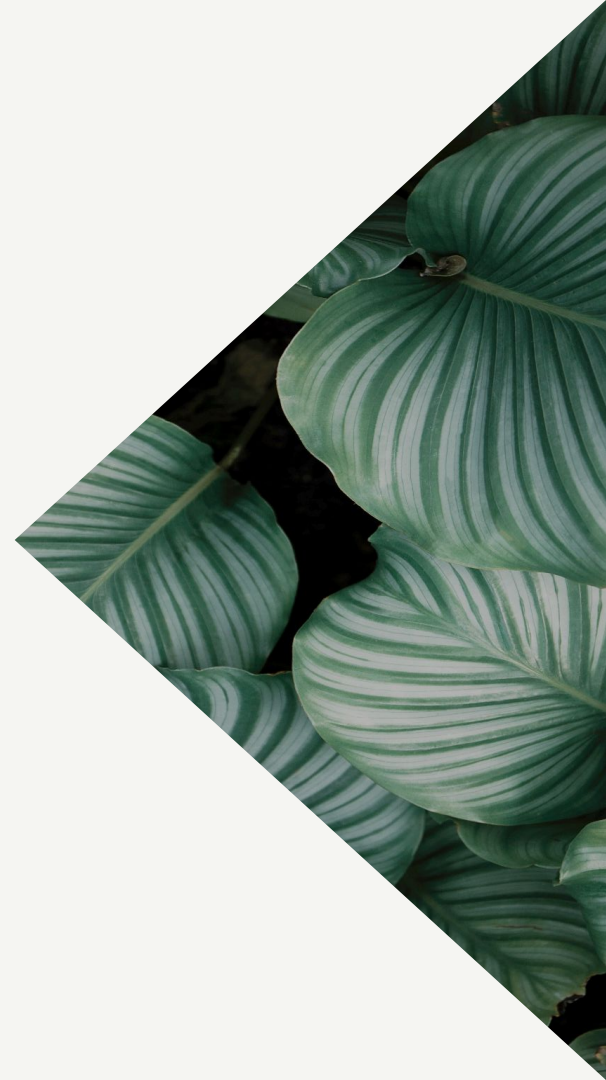


SDTM Mapping



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

Problem Statement



Non-Standards in a Study:

- Roche's MDR (Metadata Repository) has a wealth of SDTM mappings defined for Data Collection standards.
- Often to meet the Clinical Protocol Requirements, Clinical studies need to collect data that is not defined in MDR. Roche calls this Non-Standard data.

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.

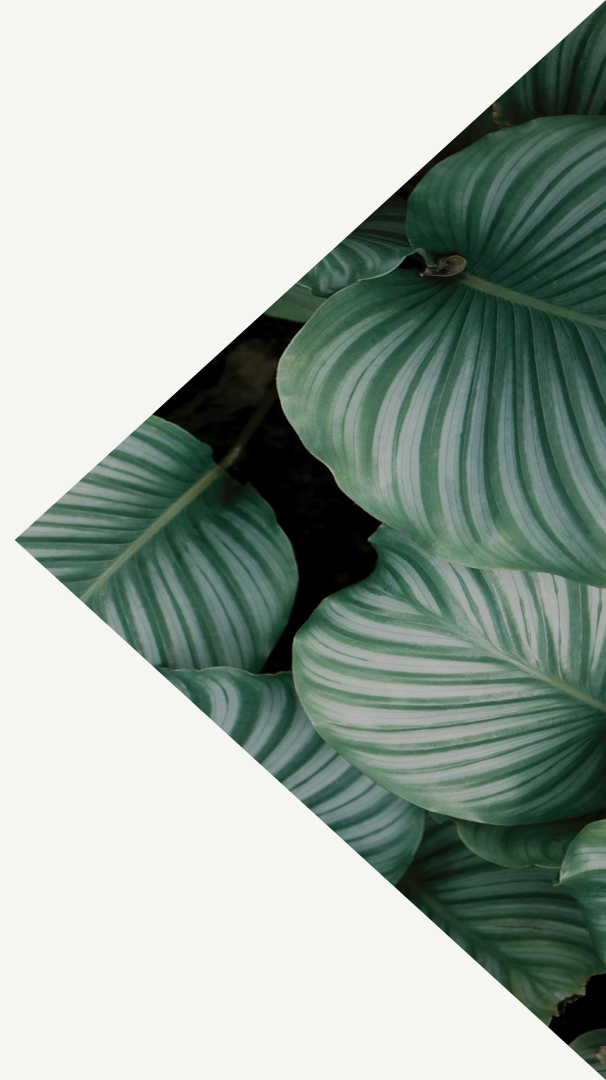
Leading to Inconsistency

- 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).

Status Quo



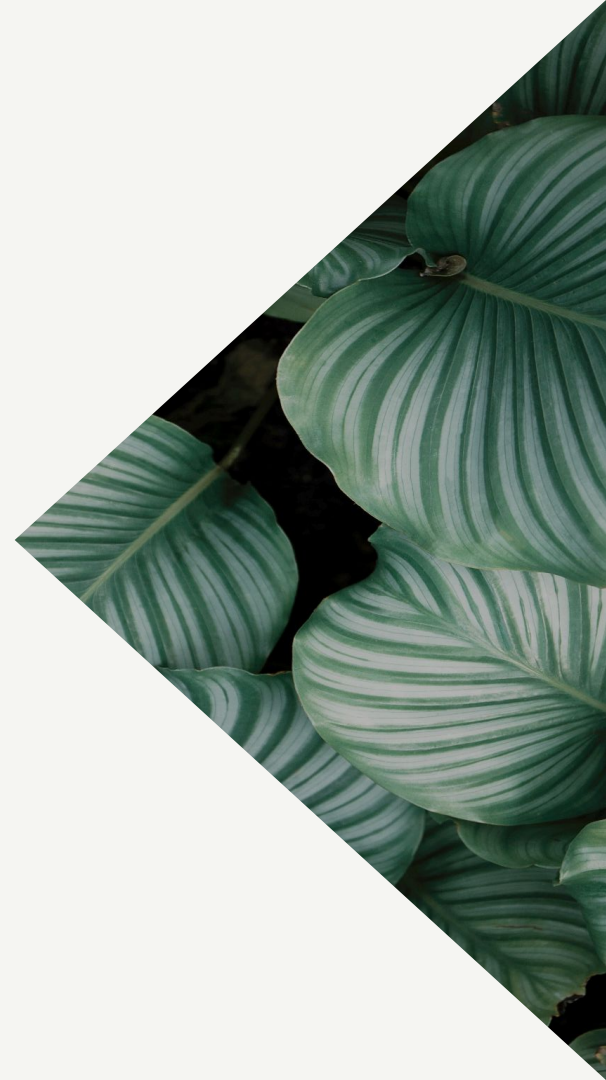
Previous Work:

- Requires actual patient data along with the metadata
- Model SDTM mapping as a Classification Problem with SDTM variables as Classes
- Either domain to be mapped is a prerequisite or work with limited number of domains, usually 5-6.

Caveats:

- Actual patient data may not be always available
- Not enough data to train as a classification task
- 5-6 domains amount to ~200 classes (variables in the domains), when expanded to all CDISC domains (~70), this number would reach ~2000 classes.

ACORN Introduction



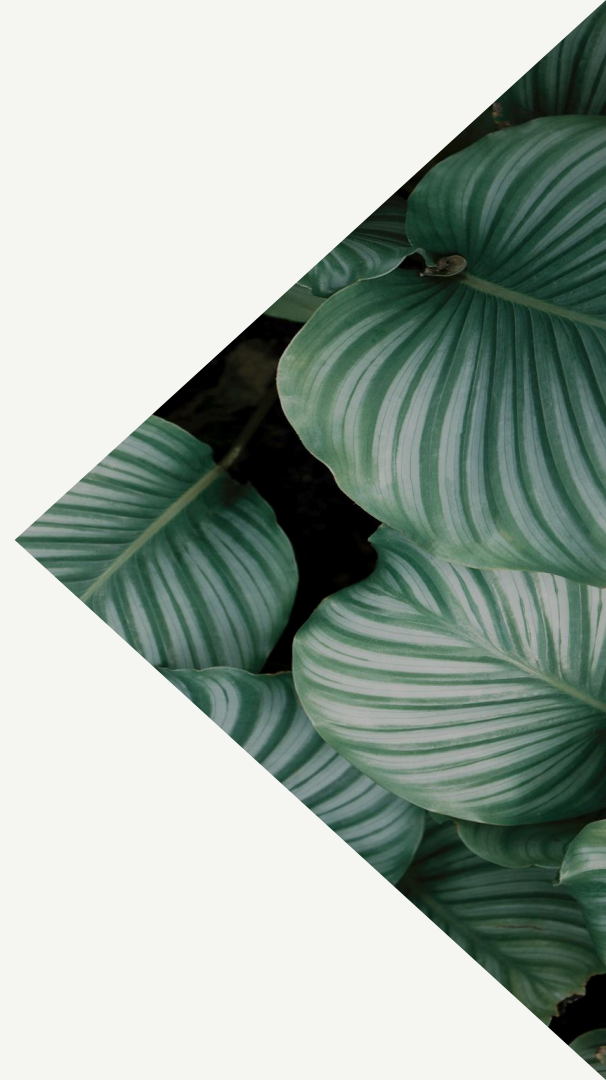
What is ACORN?

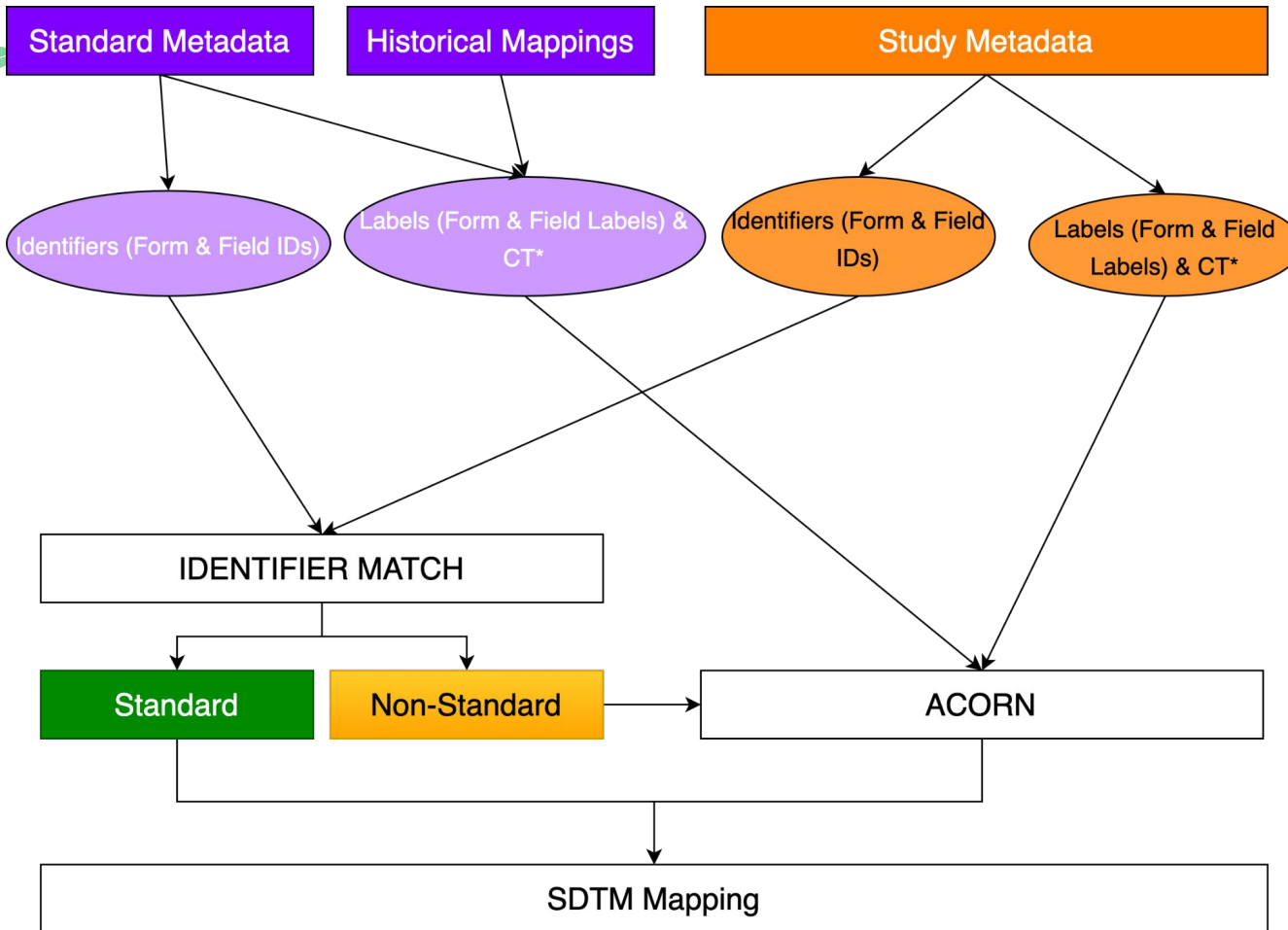
- ACORN is a part of Roche's SDTM automation flow (OAK) and will recommend the SDTM mappings to the users.
- ACORN aims to recommend the SDTM Mappings for Non-standards in the study by comparing the Non-Standard metadata in the study to the Historical mappings done by SMEs for the same molecule, TA or Indication and Standard mappings present in MDR
- The users will be able to review the recommendations while working on Non-standard SDTM mappings in their study and decide to either keep the suggestions or change them to suit their study needs.

Potential Benefits -

- ACORN strives to reduce inconsistencies in mapping the Non-standard Data collection elements in Roche studies.
- Eliminating the inconsistencies will create time savings in mapping and programming of Non-Standards in every clinical study.
- Consistent SDTM data enables easier pooling of SDTM data across studies.
- Upon implementation, ACORN will learn from user behaviour and recommendations will get more and more accurate as more user start interacting with ACORN.

Methodology





ACORN development involves

- ❖ SDTM Mappings Metadata Preparation
- ❖ Development of Recommendation Pipeline
- ❖ Review Results

SDTM Mapping - Metadata Preparation

Available

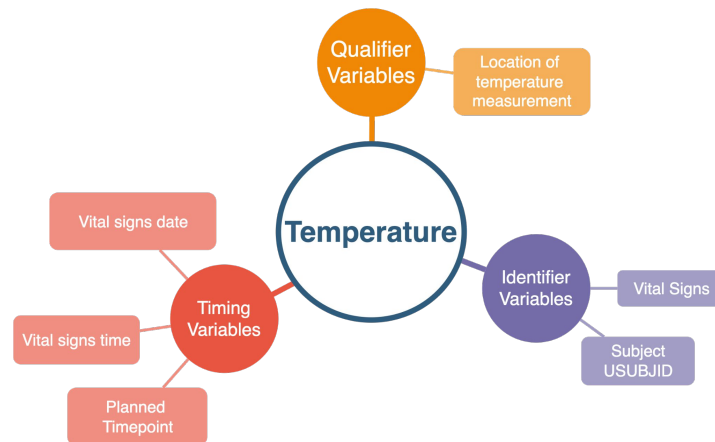
Metadata

- Historical Metadata - 110 clinical studies with CRFs SDTM mappings by Metadata Scientists
 - Reference Data - 84 studies (934 annotated CRFs)
 - Test Data - 26 studies (286 annotated CRFs)
- Standard Metadata
 - MDR (Roche's Metadata Repository) has ~12000+ unique SDTM mappings.

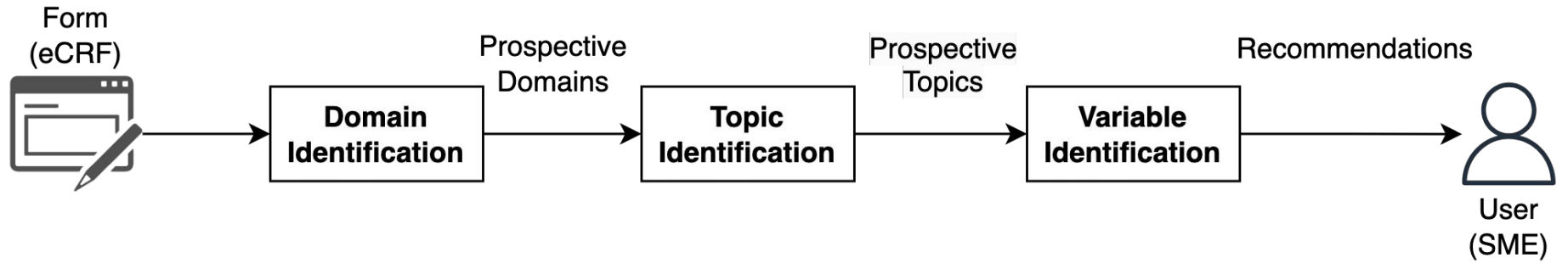
Standard Set

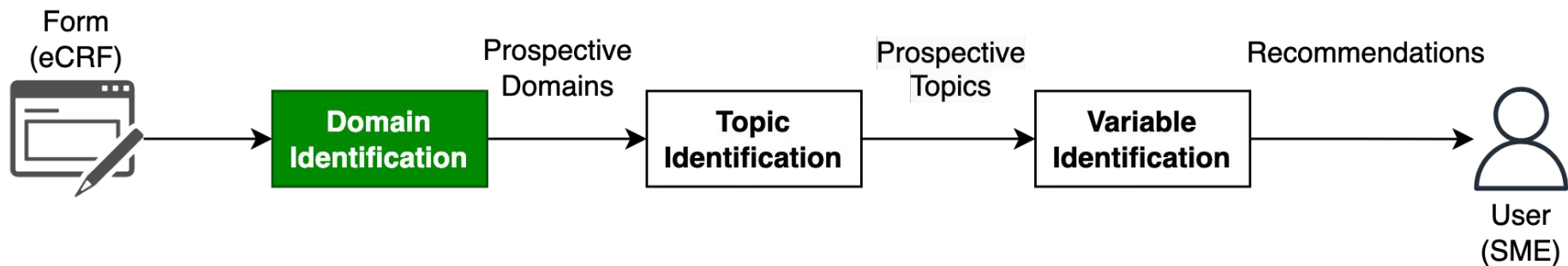


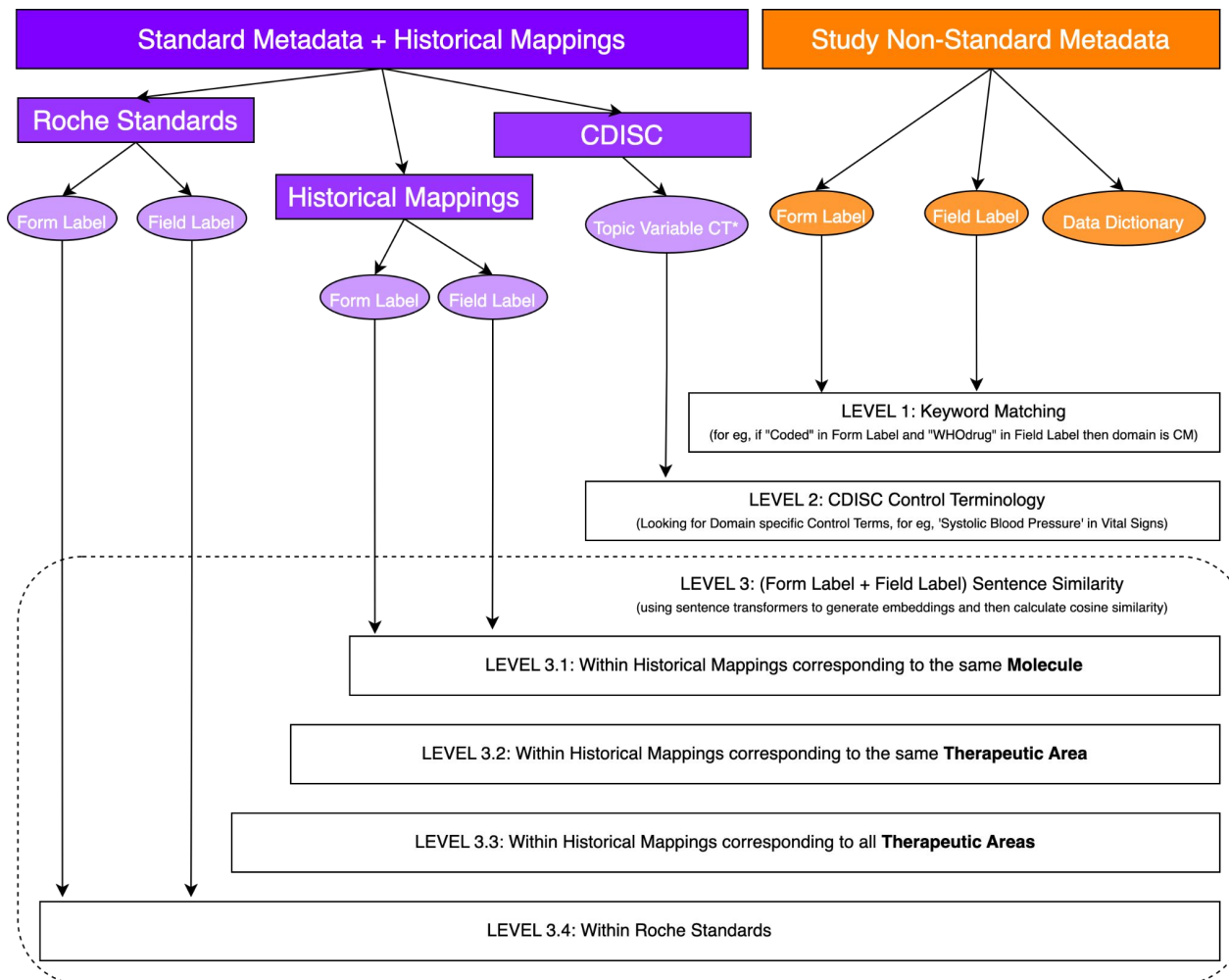
Example Standard Set



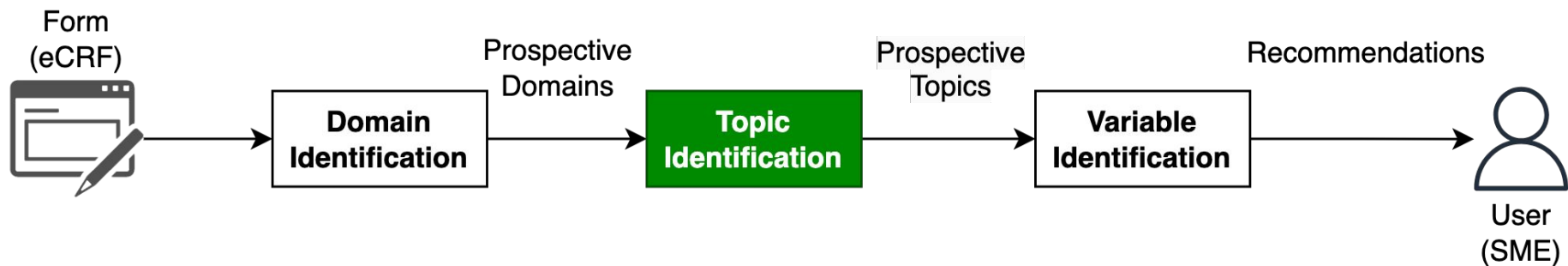
Recommendation Pipeline





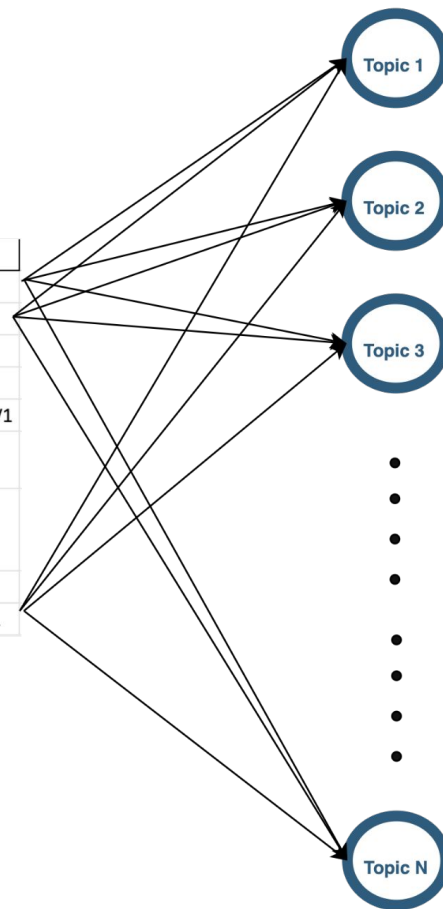


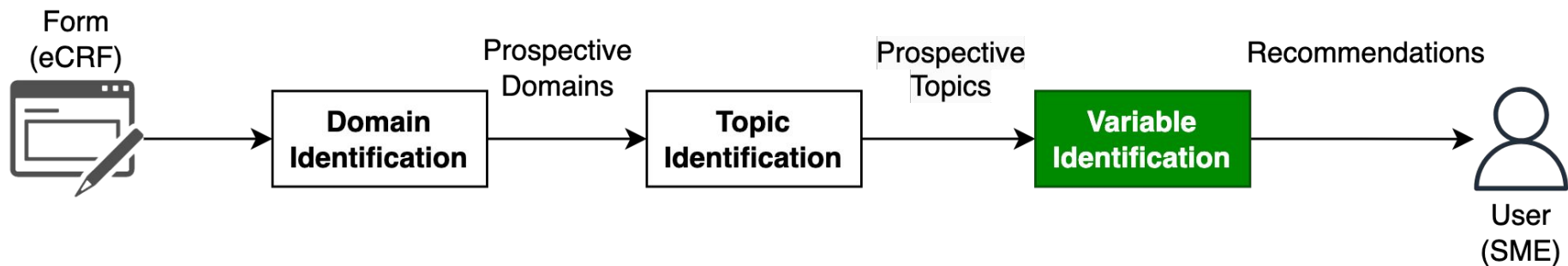
* Control Terminology



COVID-19 Non-Protocol-Specified Visual Acuity Testing

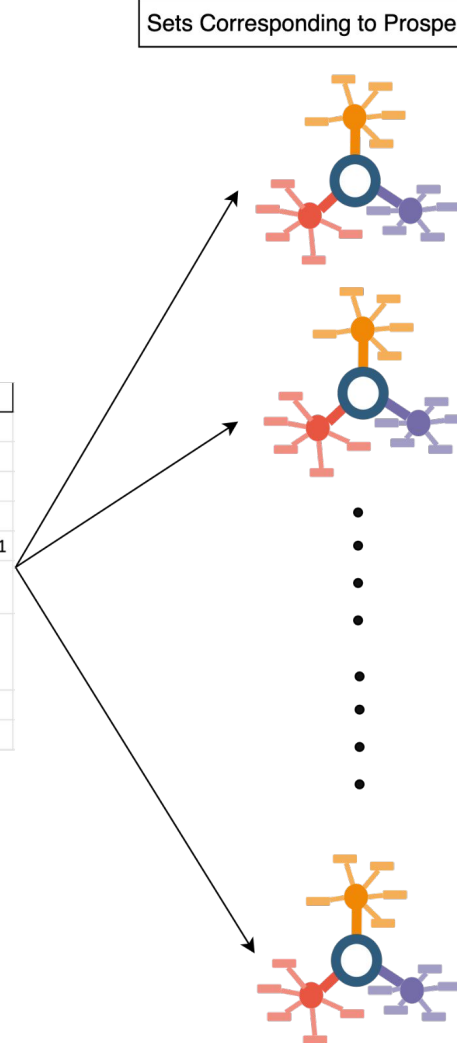
FormOID	FieldOID	Field Label	Format	Dictionary
COVNPVA	STYVNUM	Study Visit#	\$50	
COVNPVA	ASMD	Date of assessment	dd MMM yyyy	
COVNPVA	LOC DV	Site	\$40	
COVNPVA	LAT DV	Side	\$25	
COVNPVA	RFNMT HD	Refraction method used	\$35	RFNMT HD_CV1
COVNPVA	VATTYP	Which type of VA testing was done?	\$10	VATTYP_CV1
COVNPVA	MAXDST	Distance (in meters/inches) at which VA testing was performed:	\$13	MAXDST_CV1
COVNPVA	LETNUM	ETDRS BCVA total letter score	3	
COVNPVA	SNLEQVL	Snellen	\$45	SNLEQVL_CV1



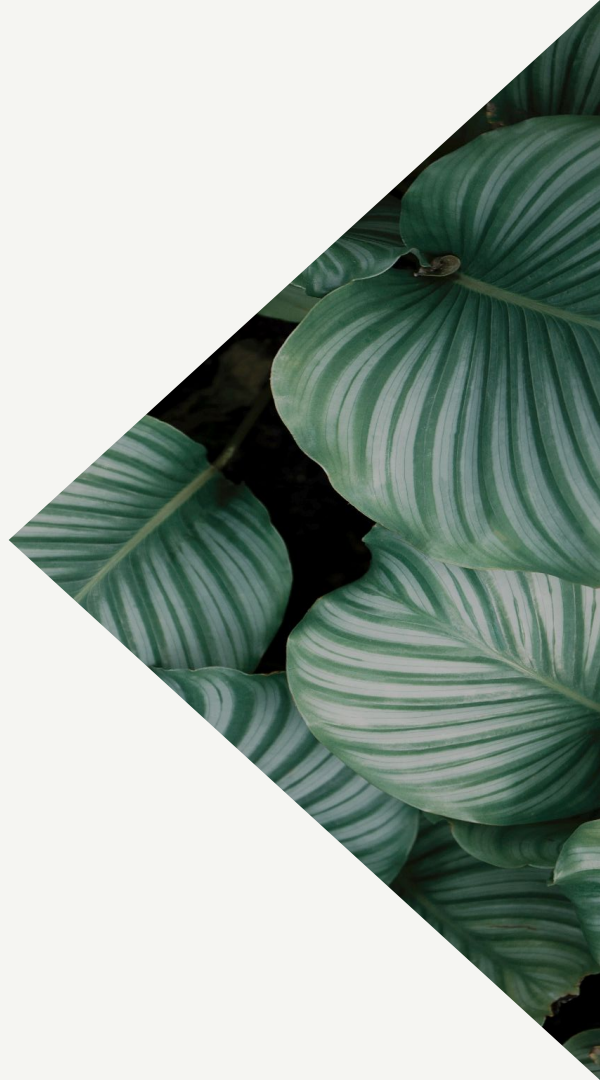


COVID-19 Non-Protocol-
Specified Visual Acuity
Testing

FormOID	FieldOID	Field Label	Format	Dictionary
COVNPVA	STVNUM	Study Visit#	\$50	
COVNPVA	ASMD	Date of assessment	dd MMM yyyy	
COVNPVA	LOCDV	Site	\$40	
COVNPVA	LATDV	Side	\$25	
COVNPVA	RFMNTHD	Refraction method used	\$35	RFNMTHD_CV1
COVNPVA	VATTYP	Which type of VA testing was done?	\$10	VATTYP_CV1
COVNPVA	MAXDST	Distance (in meters/inches) at which VA testing was performed:	\$13	MAXDST_CV1
COVNPVA	LETNUM	ETDRS BCVA total letter score	3	
COVNPVA	SNLEQVL	Snellen	\$45	SNLEQVL_CV1

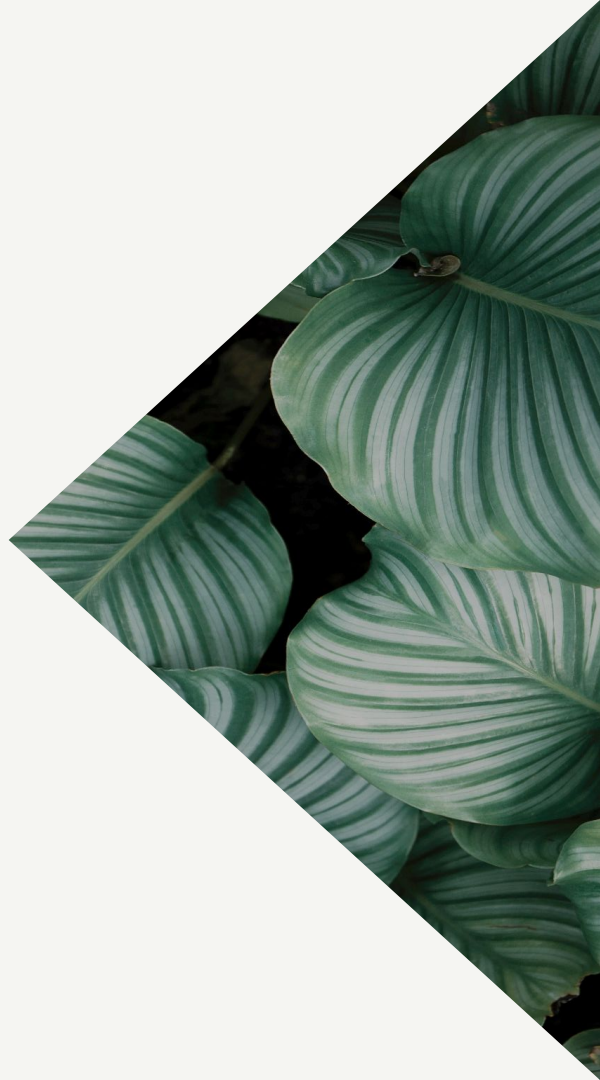


Results



Field Label	SME defined SDTM Annotation	score	Predicted Variables	Predicted Annotation Text
Date of assessment	OE.OEDTC; FA.FADTC	1	['OEDTC']	['Date part of OE.OEDTC;Date part of FA.FADTC']
Site	OE.OELOC	1	['OELOC']	['OE.OELOC']
Side	OE.OELAT	1	['OELAT']	['OE.OELAT']
Refraction method used	FA.FAORRES when FA.FATESTCD = METHOD; FA.FATEST = 'Method'; FA.FAOBJ = 'Refraction'	1	['FATEST', 'FATESTCD', 'FAORRES', 'FAOBJ']	["FA.FAORRES when FATESTCD = METHOD;\nFATEST = Method;FA.FAOBJ = 'Refraction'"]
Which type of VA testing was done?	OE.OEMETHOD	0.99	['OEMETHOD']	['OE.OEMETHOD "EYE CHART"]
Distance (in meters/inches) at which VA testing was performed:	OE.OEORRES when OE.OETESTCD = MAXDIST; OE.OETEST = 'Testing Distance'; OE.OESCAT = DISTANCE AT VA TESTING	0.99	['OETESTCD', 'OETSTDTL', 'OETEST', 'OEORRESU', 'OEORRES']	["OE.OEORRES where OE.OETESTCD = CFDISTM;OE.OETEST = 'Count Fingers Maximum Distance';OE.OEORRESU = 'ft';OE.OETSTDTL = TESTING DISTANCE: <MAXDST>"]
ETDRS BCVA total letter score	OE.OEORRES when OE.OETESTCD = NUMLCOR; OE.OETEST = 'Number of Letters Correct'; OE.OESCAT = TOTAL	1	['OETEST', 'OETESTCD', 'OESCAT', 'OEORRES']	["OE.OEORRES when OETESTCD = NUMLCOR;OE.OETEST = 'Number of Letters Correct';OE.OESCAT = TOTAL."]
Snellen	OE.OERESCAT	1	['OERESCAT']	["OE.OERESCAT = 'ROW [Row number] - SNELEN [Snellen equivalent]"]

Next Steps



Next Steps -

- Not Submitted Fields - Identifying which data need not be submitted to the health authorities
- Using Named Entity Recognition to identify drugs and diseases in the form and field label (which can help identify domains like EC, CM)
- Identify the source data that does not fit to any domains, that can be mapped to Findings About domains (FA,SR)
- Identify questionnaires mappings my looking into Questionnaire label, Question label and Question Responses.
- Extend the model to find SDTM mappings for Non-standard Fields in Standard Forms.

Doing now what patients need next