cdisc

2023
US
INTERCHANGE
FALLS CHURCH, VA | 18-19 OCTOBER

Digital Transformation – Will AI Be Our New Best Friend in Creating CDISC Compliant Workflows?

Michelle Lumicao
Principal Data Standards Specialist
Syneos Health

Meet the Speaker

Michelle Lumicao

Title: Principal Data Standards Specialist

Organization: Syneos Health

Michelle Lumicao is a Principal Data Standards Specialist at Syneos Health, where she is a member of the Data Standards and Governance group as an SDTM SME. She has more than 17 years of experience in clinical research. She worked in Data Management for 4 years and in Statistical Programming 13 years. She has been a CDISC volunteer for CORE and the Digital Health Technologies Team.

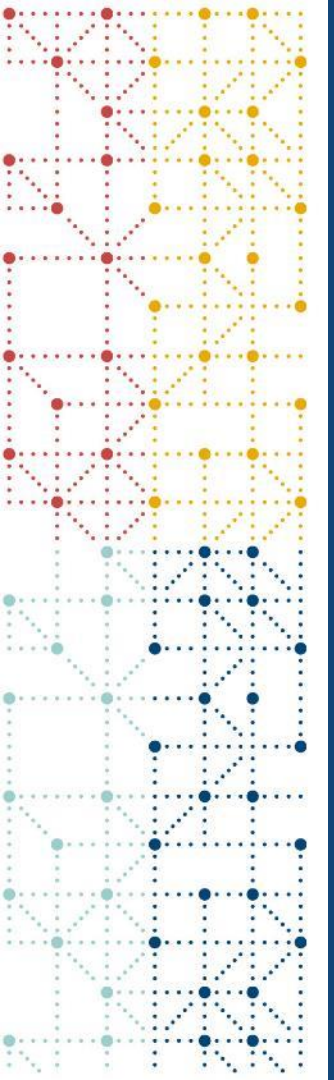
Disclaimer and Disclosures

- *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.*

- *The author has no real or apparent conflicts of interest to report.*

Agenda

1. Current Challenges for CDISC Workflows
2. Our Experience with Some Technological Solutions
   - Metadata Repositories (MDR)
   - AI/ML SDTM Transformation
3. Assessment of a New Tool

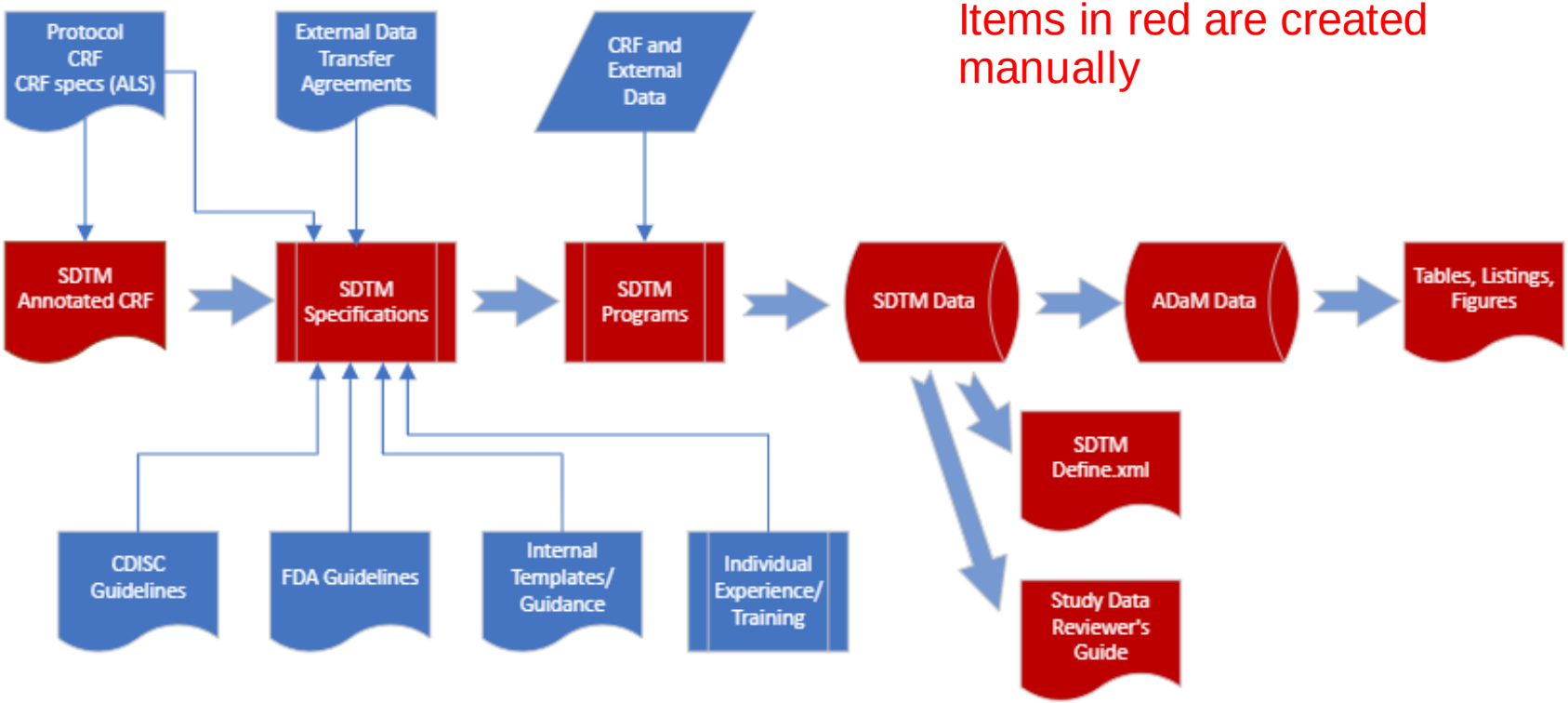


Current Challenges for CDISC Workflows

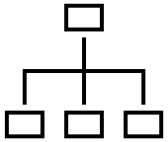
How have we traditionally created CDISC outputs?

Traditional Process for Creating CDISC Compliant Deliverables for SDTM

Items in red are created manually

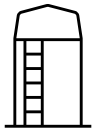


Challenges in Completely Manual Process



- Difficult to implement governance of data standards
  - Different level of SDTM expertise in SDTM programmers
  - Different experiences in previous studies
  - New versions of guidance, best practices, controlled terminology

- Increase in external data sources:
  - Digital Health Technology (e.g., wearables, eDiary, medical devices)
  - Real World Evidence (e.g., insurance claims data, EHR)
  - Decentralized Clinical Trials
  - Biomarkers, Genetic Information



- Difficult to promote efficiency by using existing tools
  - Silos in lessons learned by individuals/studies

- Mistakes in SDTM mapping can trickle down to ADaM and TLFs
  - Rework in SDTM can lead to delays and budget overburn

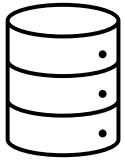
A technological solution for a CRO must be able to:

- Create submission-ready, compliant outputs
- Allow governance of data standards
- Decrease the time to create deliverables
- Increase reusable transformations into CDISC standards
- Remove silos of knowledge (clear traceability of downstream outputs)
- Be easy to use in order to promote tool adoption
- Be flexible enough for multiple TA and sponsor requirements

Technological Solutions

Metadata Repository, AI/ML Enabled SDTM Transformations

Metadata Repository



- What is metadata?
  - Data about data. It describes the format and content of the data for a study, including datasets, variables, value-level metadata, and codelists.
  - Usually in Excel specs. For example:

CRF Specs (ALS) metadata

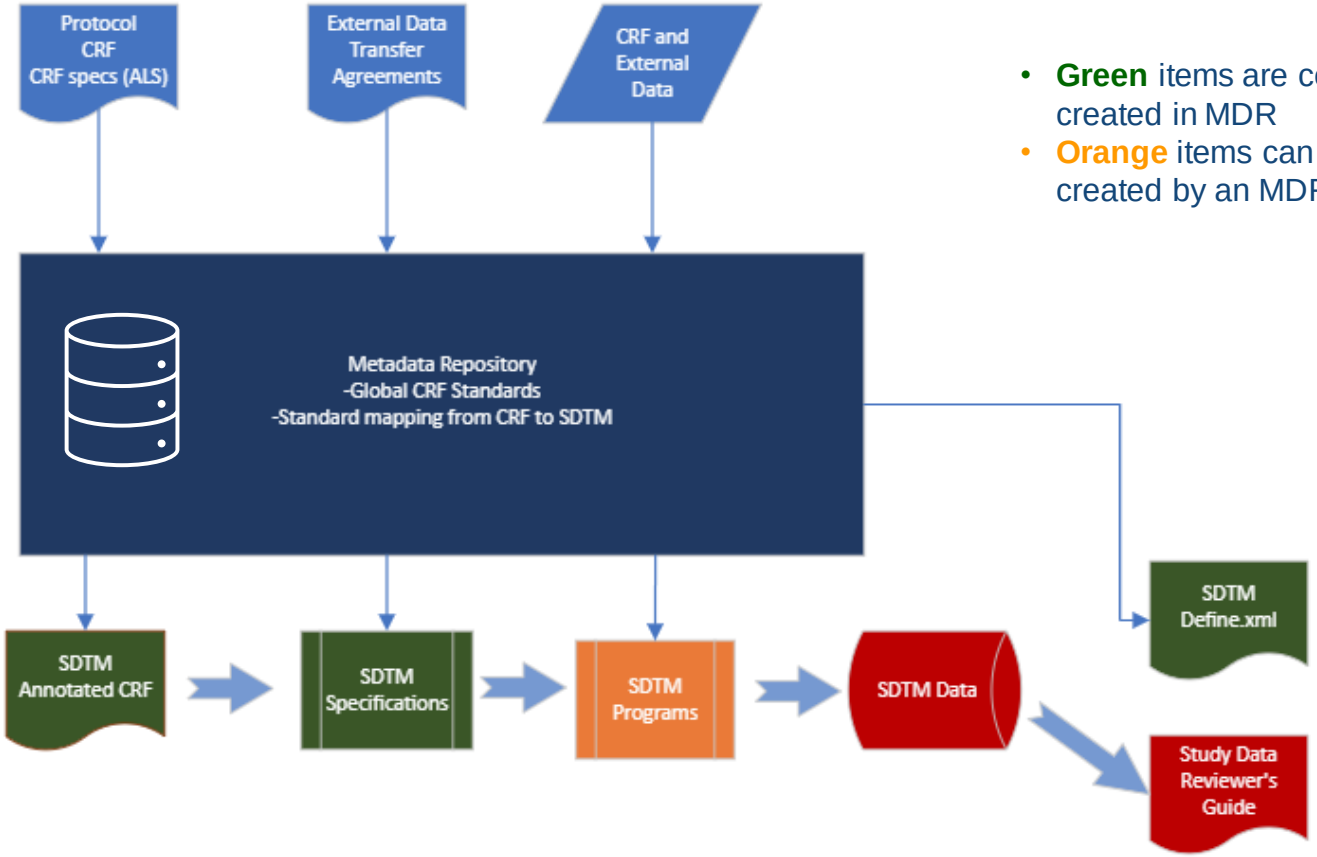
Form OID	FieldOID	DataFormat	DataDictionary Name	ControlType	PreText
VS	VSWT	5.2		Text	Weight
VS	VSWTU	$2	BODYMASSUNIT	RadioButton (Vertical)	Weight Unit
VS	VSBPSYS	3		Text	Blood Pressure - Systolic
VS	VSBPSYSU	$4		Text	Blood Pressure - Systolic Unit
VS	VSBPDIA	3		Text	Blood Pressure - Diastolic
VS	VSBPDIAU	$4		Text	Blood Pressure - Diastolic Unit
VS	VSPULSE	3		Text	Pulse Rate

SDTM metadata

Dataset	Variable	Label	Data Type	Length	Format	Codelist Keyword	Origin
VS	STUDYID	Study Identifier	text	7			Protocol
VS	DOMAIN	Domain Abbreviation	text	2		DOMAIN	Assigned
VS	USUBJID	Unique Subject Identifier	text	17			Derived
VS	VSSEQ	Sequence Number	integer	8	BEST.		Derived
VS	VSTESTCD	Vital Signs Test Short Name	text	8		VSTESTCD	Assigned

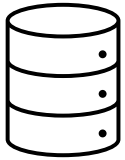
- A Metadata Repository (MDR) is a database created to store metadata

Standardized Workflow with an MDR



- **Green** items are centrally created in MDR
- **Orange** items can potentially be created by an MDR

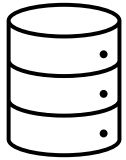
Implementing MDR



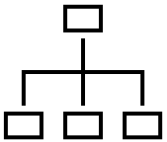
- Select TA
- Create Standard CRFs
- Input all metadata for each asset.
- Assets are content components which include:

	Example: Domain	Example: Variable	Example: Value Level
Identification Information	Name = AE Version = Study ABC SDTM IG 3.2	Name = AEACN Version = Study ABC SDTM IG 3.2	Name = VLM.ACN Version = Study ABC SDTM IG 3.2
Classification Information	Object Type = Domain Class = Events	Label = Action Taken with Study Treatment Type = Text Core = Exp	
Artifact Information		Computation Method	
Relationship Information	Contains = AESER, AEACN	Contained By = AE Contains Terminologies = ACN	Comprised of Values = DOSE INCREASED, DOSE REDUCED, DOSE NOT CHANGED Maps to Variables = AEACN

Advantages of an MDR



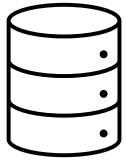
- Efficiency
  - No need to re-invent the wheel with every study

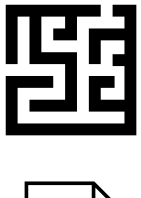


- Better Quality
  - Governance of data standards
  - More consistently implement new SDTM IG changes

- Time Saved
  - Only need to update if there are study-specific domains or variables that are not included in the global standards

Challenges/Limitations of an MDR



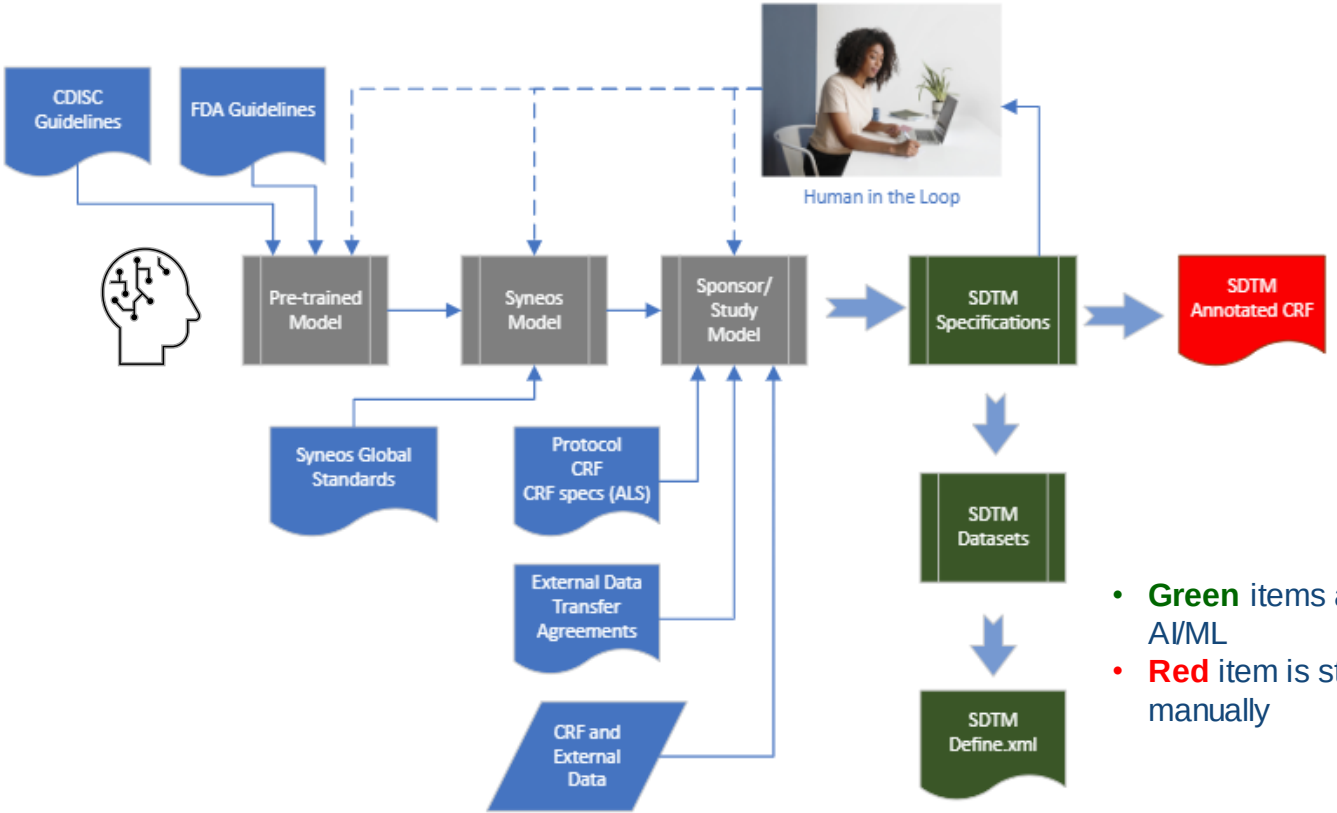


- Takes a lot of time to set up Global Standards
- System latency can be very slow depending on location
- Dependence on external resource to fix MDR issues
- Easier to update Excel than to make updates in MDR
- Generally, only standardizes data from CRFs

AI/ML SDTM Transformation

- We have received requests from sponsors to use AI/ML
- AI/ML tools are available to assist in the mapping of raw datasets and external data into SDTM.
- A pre-trained AI/ML model can predict the following:
  - Domain
  - Variable
  - Transformation function
- There are out of the box transformation functions available (e.g., convert to ISO date, applying codelists)
- Option to manually create custom transformation code using SAS/Python which can be reused in the future

Workflow using AI/ML SDTM Transformation



- **Green** items are created by AI/ML
- **Red** item is still created manually

Advantages of AI/ML

- Relatively quicker to implement than MDR
  - No need to standardize everything at the start
- Time Saved
  - Each prediction comes with a confidence score. Human in the loop can focus on the predictions with low reliability score.

- Better Quality
  - Using previously validated transformations should lead to better quality and less time needed for validation.

- Continued improvement in predictions
  - Feedback on predictions will improve the model

- Predicted mapping for a variety of external data (not just CRF)

Challenges/Limitations of AI/ML

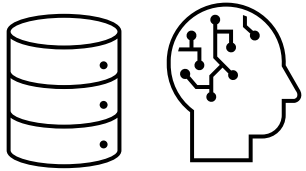
- Unknown reliability of predictions
  - Haven't had the chance to review the out of the box prediction of the pre-trained model
- New external data sources
  - AI/ML works best when the new data is similar to training data

- No portability of improved model based on our feedback
  - If we decide to switch to a different vendor, we would need to retrain the new system

- Still some manual creation of TDM and SDTM aCRF
- Does not create SAS program

- More expensive than MDR
  - Charged per study

Hybrid MDR and AI/ML

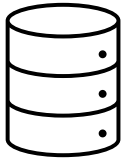


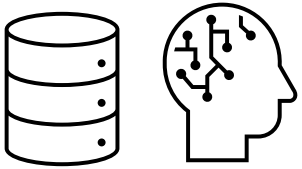
- Some new tools have components of both MDR and AI/ML
- In addition to advantages from traditional MDR, additional features:
  - AI/ML reads in protocol and partially create the TDM datasets
  - Creates dummy EDC data
  - Makes SDTM mapping suggestions
  - Creates SDTM aCRF
  - Creates SAS program
- We have not had as much time to assess the Hybrid MDR with AI/ML capabilities.

Assessment of a New Tool

How do we select the right tool for our process?

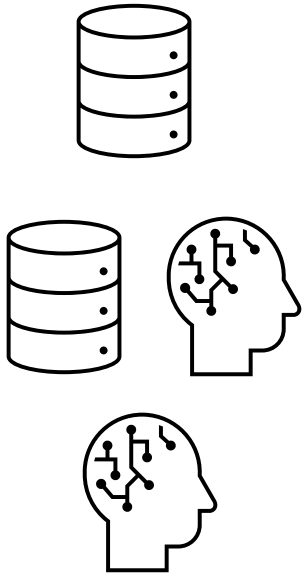
How much time will we save?





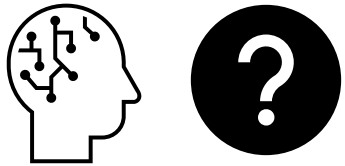
- MDR
  - Re-using standards will decrease the time to create SDTM specs, aCRF, and SDTM datasets from CRF data.
  - Define.xml can be easily created from the metadata stored in the MDR
- Hybrid MDR with AI/ML will additionally improve:
  - More quickly create the TDM from protocol
  - Creates SAS programs reflecting the SDTM specs and transformation functions
- AI/ML SDTM Transformation
  - SDTM specs will be much easier to create, even if no pre-defined standards for non-CRF inputs

How quickly can we implement?



- MDR will take the longest to set up
  - Select TA or subset of CRFs to create standards
  - Enter metadata in MDR
- Hybrid MDR with AI/ML will take less time
  - AI/ML can auto-populate some of the metadata and assist in transformation functions to convert raw datasets to SDTM
  - If only using AI/ML SDTM Transformation, it will take less time to set up
- AI/ML SDTM Transformation will be relatively quicker to set up
  - Pre-trained model will be able to ingest CRF specs and raw datasets to predict SDTM transformation

Unknowns with AI/ML



- How well will AI/ML predictions work for non-CRF data?
- How accurate is the confidence score for each prediction?
  - If the model encounters new sources of data that were not similar to training data, will that be flagged with a low confidence score?
- How long will it take to improve the pre-trained model?
- How much validation is needed for outputs?
  - What is the risk if double programming is not done?

General Impressions Based on Our Experience So Far

	MDR	AI/ML SDTM Transformation	Hybrid MDR and AI/ML
Time to initial set up	Longest	Fairly quick	Long for MDR, quicker for AI/ML only
Time to create TDM	N/A	N/A	Quickest
Time to create SDTM specs	Moderate improvement	Quickest	Quick
Time to create SDTM aCRF	Moderate improvement	N/A	Quick
Time to create SDTM datasets	Moderate improvement	Quickest	Quick
Time to create define.xml	Quickest	N/A	Quick
Time to validate outputs	Moderate improvement	Quickest	Quick
Additional manual work	Moderate	Least	Moderate
Rework Due to Quality	Lowest Risk	Moderate Risk	Low Risk
Governance of Data Standards	Excellent	Moderate	Very good
Cost of tool	Least expensive (annual)	Most expensive (per study)	Less expensive (per study)
Improved traceability and KPI measurement	Excellent	Excellent	Excellent

Thank You!