WITH STANDARDS – UNLOCK THE POWER OF DATA

cdisc

2022
US
INTERCHANGE
26-27 OCTOBER | AUSTIN

Enable 10X Efficiency to Your Clinical Data Standardization and SDTM Submission Processes

Presented by Srinivasan Anandakumar, Head of Product Management, Saama

Meet the Speaker

Srinivasan Anandakumar

Title: Head of Product Management

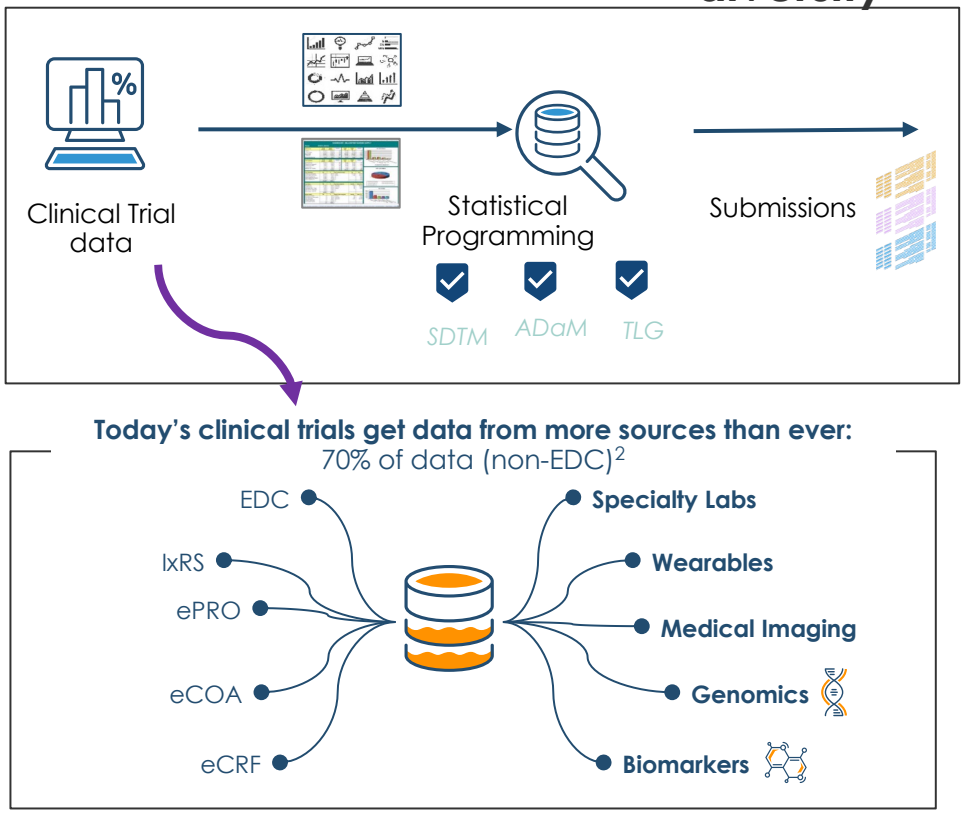
Organization: Saama

Srini heads the Product management at Saama for a portfolio of products that accelerate clinical data management and submissions. He is also responsible for clinical trials data platform strategy that enables both analytics and submission pathways. Srini has 20+ years of clinical trials process and technology experience and is passionate about building next generation products to accelerate clinical trials.

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.

Clinical submission process efficiency is driven by data volume and diversity



929,203

Data points per clinical trial in 2015[1], **up 88%** from 2005 *

10 Million*

60% of trials have such data points

- **Increase in time from** LPLV to database lock (2017-2020)[3] **- 32%**
- **Avg time to standardize a non-standard study – 35 days****

1. *CISCRP and Tufts Center for the Study of Drug Development, "Industry Practices That Will Benefit by Patient Engagement," by Ken Getz, October 2017*
2. *https://www.pharmavoice.com/wp-content/uploads/Opportunities-to-Accelerate-Decentralized-and-Digital-Trials.pdf?tracker_id=1630241131*
3. *https://www.eclinicalsol.com/newsroom/tufts-csdd-eclinical-solutions-clinical-data-study-shows-32-percent-increase-in-trial-cycle-time-metric*

- *Based on sample analysis of data in Saama's platform*
- *Based on baselining done by Saama before solution implementation, medium size trial*

Current challenges in SDTM generation

Maintenance of Global Standards
User: Global Librarian

Study environment setup
User: SDTM Programmer, Biostats & programming IT Analyst

Study programming
User: SDTM Programmer

Data extracts
User: SDTM Programmer, Biostats & programming IT Analyst

Standards maintenance
CRF Metadata, Third Party Metadata, Code Lists

Macros maintenance
Global Macros Maintenance

Study documents review
Protocol, CRF's, Data Management Plan, etc.

Source data acquisition
Source System Data Flow Setup

Study transformation Metadata management
Global Macros Copy

Custom programming & dry run
Custom domains, Macros, Quality checks

Re-programming
Re-programming due to Metadata changes

Blinding & un-blinding data extraction
Dummy and real datasets for data transformation

QC & submission package generation
(P21 checks, define.XML, SDRG, aCRF)

Challenges

- **Manual process** in maintaining the **global macros library**
- **Higher lead time** in setting **up data and metadata for SDTM generation**
- **Significant effort and lead time in SDTM generation, QC** for both standardized and non-standardized raw data sets
- **Lineage and data traceability**

Solution Requirements

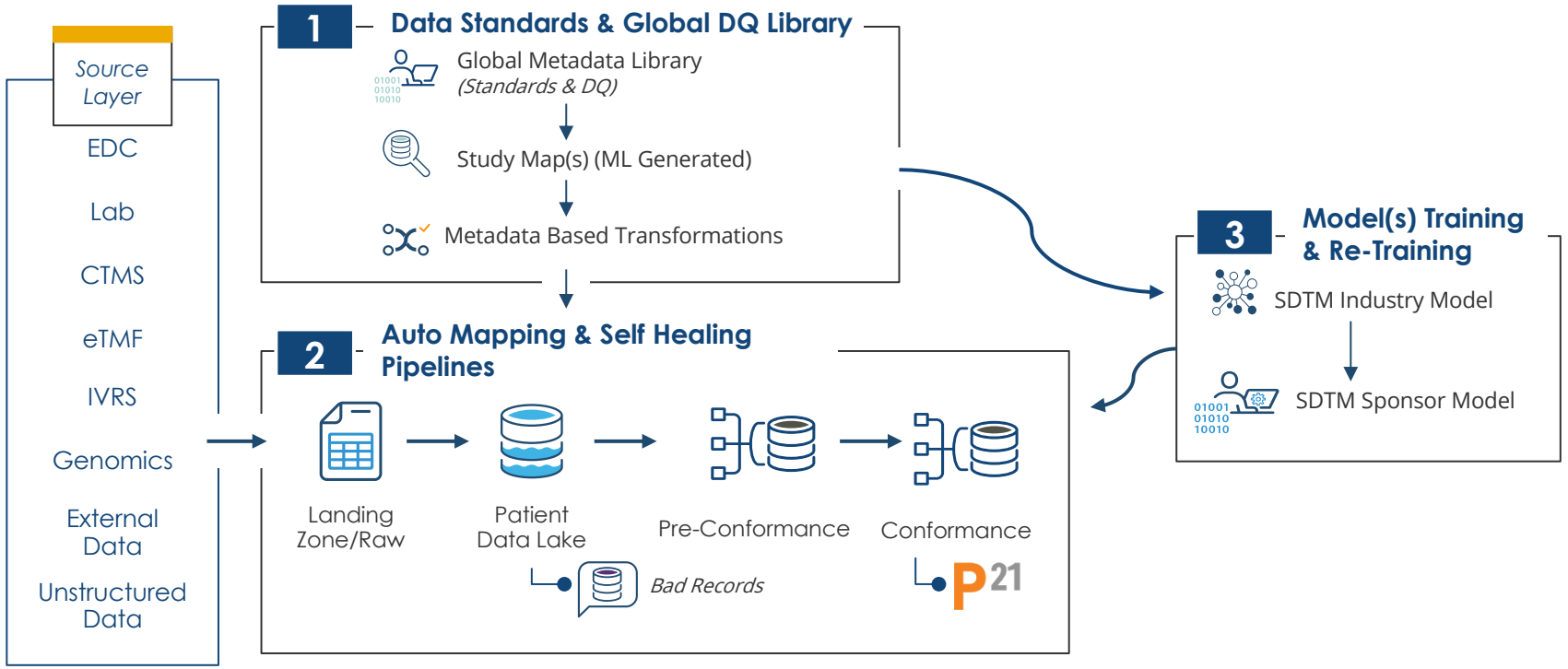
- **Faster source system Set up including Transformation metadata et up** enabled by self service Global Library, data standards & low code platform
- **On demand execution** of data standardization process with mechanism to handle **Metadata changes**
- Reduced effort in QC, **submission package generation and data provisioning (Blinded/Un-blinded)**

Process: Spreadsheets Manual Processes Custom Scripts

Building blocks for solution – SDTM generation

	PEOPLE Clinical Programmer □ Clinical Data Scientist	PROCESS SOPs on Human-Algorithm Interface	PRODUCT Composite Architecture
Data Standards & Global Transformations Library	• **Data Standards training** – Global Standards & Transformation Metadata • Transition to **Augmented Study Set up** • Custom functions Programming – **Low Code** experience	• **Standards maintenance process** – Global Standards & Transformation Metadata • SOP & Process – **automated study set up** • **Promotion Process** – Study Variances	• **Standards Library –** Data element, Contextual and Transformation Metadata • **Global Standards Library** – Promote, demote, Propagate • **Wizard** based Study Set up
Auto Mapping & Self Healing Pipelines	• Change management to act on **intelligent pipelines** output • Training on **team collaboration & workflow** using system features • Enablement on **Operational Metrics continuous Improvement**	• **SOP & Process on Study conduct** impact (e.g. Metadata changes) • SOP & process on **monitoring Transformation Jobs & pipelines** • Monitoring & **CAPA on Operational Metrics, SDTM quality process**	• **Hybrid Cloud** & Source System readiness • **SDTM Transformation Execution** Engine • **Intelligent self – healing pipelines** and dynamic job management framework • Adaptive Task & workflow Engine
Model(s) Training & Re-training	• "**Human in the Loop**" training & re-training for Model output Review • Change Management to **trust ML/AI** Algorithms • Training & enablement to act on **Smart Suggestions**	• SOP& Process on **User feedback** – Sponsor model, Study Implementation • SOP& Process on **Sponsor SDTM Model training/Re-training** • SOP & Process on ML/AI training – **Industry SDTM model**	• **Model Performance Management** – Continuous tracking with thresholds • **Automated Model training/re-training** & deployment framework • Federated ML/AI Architecture

Data flow view representation – SDTM generation (Metadata & AI Driven)



ML/AI based mapping – An Example

1 Source Data

STDAT	STPDAT	SER	SEV	ACN	OUT
9-11-19	9-11-19	No	Mild	NA	RECOVERED/RESOLVED
7-11-19	7-11-19	No	Moderate	NA	RECOVERED/RESOLVED
5-11-19	5-11-19	No	Moderate	NA	RECOVERED/RESOLVED

2 Source Metadata

Name	Type	Length	Label
STDAT	Char	20	start date of AE
STPDAT	Char	20	End date of AE
SER	Char	2	Serious event

- Sources for Model Inference
- Components for Model Training/Re-training
- Model Inference Components

11. SME Input

3 Codelists, Annotated CRFs, Target Standards *(Sponsor)*, Historical Study Maps

5 Input Sources

6. Domain Class Prediction

7. Domain Prediction

8. Column mapping Prediction

9. Transformation Function and Parameter Prediction

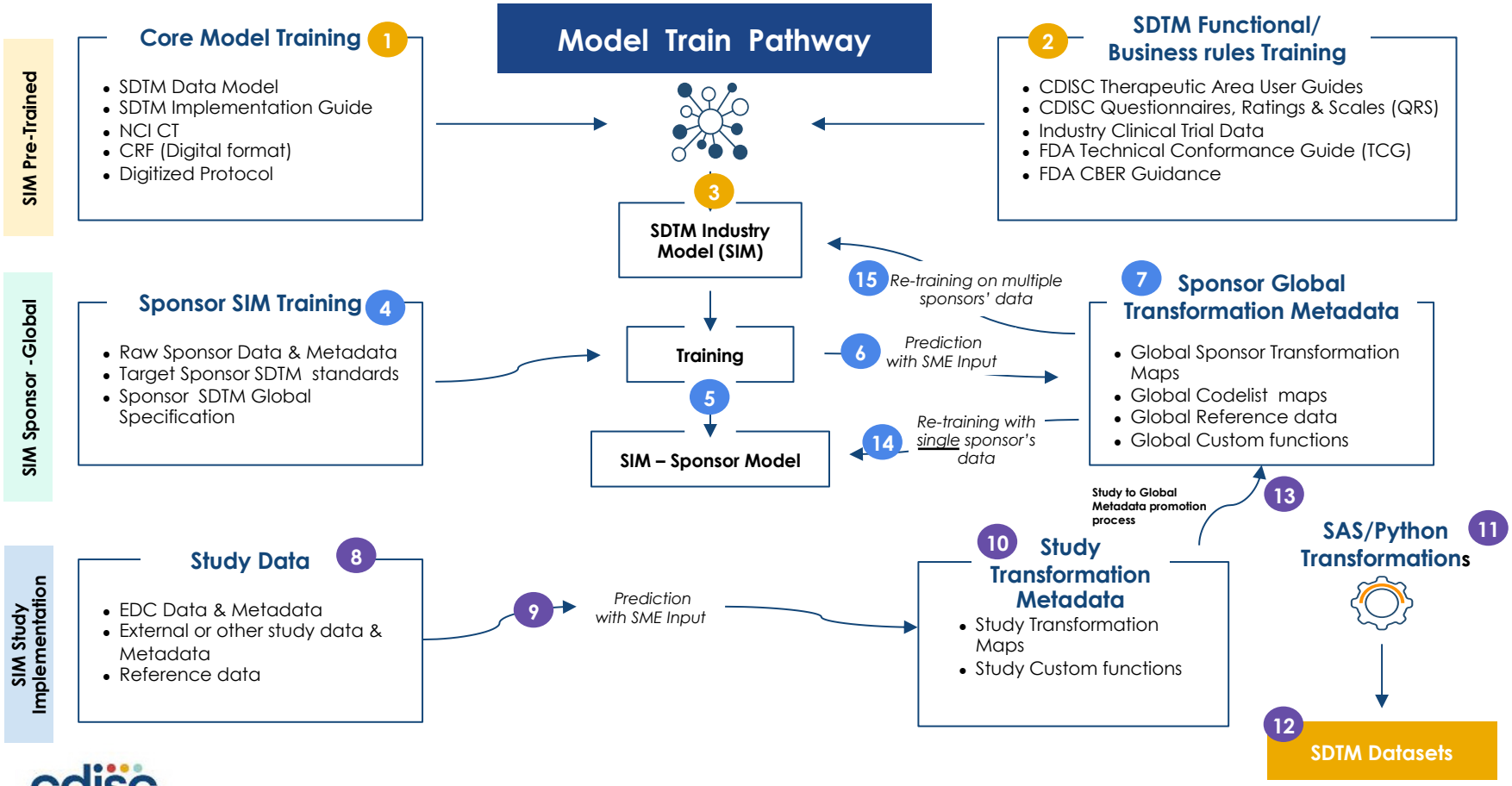
4 Pre-trained ML /AI SIM Model

12 Model Re-Training

10 Metadata Mapping *(ML Generated)*

CLASS	DOMAIN	SOURCE VARIABLE	TARGET VARIABLE	FUNCTION	PARAMETERS
Events	AE	STDAT	AESTDTC	date_convert	{format:"dd-mm-yy"}
Events	AE	STPDAT	AEENDTC	date_convert	{format:"dd-mm-yy"}

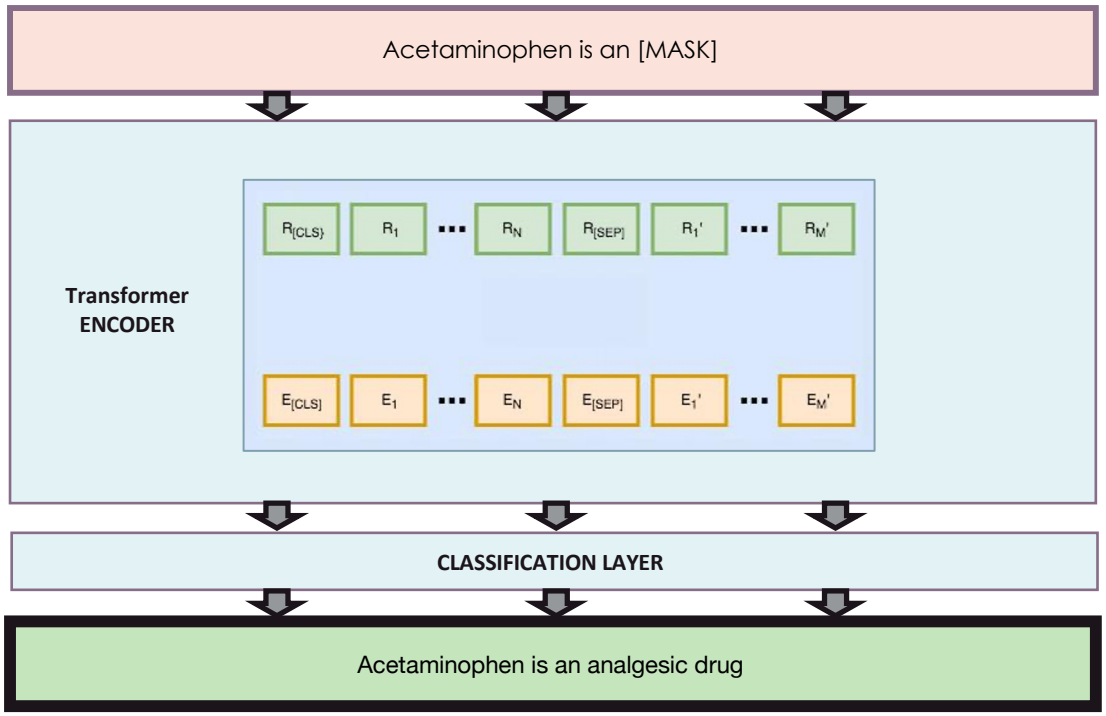
Training & re-training the pre-trained model



Language model pre-training on Pubmed abstracts and full text

Technical details:

- The models are Deep learning based NLP Models
- RoBERTa / BERT flavor of transformer architectures are used as base models
- The models were pre-trained with Masked language model objective on Pubmed abstracts and full text



Masked Language model (MLM) Training:

- MLM objective helps the model with Biomedical domain understanding which is leveraged in the next stages of training (task specific training).
- **Example**: Understanding that Acetaminophen is a drug could help the model to predict that the column could map to Interventions class (in class prediction task) , CM domain (in domain prediction task) and CMTRT column (in column prediction task)

Business metrics – Industry baseline & next generation solution

Category	Metric	Industry Baseline*	ML/AI solution**	Efficiency Improvement
Schedule	EDC Go-Live/FPFV to 1st draft SDTM Development	**30 to 35 Days**	5 to 10 days	6X
	SDTM refresh and validation (for key study milestones)	**5 to 7 days**	2 to 3 days	3X
	DBL to un-blinded SDTM datasets (End of Study)	**2 to 3 days**	1 to 2 days	2X
	Submission Package	**4 to 5 weeks**	1 to 2 weeks	4X
Effort	# Effort for EDC Go-Live/FPFV to 1st draft SDTM Development	**800 to 1000 hrs**	80 to 100 hrs	10X
	# SDTM programmer (FTE) per study	**3 FTE**	1 FTE	3X

- *Based on baselining done by Saama before solution implementation - medium size trial*

Summary

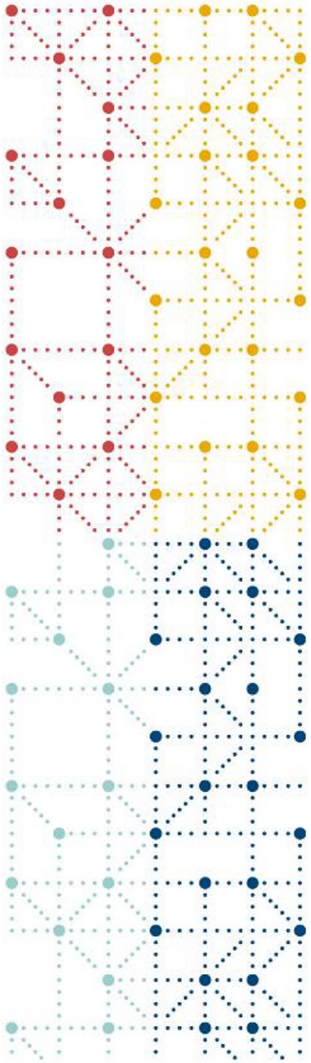
Significant increase in clinical trial data sources will continue to increase the data management and data standardization effort and lead time.

Existing process and systems – with silos & manual elements tends to compound the challenges, thus affecting Business outcomes.

Data standards & global transformation library, auto mapping & self – healing pipelines, model (s) training & re-training are building block concept streams for next generation clinical data quality solution

Continuous learning systems – with "Human in the loop" enables higher quality, effort and schedule reduction in Submission (SDTM generation) Process

Standardized metrics across schedule and effort can measure the improvement of the efficiency in the SDTM generation process. These metrics show **2x to 10x potential increase** in efficiency when the process is enabled by a smart ML/AI solution.



Thank You!

cdisc