

AUTOMATING SDTM

GET IT PREPARED WITH YOUR VOICE

GENPRO

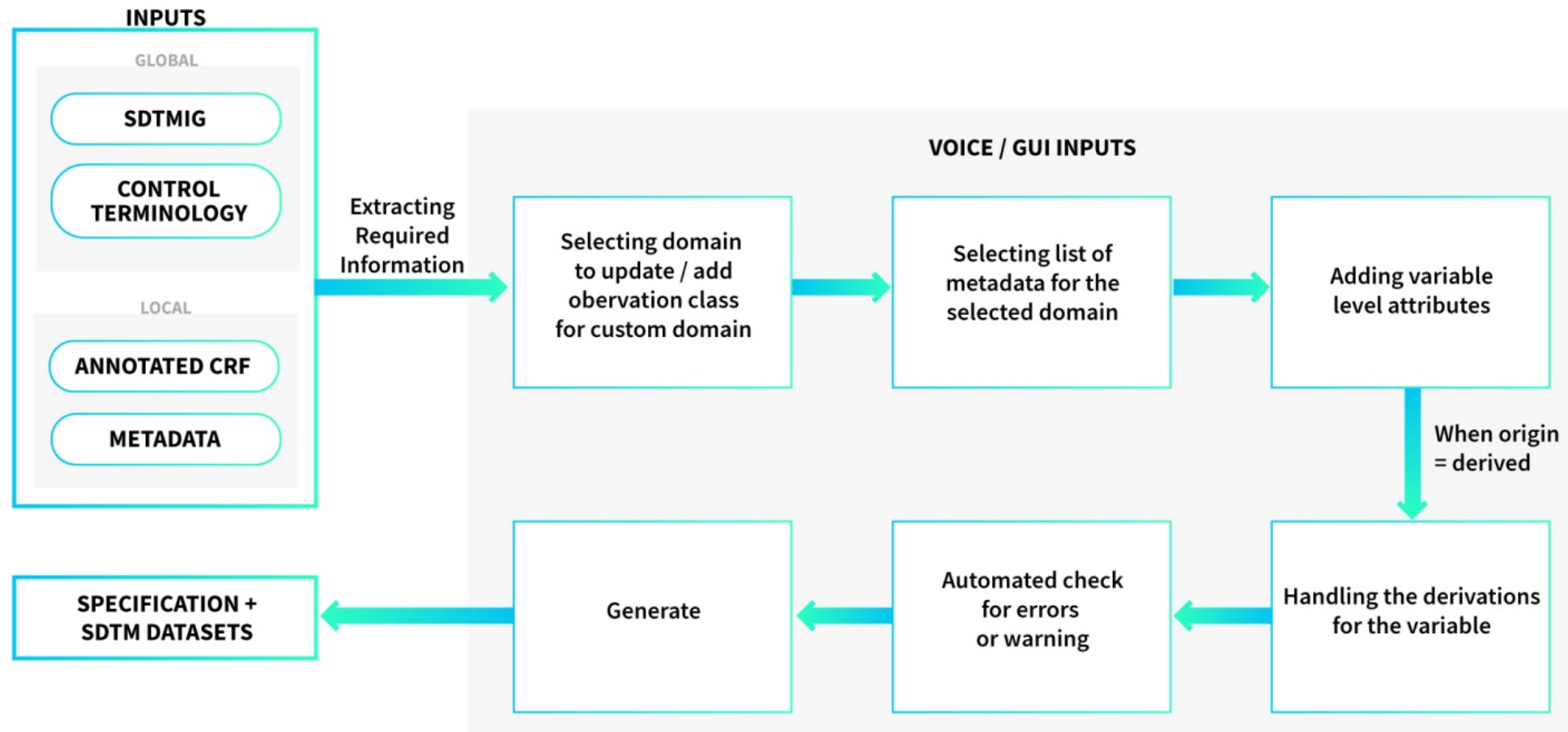
SDTM Challenges

- ▶ FDA mandates that data submissions for all studies must be in SDTM format
- ▶ Current generation cycle comprises significant human effort
- ▶ Cost and Quality issues are matters of concern
- ▶ Sponsors are looking for ways to generate the data sets in a cost effective process



YESDTM is a tool which semi-automatically generates the SDTM specifications and datasets using raw datasets, SDTM IG, SDTM control terminology and annotated CRF in a shorter time span using NLP and AI algorithms.

YESDTM workflow



Domain Extraction

YESDTM PDF reader augmented by our NLP engine will automatically extract the domains and associated variables using the Annotated CRF and the SDTM IG

YESDTM

Christohper

RequestsUsersStudy ManagementUser Level AdminsQueriesTask Management

Assign Task Here..

Domain Type	Domain Name	Title	Specs Author	Specification Target Completion Date	Specs Actual Completion Date	Status	Specs Reviewer
Events	AE	Adverse Events	Select User	Set Target Date	--/--	---	Select User
Interventions	EX	Exposure	Select User	Set Target Date	--/--	---	Select User
Special Domains	DM	Demographics	Select User	Set Target Date	--/--	---	Select User
Interventions	EC	Exposure as Collected	Select User	Set Target Date	--/--	---	Select User
Special Purpose	SV	Subject Visits	Select User	Set Target Date	--/--	---	Select User

ADD DOMAIN

Domain Creation

Domain Level information

- ▶ Inputs required from user
 - Description
 - Class
 - Structure
 - Key Variables
 - Repeating
 - Reference Data
- ▶ YESDTM will automatically populate this next time onwards

Domain Creation

YESDTM Will automatically populate fields like Variables, label, data type, length, mandatory and role for standard domains

Order	Variable	Label	Data Type	Length	Significant Digits	Mandatory	Codelist	Origin	Pages	Method	Predecessor	Role	Comment
1	STUDYID	Study Identifier	text			Yes		Assigned ▼				Identifier	<button>ADD</button>
2	DOMAIN	Domain Abbreviation	text			Yes	DOMAIN	Assigned ▼				Identifier	
3	USUBJID	Unique Subject Identifier	text			Yes		Derived ▼		<button>ADD</button>		Identifier	
4	VISITNUM	Visit Number	float	2		Yes	VISITNUM	Assigned ▼				Topic	<button>ADD</button>
5	VISIT	Visit Name	text			No	VISIT	Assigned ▼				Timing	<button>ADD</button> REMOVE
6	VISITDY	Planned Study Day of Visit	integer			No		Derived ▼		<button>ADD</button>		Timing	REMOVE
7	EPOCH	Epoch	text			No	EPOCH	Derived ▼		<button>ADD</button>		Timing	REMOVE
8	SVSTDTC	Start Date/Time of Visit	text			No	ISO8601	CRF ▼	24			Timing	<button>ADD</button> REMOVE
9	SVENDTC	End Date/Time of Visit	text			No	ISO8601	CRF ▼	24			Timing	<button>ADD</button> REMOVE
10	SVSTDY	Study Day of Start of Visit	integer			No		▼				Timing	REMOVE

ADD VARIABLEVALIDATEGENERATE

Domain Creation

Automation Logic

- ▶ For all the data types, populated as 'text', YESDTM will use the maximum length from the variable and populate that as the length.
- ▶ "Core" column of the domain models in the SDTM IG is used to populate the field 'Mandatory'
- ▶ After the generation of the dataset, YESDTM will compare the data value with the selected control terminology value based on the codelist name to populate code list

Domain Creation

Handle Derivations of the variable

Example for comment/Derivation:

Comment:

Set to 'MULTIPLE' when RAW.DM.XX="Yes" Where XX refers to RACEAM, RACEAS and RACEBL;

Set to 'AMERICAN INDIAN OR ALASKA NATIVE' when RAW.DM.RACEAM="Yes";

Set to 'ASIAN' when RAW.DM.RACEAS="Yes";

Set to 'BLACK OR AFRICAN AMERICAN' when RAW.DM.RACEBL="Yes";

Derivation:

```
Data dmrace;  
  Length race $200;  
  Set raw.dm;  
  If strip(raceam)="yes" and strip(raceas)="yes"  
    and strip(racebl)="yes" then race="MULTIPLE";  
  else if strip(raceam)="yes" then race="AMERICAN INDIAN OR ALASKA NATIVE";  
  else if strip(raceas)="yes" then race="ASIAN";  
  else if strip(racebl)="yes" then race="BLACK OR AFRICAN AMERICAN";  
  keep subject race;  
run;
```

Custom Domains

- ▶ How YESDTM identifies custom domains
 - When a domain is identified from the CRF which is not present in SDTM IG, YESDTM will identify it as a custom domain.
- ▶ Additional inputs required from user for domain Mapping
 - Domain Level information
 - List of variables
 - Labels

SAS/Pinnacle Interface

- ▶ SAS Interface to do the derivations
 - Currently SAS code will be run through the command prompt
 - The log, SAS and XPT dataset will be the output for each domain.
 - In the future, YESDTM will be integrated with SAS VIA APIs
- ▶ Pinnacle 21 interface will be used to do the validations
 - Pinnacle API is a RESTful service
 - The developer can validate the specification with the help of pinnacle.

Outputs

- ▶ Once the specification is finalized user can generate
 - SAS datasets
 - XPT datasets
 - Define XML
 - Pinnacle Report

Visualization

- ▶ MOSS Interface will be used to do the Visualization
 - MOSS is a dynamic software framework for NLP based data visualization
 - Intended to eliminate the dependency on the data team
 - Allow clinicians to analyze trial data independently
 - Supports a variety of Table and Graph presentation formats for the users

REFERENCE

- ▶ LAWRENCE, R. (2008). FUNDAMENTALS OF SPEECH RECOGNITION. PEARSON EDUCATION INDIA.
- ▶ WOOD, F. (2008). THE CDISC STUDY DATA TABULATION MODEL (SDTM): HISTORY, PERSPECTIVE, AND BASICS. PHARMASUG PROCEEDINGS, JUNE.
- ▶ [HTTPS://MOOS.CLAPRESEARCH.COM/](https://moos.clapresearch.com/)

THANK YOU