Supercharging SDTM Automation with AI - MAPIT

PHUSE SDE, Beijing 2025

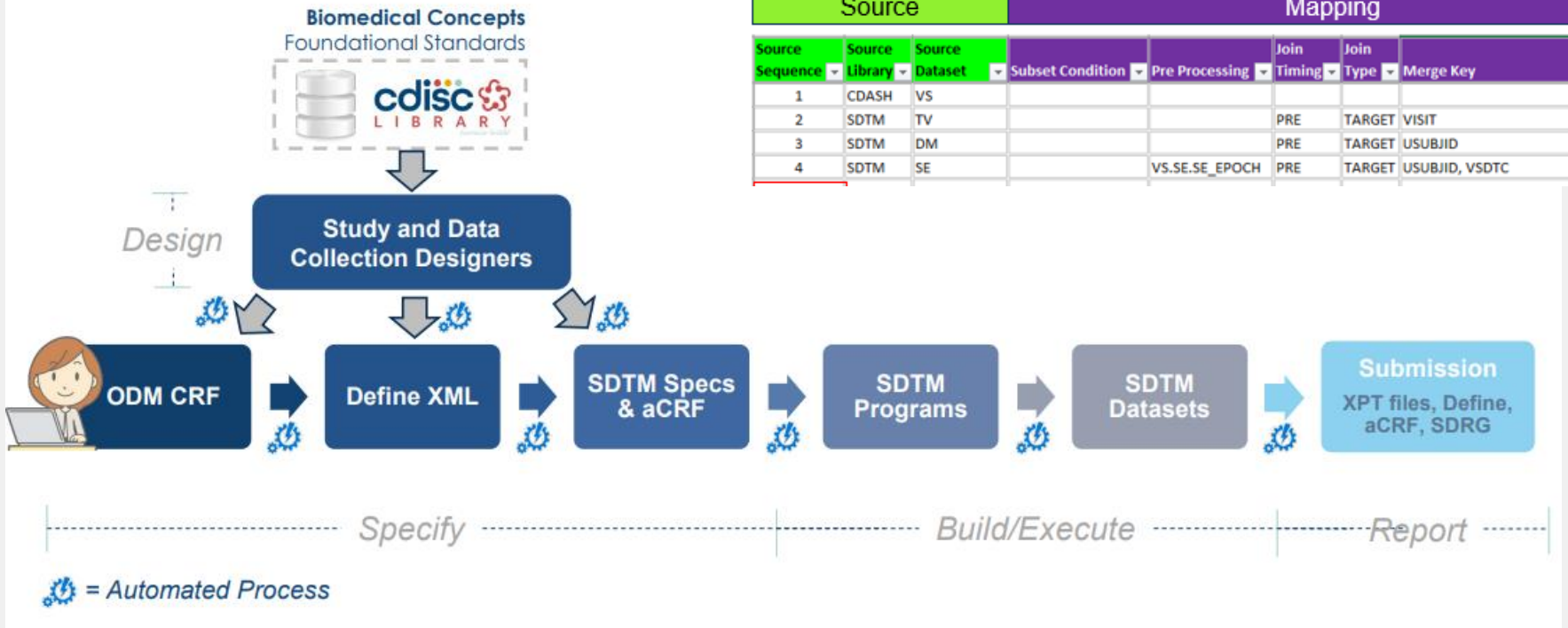
phastar

Agenda

- CDISC SDTM Automation
- Phastar SDTM Automation
- Practical Use SDTM Metadata
- Sentence Transformer Model
- Mapping Results
- Summary
- Q&A

SDTM Automation Concept - CDISC 360

Source				Mapping					Target						
Source Sequence	Source Library	Source Dataset	Source Variable	Map Sequence	Origin	Method	Comment	Code List	Target Library	Target Dataset	Target Variable	Target Description	Target Data Type	Target Length	Target Sorting Order
1	CDASH	VS			Assigned		CDISC360-2		SDTM	VS	STUDYID	Study Identifier	text	10	1
1	CDASH	VS			Assigned		VS	DOMAIN	SDTM	VS	DOMAIN	Domain Abbreviation	text	2	2
1	CDASH	VS	SUBJID		Assigned	ALL.USUBJID			SDTM	VS	USUBJID	Unique Subject Identifier	text	14	3
1	CDASH	VS			Assigned	VS.VSSPID			SDTM	VS	VSSPID	Sponsor-Defined Identifier	text	4	5
1	CDASH	VS	VISIT		Convert			VISITNUM	SDTM	VS	VISITNUM	Visit Number	integer	8	16
1	CDASH	VS	VISIT		Predecessor			VISIT	SDTM	VS	VISIT	Visit Name	text	18	17
1	CDASH	VS	VSDAT		Assigned	VS.VSDTC			SDTM	VS	VSDTC	Date/Time of Measurements	date	10	19

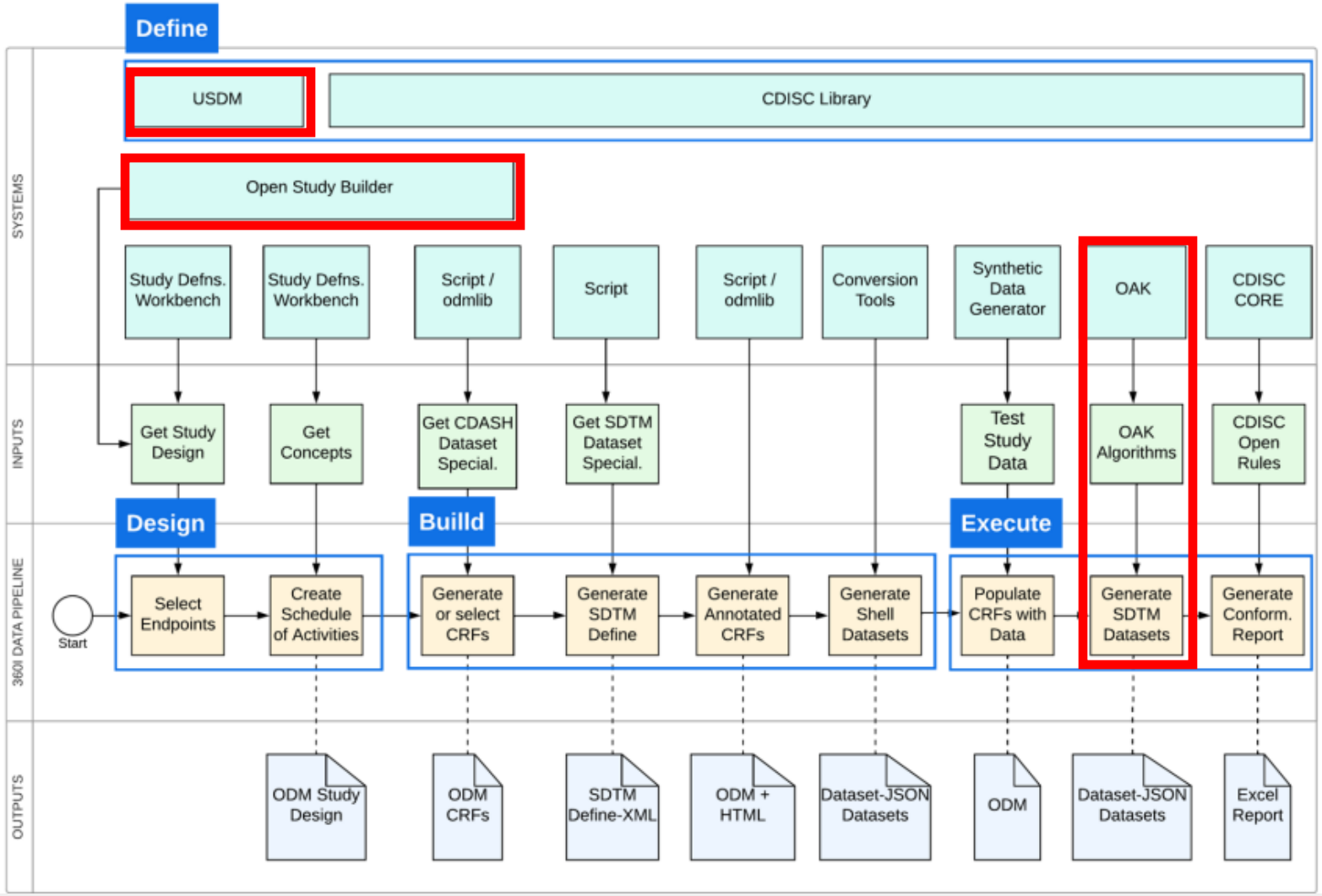


Source			Mapping					Target		
Source Sequence	Source Library	Source Dataset	Subset Condition	Pre Processing	Join Timing	Join Type	Merge Key	Target Sequence	Target Library	Target Dataset
1	CDASH	VS						5	SDTM	VS
2	SDTM	TV			PRE	TARGET	VISIT	5	SDTM	VS
3	SDTM	DM			PRE	TARGET	USUBJID	5	SDTM	VS
4	SDTM	SE		VS.SE.SE_EPOCH	PRE	TARGET	USUBJID, VSDTC	5	SDTM	VS

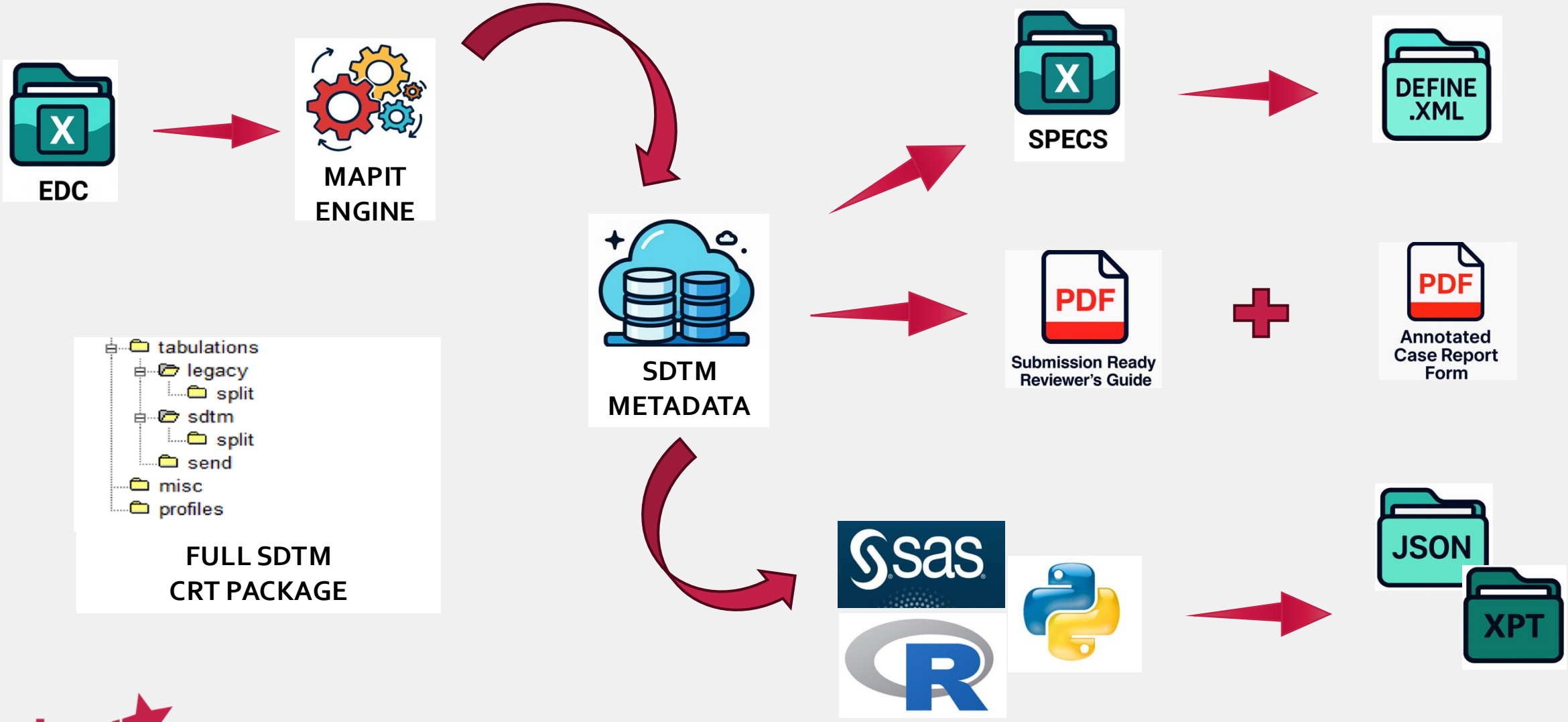
Current CDISC Approach – CDISC 360i

CDISC 360i Phase 1: Generating SDTM

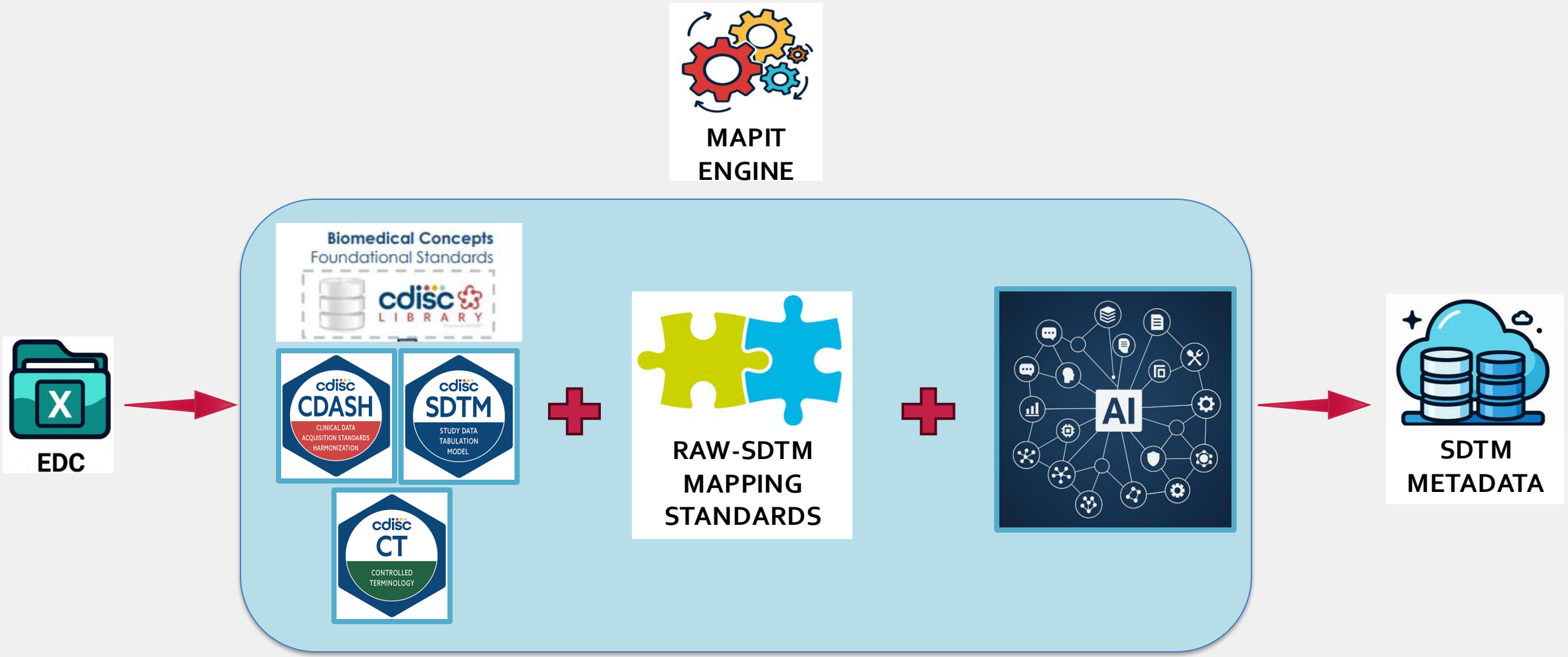
Technical Process



Phastar SDTM Automation – MAPIT



Phastar SDTM Automation – MAPIT



Using SDTM Metadata - Code

RAW_TABLE	RAW_COLUMN	RAW_TYPE	RAW_LENGTH	RAW_LABEL	RAW_FORMAT	RAW_COLUMN_WHRVALUE	SDTM_TAB	SDTM_COLUMN	SDTM_WHRCLAUSE	SDTM_ASSIGNED_VALUE
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"0"	AE	AEACN		"DOSE NOT CHANGED"
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"1"	AE	AEACN		"DOSE INCREASED"
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"2"	AE	AEACN		"DOSE REDUCED"
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"3"	AE	AEACN		"DRUG INTERRUPTED"
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"4"	AE	AEACN		"DRUG WITHDRAWN"
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"98"	AE	AEACN		"NOT APPLICABLE"
AE	AEOUT	C	2	Outcome of AE	$AEOUT.	"0"	AE	AEOUT		"RECOVERED/RESOLVED"
AE	AEOUT	C	2	Outcome of AE	$AEOUT.	"1"	AE	AEOUT		"RECOVERING/RESOLVING"
AE	AEOUT	C	2	Outcome of AE	$AEOUT.	"2"	AE	AEOUT		"RECOVERED/RESOLVED WITH SEQUELAE"
AE	AEOUT	C	2	Outcome of AE	$AEOUT.	"3"	AE	AEOUT		"NOT RECOVERED/NOT RESOLVED"
AE	AEOUT	C	2	Outcome of AE	$AEOUT.	"4"	AE	AEOUT		"FATAL"

```
*** RAW column =AEACN ***;
if AEACN= "0" then S__AEACN = "DOSE NOT CHANGED";
else if AEACN= "1" then S__AEACN = "DOSE INCREASED";
else if AEACN= "2" then S__AEACN = "DOSE REDUCED";
else if AEACN= "3" then S__AEACN = "DRUG INTERRUPTED";
else if AEACN= "4" then S__AEACN = "DRUG WITHDRAWN";
else if AEACN= "98" then S__AEACN = "NOT APPLICABLE";
```

```
# RAW column = AEACN
mutate(S__AEACN =
         case_when(
           AEACN=="0" ~ "DOSE NOT CHANGED",
           AEACN=="1" ~ "DOSE INCREASED",
           AEACN=="2" ~ "DOSE REDUCED",
           AEACN=="3" ~ "DRUG INTERRUPTED",
           AEACN=="4" ~ "DRUG WITHDRAWN",
           AEACN=="98" ~ "NOT APPLICABLE"
       ))  %>%
```

```
# Map the "AEACN" column to a new column called "S__AEACN"
mapping = {
    "0": "DOSE NOT CHANGED",
    "1": "DOSE INCREASED",
    "2": "DOSE REDUCED",
    "3": "DRUG INTERRUPTED",
    "4": "DRUG WITHDRAWN",
   "98": "NOT APPLICABLE",
}

# Apply mapping to create a new column "s__aeacn"
df['s__aeacn'] = df['aeacn'].map(mapping)
```

```
*** RAW column =AEOUT ***;
if AEOUT= "0" then S__AEOUT = "RECOVERED/RESOLVED";
else if AEOUT= "1" then S__AEOUT = "RECOVERING/RESOLVING";
else if AEOUT= "2" then S__AEOUT = "RECOVERED/RESOLVED WITH SEQUELAE";
else if AEOUT= "3" then S__AEOUT = "NOT RECOVERED/NOT RESOLVED";
else if AEOUT= "4" then S__AEOUT = "FATAL";
```

```
# RAW column = AEOUT
mutate(S__AEOUT =
         case_when(
           AEOUT=="0" ~ "RECOVERED/RESOLVED",
           AEOUT=="1" ~ "RECOVERING/RESOLVING",
           AEOUT=="2" ~ "RECOVERED/RESOLVED WITH SEQUELAE",
           AEOUT=="3" ~ "NOT RECOVERED/NOT RESOLVED",
           AEOUT=="4" ~ "FATAL"
       ))  %>%
```

```
# Map the "AEOUT" column to a new column called "S__AEOUT"
mapping = {
    "0": "RECOVERED/RESOLVED",
    "1": "RECOVERING/RESOLVING",
    "2": "RECOVERED/RESOLVED WITH SEQUELAE",
    "3": "NOT RECOVERED/NOT RESOLVED",
    "4": "DRUG WITHDRAWN",
}

# Apply mapping to create a new column "s__aeout"
df['s__aeout'] = df['aeout'].map(mapping)
```

Using SDTM Metadata – Specifications / Define

RAW_TABLE	RAW_COLUM	RAW_TYP	RAW_LENGT	RAW_LABEL	RAW_FORMAT	RAW_COLUMN_WHRVALUE	SDTM_TAB	SDTM_COLUM	SDTM_WHRCLAUSE	SDTM_ASSIGNED_VALUE
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"0"	AE	AEACN		"DOSE NOT CHANGED"
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"1"	AE	AEACN		"DOSE INCREASED"
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"2"	AE	AEACN		"DOSE REDUCED"
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"3"	AE	AEACN		"DRUG INTERRUPTED"
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"4"	AE	AEACN		"DRUG WITHDRAWN"
AE	AEACN	C	2	Action Taken, Investig	$AEACN.	"98"	AE	AEACN		"NOT APPLICABLE"
AE	AEOUT	C	2	Outcome of AE	$AEOUT.	"0"	AE	AEOUT		"RECOVERED/RESOLVED"
AE	AEOUT	C	2	Outcome of AE	$AEOUT.	"1"	AE	AEOUT		"RECOVERING/RESOLVING"
AE	AEOUT	C	2	Outcome of AE	$AEOUT.	"2"	AE	AEOUT		"RECOVERED/RESOLVED WITH SEQUELAE"
AE	AEOUT	C	2	Outcome of AE	$AEOUT.	"3"	AE	AEOUT		"NOT RECOVERED/NOT RESOLVED"
AE	AEOUT	C	2	Outcome of AE	$AEOUT.	"4"	AE	AEOUT		"FATAL"

ID	Name	NCI Codelist Code	Data Type	Order	Term	NCI Term Code
AEACN	Action Taken with Study Treatment	C66767	text	1	DOSE NOT CHANGED	C49504
AEACN	Action Taken with Study Treatment	C66767	text	2	DOSE INCREASED	C49503
AEACN	Action Taken with Study Treatment	C66767	text	3	DOSE REDUCED	C49505
AEACN	Action Taken with Study Treatment	C66767	text	4	DRUG INTERRUPTED	C49501
AEACN	Action Taken with Study Treatment	C66767	text	5	DRUG WITHDRAWN	C49502
AEACN	Action Taken with Study Treatment	C66767	text	6	NOT APPLICABLE	C48660
AEOUT	Outcome of Event	C66768	text	1	RECOVERED/RESOLVED	C49498
AEOUT	Outcome of Event	C66768	text	2	RECOVERING/RESOLVING	C49496
AEOUT	Outcome of Event	C66768	text	3	RECOVERED/RESOLVED WITH SEQUELAE	C49495
AEOUT	Outcome of Event	C66768	text	4	NOT RECOVERED/NOT RESOLVED	C49494
AEOUT	Outcome of Event	C66768	text	5	FATAL	C48275

Action Taken with Study Treatment [*C66767*]

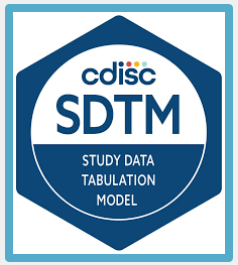
Permitted Value (Code)
DOSE NOT CHANGED [*C49504*]
DOSE INCREASED [*C49503*]
DOSE REDUCED [*C49505*]
DRUG INTERRUPTED [*C49501*]
DRUG WITHDRAWN [*C49502*]
NOT APPLICABLE [*C48660*]

Outcome of Event [*C66768*]

Permitted Value (Code)
RECOVERED/RESOLVED [*C49498*]
RECOVERING/RESOLVING [*C49496*]
RECOVERED/RESOLVED WITH SEQUELAE [*C49495*]
NOT RECOVERED/NOT RESOLVED [*C49494*]
FATAL [*C48275*]

Methods of Applying SDTM Mapping

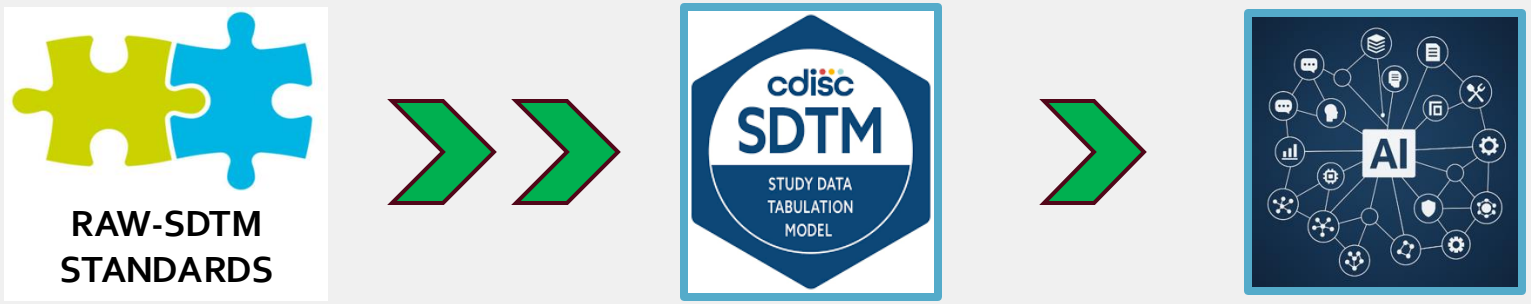
- Client Standards
  - High accuracy
  - Bespoke / complex mapping
- RAW - SDTM IG look-up
  - Useful when standards don't exist
  - Limited to strict matching
- AI methods
  - Semantic matching – more flexible
  - Trained on historical data – improves over time



Methods of Applying SDTM Mapping

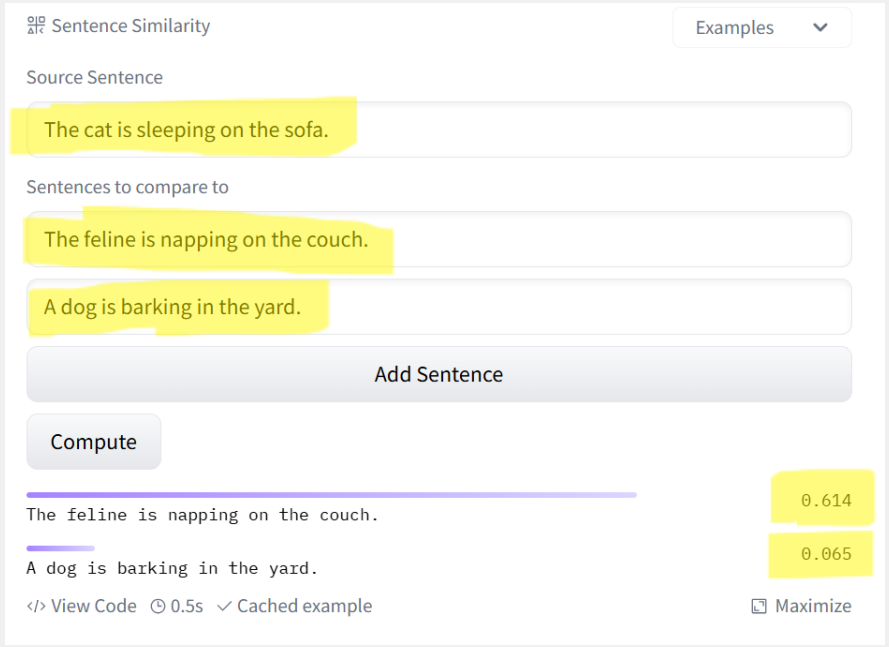
- None of the approaches are mutually exclusive
- Known mappings are better when available
- AI supplements deterministic approach i.e. fills gaps

Deterministic **Predictive**



Semantic Matching - Introduction

- Use a Sentence Transformer Model - all-MiniLM-L6-v2
- Vectorize and compare meanings between sentences for similarity
- How does this work in practice?




```python
from sentence_transformers import SentenceTransformer, util

model = SentenceTransformer('all-MiniLM-L6-v2')
embeddings = model.encode([
    "The cat is sleeping on the sofa.",
    "The feline is napping on the couch.",
    "A dog is barking in the yard."
], convert_to_tensor=True)

score1 = util.cos_sim(embeddings[0], embeddings[1])
score2 = util.cos_sim(embeddings[0], embeddings[2])
print(score1, score2)
```

```
tensor([[0.6138]]) tensor([[0.0653]])
```

https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2

Semantic Matching - SDTM

raw_table_name	raw_table	raw_colun	raw_label	raw_code_	decode
Adverse Events (AE)	AE	AESER	Serious AE	0	No
Adverse Events (AE)	AE	AESER	Serious AE	1	Yes
Adverse Events (AE)	AE	AEACN	Action Taken, Inv	0	Dose not changed
Adverse Events (AE)	AE	AEACN	Action Taken, Inv	1	Dose increased
Adverse Events (AE)	AE	AEACN	Action Taken, Inv	2	Dose reduced
Adverse Events (AE)	AE	AEACN	Action Taken, Inv	3	Drug interrupted
Adverse Events (AE)	AE	AEACN	Action Taken, Inv	4	Drug permanently discontinued
Adverse Events (AE)	AE	AEACN	Action Taken, Inv	98	Not applicable
Adverse Events (AE)	AE	AEOUT	Outcome of AE	0	Recovered/resolved
Adverse Events (AE)	AE	AEOUT	Outcome of AE	1	Recovering/resolving
Adverse Events (AE)	AE	AEOUT	Outcome of AE	2	Recovered/resolved with sequelae
Adverse Events (AE)	AE	AEOUT	Outcome of AE	3	Not recovered/not resolved
Adverse Events (AE)	AE	AEOUT	Outcome of AE	4	Fatal
Demographics (DM)	DM	SEX	Sex	C16576	Female
Demographics (DM)	DM	SEX	Sex	C20197	Male

Context is used from multiple fields from raw and SDTM metadata in a process called **enriching** to enhance semantic matching

raw_text = f"{raw_label}: {decode}"
sdtm_text = f"{label}: {preferred_term} (submission_value)"

domain	dataset	name	label	code	codelist_code	cdisc_submission_valu	cdisc_synonym	nci_preferred_term	cdisc_definition
AE	Adverse Events	AESER	Serious Event	C17998	C66742	U	U; UNK; Unknown	Unknown	Not known, not observed, not recorded, or refused. (NCI
AE	Adverse Events	AESER	Serious Event	C48660	C66742	NA	NA; Not Applicable	Not Applicable	Determination of a value is not relevant in the current co
AE	Adverse Events	AESER	Serious Event	C49487	C66742	N	No	No	The non-affirmative response to a question. (NCI)
AE	Adverse Events	AESER	Serious Event	C49488	C66742	Y	Yes	Yes	The affirmative response to a question. (NCI)
AE	Adverse Events	AEACN	Action Taken with Study Treatment	C150826	C66767	DOSE RATE REDUCED		Dose Rate Reduced	An indication that the administration of a medication w
AE	Adverse Events	AEACN	Action Taken with Study Treatment	C17998	C66767	UNKNOWN	U; UNK; Unknown	Unknown	Not known, not observed, not recorded, or refused. (NC
AE	Adverse Events	AEACN	Action Taken with Study Treatment	C48660	C66767	NOT APPLICABLE	NA; Not Applicable	Not Applicable	Determination of a value is not relevant in the current c
AE	Adverse Events	AEACN	Action Taken with Study Treatment	C49501	C66767	DRUG INTERRUPTED		Drug Interrupted	An indication that a medication schedule was modified
AE	Adverse Events	AEACN	Action Taken with Study Treatment	C49502	C66767	DRUG WITHDRAWN		Drug Withdrawn	An indication that a medication schedule was modified
AE	Adverse Events	AEACN	Action Taken with Study Treatment	C49503	C66767	DOSE INCREASED		Dose Increased	An indication that a medication schedule was modified
AE	Adverse Events	AEACN	Action Taken with Study Treatment	C49504	C66767	DOSE NOT CHANGED		Dose Not Changed	An indication that a medication schedule was maintain
AE	Adverse Events	AEACN	Action Taken with Study Treatment	C49505	C66767	DOSE REDUCED		Dose Reduced	An indication that a medication schedule was modified
AE	Adverse Events	AEOUT	Outcome of Adverse Event	C17998	C66768	UNKNOWN	U; UNK; Unknown	Unknown	Not known, not observed, not recorded, or refused. (NCI
AE	Adverse Events	AEOUT	Outcome of Adverse Event	C48275	C66768	FATAL	5; FATAL; Grade 5	Death Related to Advers	The termination of life as a result of an adverse event. (N
AE	Adverse Events	AEOUT	Outcome of Adverse Event	C49494	C66768	NOT RECOVERED/NOT RESOLVED		Not Recovered from Adv	One of the possible results of an adverse event outcome
AE	Adverse Events	AEOUT	Outcome of Adverse Event	C49495	C66768	RECOVERED/RESOLVED WITH SEQUELAE		Recovered from Adverse	One of the possible results of an adverse event outcome
AE	Adverse Events	AEOUT	Outcome of Adverse Event	C49496	C66768	RECOVERING/RESOLVING		Recovering or Resolving	One of the possible results of an adverse event outcome
AE	Adverse Events	AEOUT	Outcome of Adverse Event	C49498	C66768	RECOVERED/RESOLVED		Recovered from Adverse	One of the possible results of an adverse event outcome
DM	Demographics	SEX	Sex	C16576	C66731	F	Female	Female	A person who belongs to the sex that normally produces
DM	Demographics	SEX	Sex	C17998	C66731	U	U; UNK; Unknown	Unknown	Not known, not observed, not recorded, or refused. (NCI
DM	Demographics	SEX	Sex	C20197	C66731	M	Male	Male	A person who belongs to the sex that normally produces

phastar

Semantic Matching - SDTM

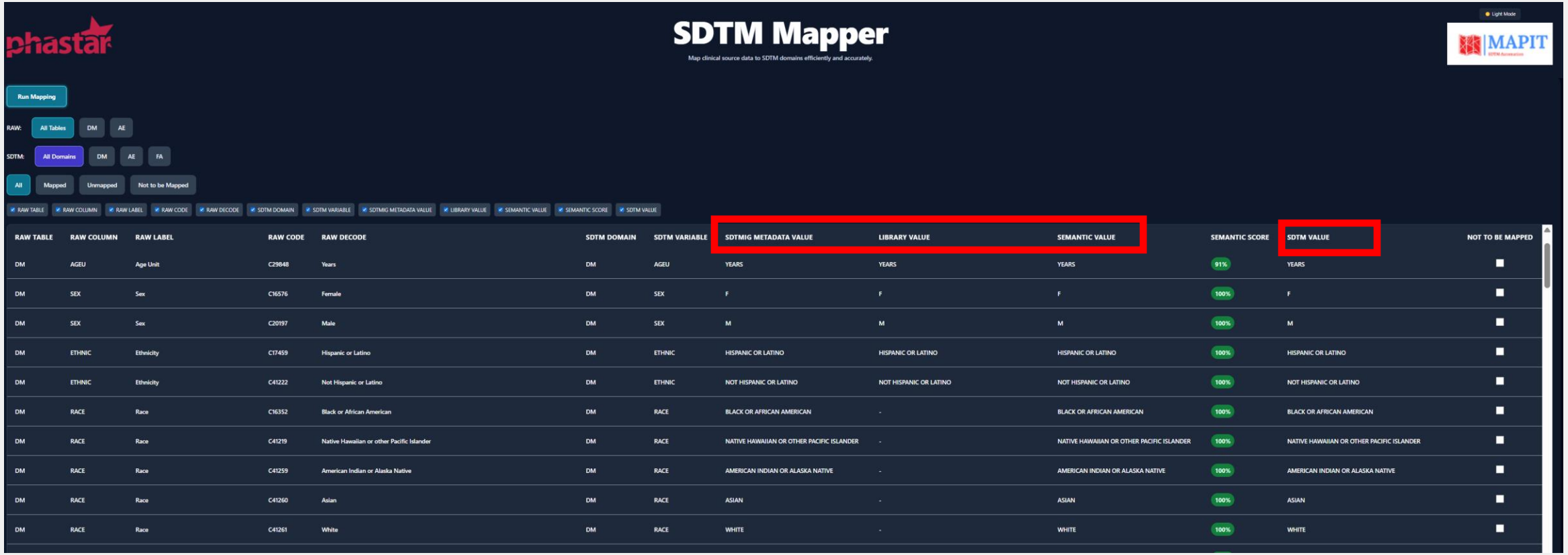

```
🔎 'Drug permanently discontinued' → Top 5 matches:
  ↳ 1. 'DRUG WITHDRAWN' (Drug Withdrawn) (score: 0.54)
  ↳ 2. 'DRUG INTERRUPTED' (Drug Interrupted) (score: 0.52)
  ↳ 3. 'DOSE NOT CHANGED' (Dose Not Changed) (score: 0.47)
  ↳ 4. 'DOSE REDUCED' (Dose Reduced) (score: 0.47)
  ↳ 5. 'UNKNOWN' (Unknown) (score: 0.47)
🔎 'Drug permanently discontinued' → Best match: 'DRUG WITHDRAWN' (score: 0.54)
```

```
🔎 'Asian' → Top 5 matches:
  ↳ 1. 'ASIAN' (Asian) (score: 1.00)
  ↳ 2. 'WHITE' (White) (score: 0.68)
  ↳ 3. 'BLACK OR AFRICAN AMERICAN' (Black or African American) (score: 0.64)
  ↳ 4. 'OTHER' (Other) (score: 0.63)
  ↳ 5. 'NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER' (Native Hawaiian or Other Pacific Islander) (score: 0.62)
🔎 'Asian' → Best match: 'ASIAN' (score: 1.00)
```

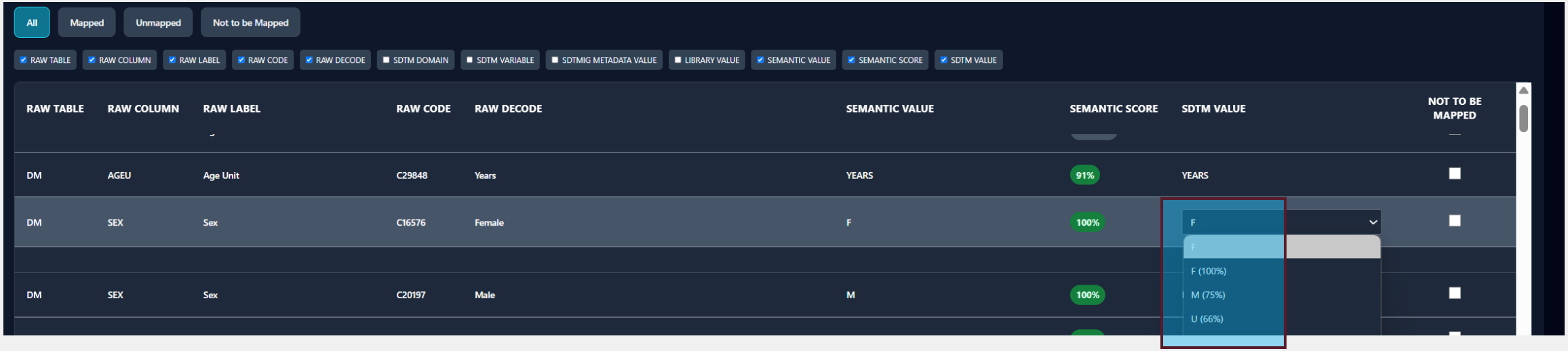
raw_table_name	raw_table	raw_colun	raw_label	raw_code_	decode	sdtm_domain	sdtm_variable	sdtm_value
Adverse Events (AI	AE	AESER	Serious AE	0	No	AE	AESER	N
Adverse Events (AI	AE	AESER	Serious AE	1	Yes	AE	AESER	Y
Adverse Events (AI	AE	AEACN	Action Taken, Investigational Product	0	Dose not changed	AE	AEACN	DOSE NOT CHANGED
Adverse Events (AI	AE	AEACN	Action Taken, Investigational Product	1	Dose increased	AE	AEACN	DOSE INCREASED
Adverse Events (AI	AE	AEACN	Action Taken, Investigational Product	2	Dose reduced	AE	AEACN	DOSE REDUCED
Adverse Events (AI	AE	AEACN	Action Taken, Investigational Product	3	Drug interrupted	AE	AEACN	DRUG INTERRUPTED
Adverse Events (AI	AE	AEACN	Action Taken, Investigational Product	4	Drug permanently discontinued	AE	AEACN	DRUG WITHDRAWN
Adverse Events (AI	AE	AEACN	Action Taken, Investigational Product	98	Not applicable	AE	AEACN	NOT APPLICABLE
Adverse Events (AI	AE	AEOUT	Outcome of AE	0	Recovered/resolved	AE	AEOUT	RECOVERED/RESOLVED
Adverse Events (AI	AE	AEOUT	Outcome of AE	1	Recovering/resolving	AE	AEOUT	RECOVERING/RESOLVING
Adverse Events (AI	AE	AEOUT	Outcome of AE	2	Recovered/resolved with sequelae	AE	AEOUT	RECOVERED/RESOLVED WITH SEQUEL
Adverse Events (AI	AE	AEOUT	Outcome of AE	3	Not recovered/not resolved	AE	AEOUT	NOT RECOVERED/NOT RESOLVED
Adverse Events (AI	AE	AEOUT	Outcome of AE	4	Fatal	AE	AEOUT	FATAL
Demographics (DM	DM	SEX	Sex	C16576	Female	DM	SEX	F
Demographics (DM	DM	SEX	Sex	C20197	Male	DM	SEX	M

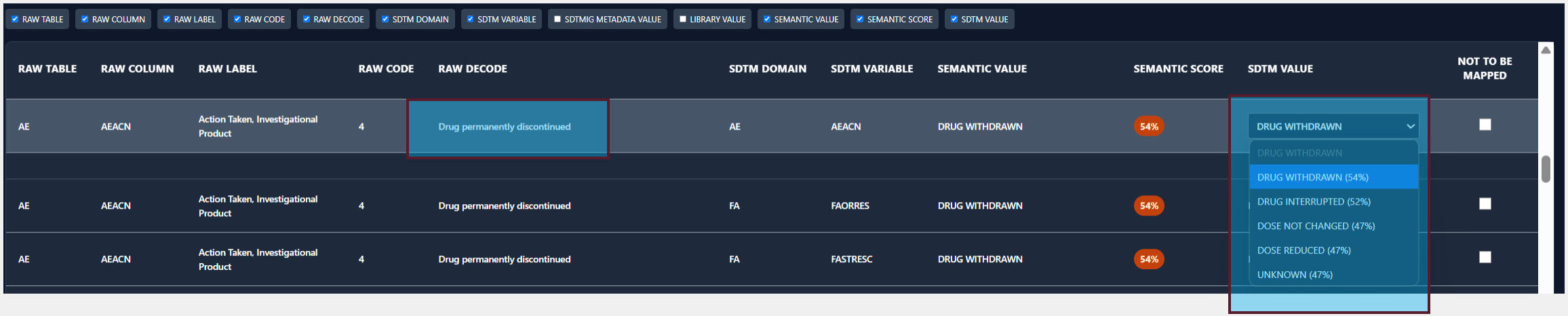
Semantic Matching - Output

- Need a practical way of interacting with AI models – inputting / extracting data
- Python web application can be an ideal ‘wrapper’
- Imagination is the only limit with open-source tools



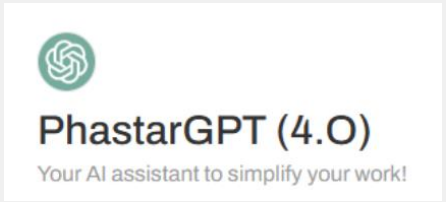
Semantic Matching - Output





Model choice and Training

- Examples are on pre-trained sentence transformers model
- Model is fast and training can enhance output but will have upper limit
- For deeper understanding and complex relationship need LLM
- Looking at balance between speed, cost, results



phastar

AI in SDTM Mapping Summary

- AI is complimentary, not a replacement for use of mapping standards
- AI options constantly evolving / improving
- AI vast integration options
- AI helps battle noise across EDC vendors – watch for USDM impact

Thank You

John.mcdade@phastar.com

phastar