



Single Day Event

An era when Data Science harnesses Data Standards

Sudharsan Dhanavel
Logesh Umapathi

Saama Technologies

PHUSE India Winter SDE (Chennai)

Hybrid Event, Saturday 10 December 2022

phuse.global



Disclaimer

The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the views of Saama.

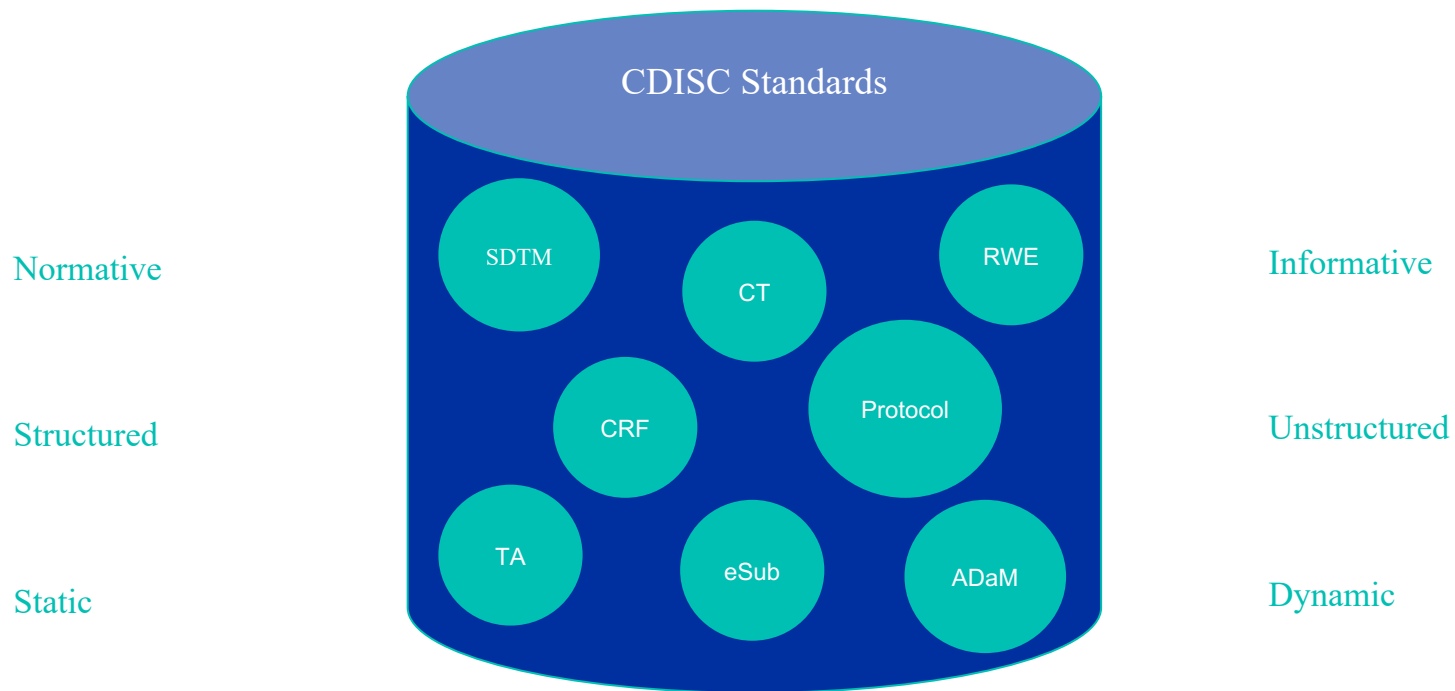


Agenda

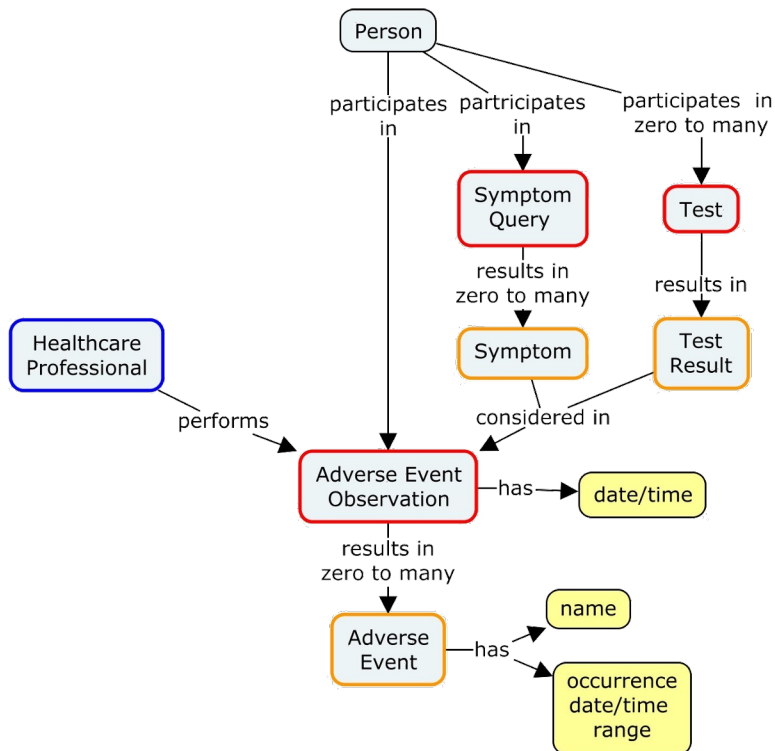
- Introduction
- Concept maps
- Knowledge, Concept & relationship
- Concept extraction
 - Entity recognition and normalization
 - Entity linking
 - Relationship extraction
- Standards ingested model
- Demo
- Conclusion



Introduction

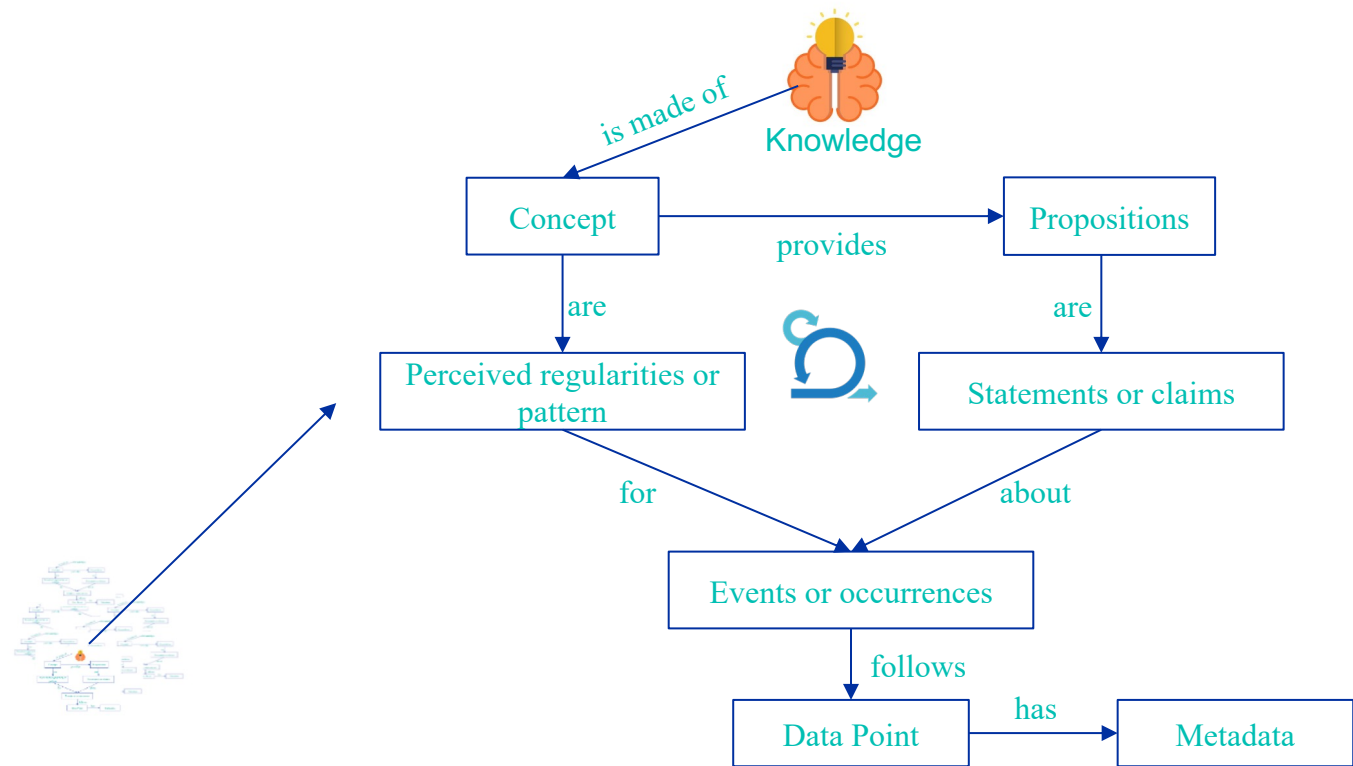


What is a Concept map?



- Knowledge representation tool
- Semantics of natural languages
- Simplify complex topics
- See the bigger picture
- Understand relationship

Knowledge, Concept & relationship...



Cognitive Learning

Compute

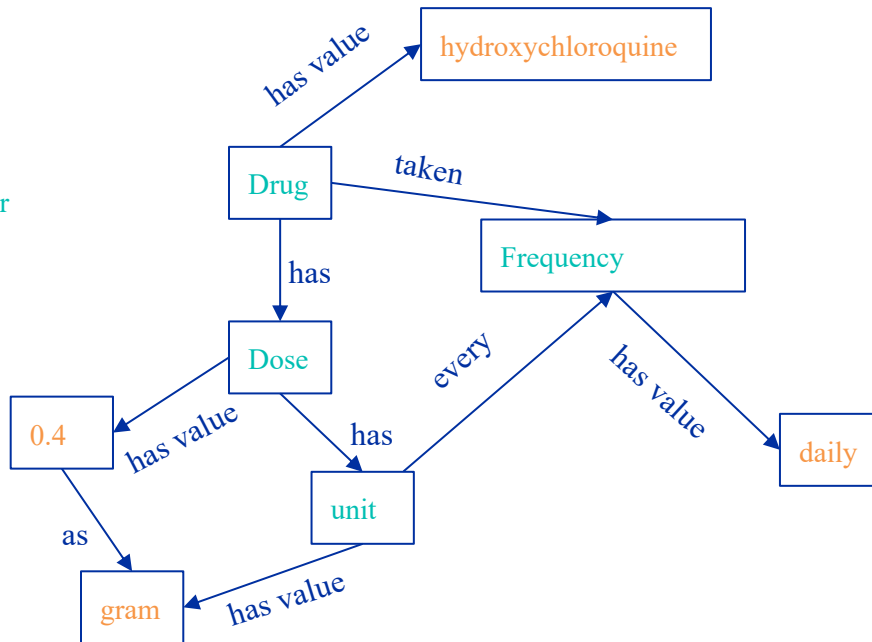
Classify

Validate



Entity recognition & normalization

- 400 mg hydroxychloroquine intervention per day
- Daily dose of 400mg hydroxychloroquine
- Hydroxychloroquine 0.4g daily





Concept extraction: SDTM Assumptions

SDTM IG v3.3 (Final)

4.2.7 Submitting Free Text from the CRF

4.2.8.2 Multiple Values for a Findings

Result Variable

4.4.9 Use of Dates as Result Variables

8.6.3 Guidelines for Differentiating

Between Events, Findings, and Findings

About Events

Pre-specified Events and Pre-specified Findings

Pre-specified Events

<https://www.cdisc.org/kb/articles/pre-specified-events-and-pre-specified-findings>

Collection of adverse event, clinical events, and medical history events can follow two approaches:

1. Were there any events?
 - a. If yes, what were the events?
2. Did event X occur?
 - a. If yes, record the details of the event(s)

The first approach is sometimes called "spontaneous" reporting. The answer to question 1 is in a CDASH --YN variable but is not recorded in SDTM. The results of the follow-up question 1a are represented in records which include the name of the event in TERM, and information about timing and other characteristics in other variables.

The second approach records pre-specified events. The fact that the question is of this type is represented by the inclusion of the variable PRESP with a value of "Y". The name of the event ("X") is in TERM and the response to question 2 is in OCCUR. If OCCUR = "Y", then the timing and other characteristics of X (the details in 2a) are recorded in other variables in the same record.

In summary:

1. In the "spontaneous" approach, TERM is an answer to question 1a, with further information about the event in other variables.
2. In the "pre-specified" approach, TERM is part of the question, the answer to question 2 is in OCCUR and further information about the event is in other variables.

Pre-specified Findings

For examinations that look for abnormalities, we have a similar situation. Questions can be asked as

1. Were there any abnormalities?
 - a. If yes, what were the abnormalities?
2. Did abnormality X occur?

In the first approach, sometimes described as an "open-ended question", the answer may or may not be represented in an SDTM record. An example where it would be represented in SDTM is for ECGs, where a record for test "Interpretation" would have a value of "NORMAL" or "ABNORMAL."



Concept extraction: Analysis Methods

2.1 Example 1: Basic Analysis of Covariance (ANCOVA)

This example illustrates an analysis dataset that supports a simple analysis of variance including covariates. Included in this illustration are identification of baseline values, change from baseline analysis, and handling of missing data.

The study measured lumbar spine bone mineral density (BMD) at baseline and then every 6 months for 3 years, with the goal of assessing the difference in effect between two treatments – Drug ABC and Placebo. Factors (i.e., machine type and subject's bone mineral density at baseline) that may influence the response are included in the model and the dataset. The response variable is the percent change from baseline in bone mineral density, with an increasing BMD indicating a positive effect. The source datasets (immediate predecessors) for the analysis dataset in this example are assumed to be ADSL and an SDTM domain containing the bone measurements, represented as “XX” in the example. Missing data are imputed using last observation carried forward (LOCF).

2.1.3 Analysis Results

The following data display illustrates an ANCOVA analysis performed using ADBMD as described above.

As described in the protocol and the SAP, the primary efficacy analysis endpoint is the percent change from baseline at Month 24 in the Bone Mineral Density (BMD) at the Lumbar Spine. The analysis is conducted using the Intent-to-Treat (ITT) population, with LOCF imputation for any missing values at Month 24. As the primary efficacy analysis method, an ANCOVA model is used, with planned treatment (as a categorical variable), baseline BMD value, machine type, and baseline BMD value by machine type interaction included as independent variables, and the percent change from baseline as the dependent variable. The adjusted least squares (LS) means and 95% confidence interval are presented.

ADaM Examples in Commonly Used Statistical Analysis Methods Version 1.0

https://www.cdisc.org/system/files/members/standard/foundational/adam/adam_examples_final.pdf

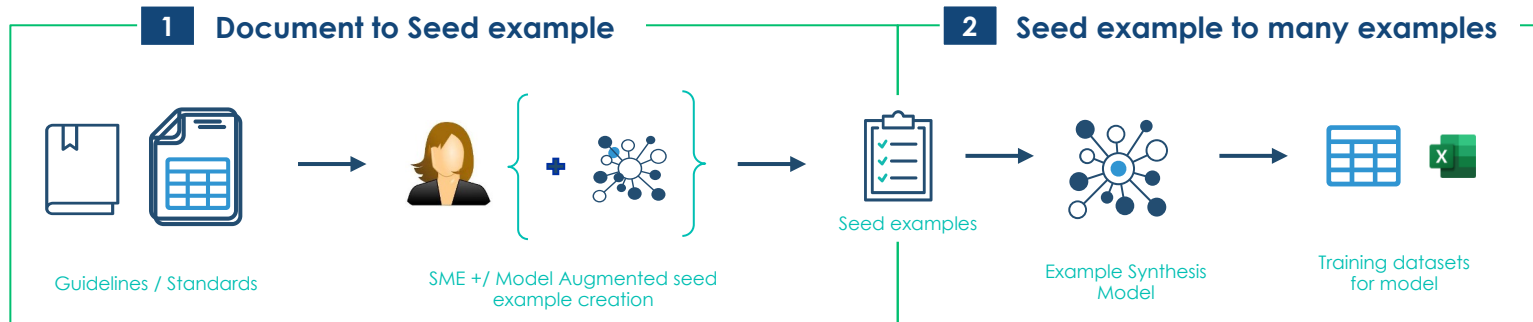
f

M24 : Month 24



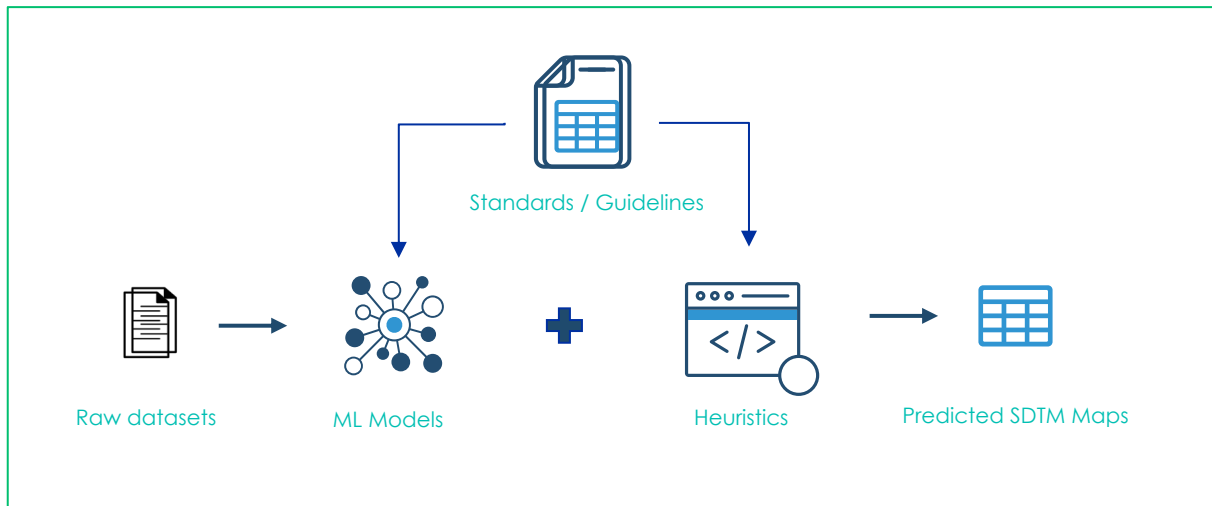


Standards to Training datasets Workflow





Grounding predictions with standards



Demo



References

https://www.cdisc.org/sites/default/files/2021-10/CDISC_360_2021_EU_Interchange.pdf

https://www.nlm.nih.gov/research/umls/new_users/online_learning/Meta_005.html

https://www.cdisc.org/sites/default/files/2021-08/Biomedical_Concepts_What_Did_We_Learn.pdf

<https://cmap.ihmc.us/docs/learn.php>

<https://www.cdisc.org/kb/articles/concept-maps-adverse-events-increasing-levels-detail>

<https://www.cdisc.org/kb/articles/magic-happens-between-cdash-and-sdtm>

https://www.cdisc.org/system/files/members/standard/foundational/SDTMIG_v3.3_FINAL.pdf

Thank You



**The Global Healthcare
Data Science Community**

Contact Channels

UK +44 1843 609600
office@phuse.global
phuse.global

Social Media

@phusetwitta
/phusebook
/phusetube
/company/phuse

Sudharsan Dhanavel
Saama Technologies
sudharsan.dhanavel@saama.com

Logesh Umapathi
Saama Technologies
logesh.umapathi@saama.com

PHUSE India Winter SDE (Chennai)

Hybrid Event, Saturday 10 December 2022