OAK for SDTM and Admiral for ADAM

Jia Liu / Data Scientist, Roche

Aug 29

Meet the Speaker

Jia Liu

Title: Data Scientist

Organization: Roche, PDD (Product Development Data Sciences)

An Analytics Data Scientist at Roche (China.Shanghai), with a focus on clinical trial data analysis and tool development. My career began with a Master's degree in Bioinformatics from Boston University, leading me to a previous role at Bayer (USA.Sacramento), where I specialized in agriculture data analysis.

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 materials provided in this presentation or any comments or information provided by the presenter are for scientific knowledge exchange purpose ONLY.*

- *This scientific material is intended for academic conferences or activities aiming at communicating and exchanging of medical information and shall be used by healthcare professionals as a reference only.*
  本资料为仅用于学术会议或活动的专业资料,旨在促进医药信息的沟通和交流,仅供医疗卫生专业人士参考。

- *All rights reserved.*

Agenda

1. OAK - SDTM Automation
2. Admiral - ADaM in R Asset Library

OAK - SDTM Automation

Driven by metadata & Global Data Standards

Open-Source {sdtm.oak}

Challenges with SDTM at the Industry Level

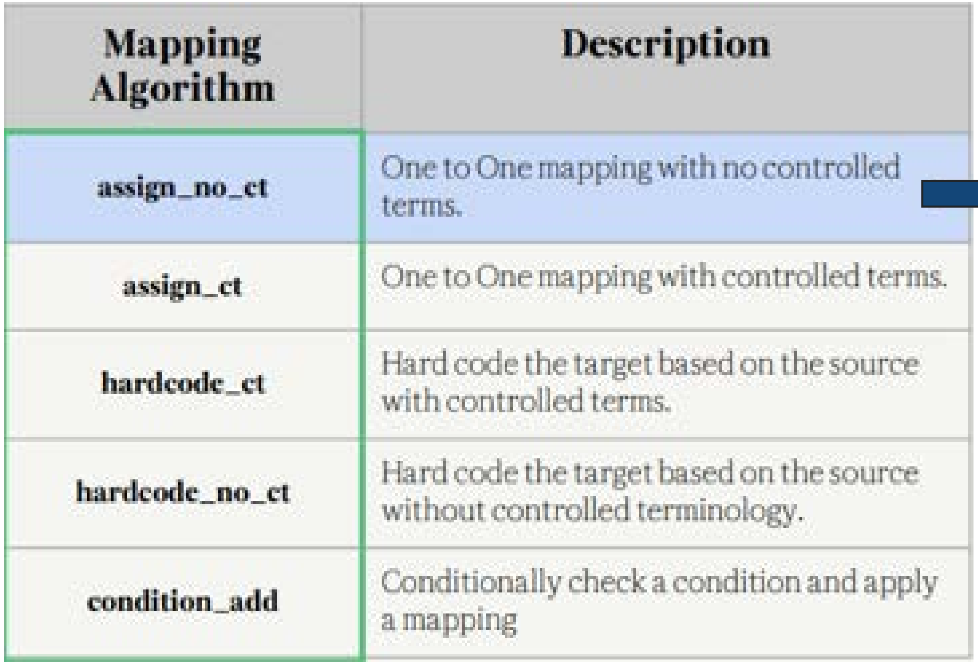
- **Raw Data Structure**: Data from different EDC systems come in varying structures, with different variable names, dataset names, etc.
- **Varying Data Collection Standards**: Despite the availability of CDASH, pharmaceutical companies still create different eCRFs using CDASH standards.

Sponsored by CDISC COSA, pharmaceutical companies, including Roche, Pfizer, Merck, GSK, and Vertex.

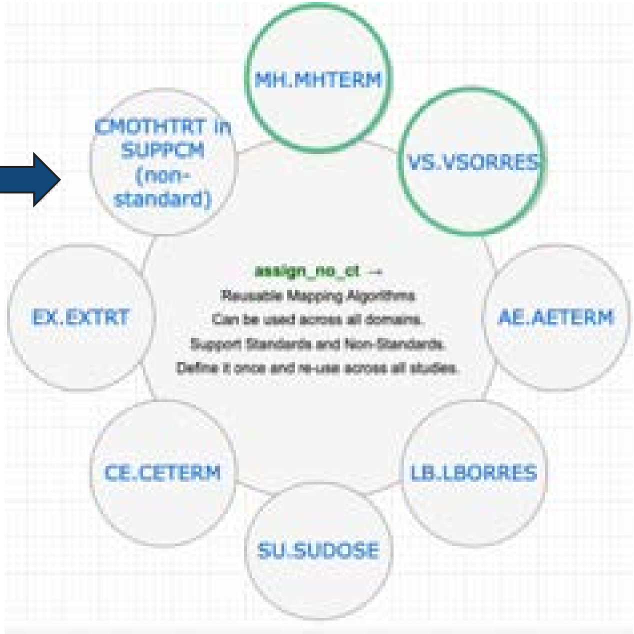
Part of the Pharmaverse Group of packages.

Goal: address this issue by providing an EDC-agnostic, standards-agnostic solution. It is an open-source R package that offers a framework for the modular programming of SDTM in R.

Core Concept - Reusable Algorithms



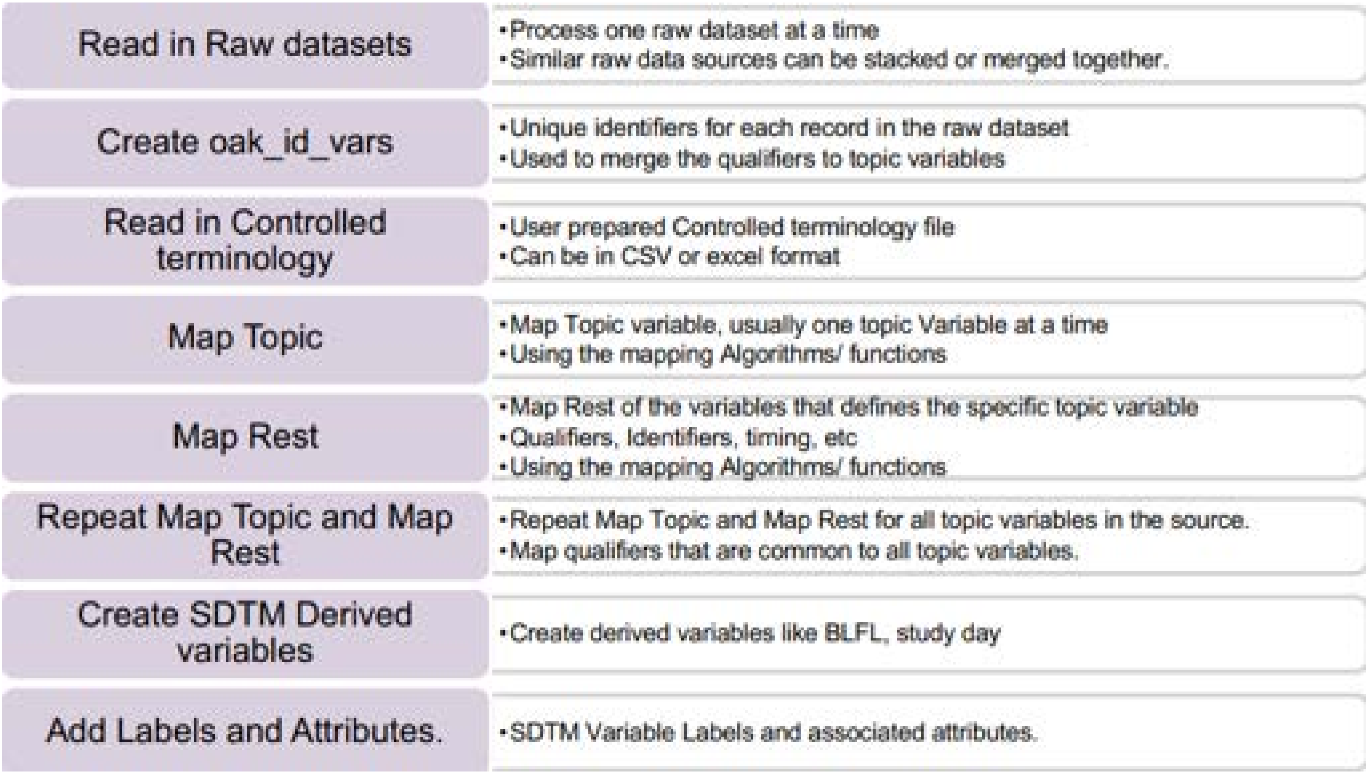
Mapping Algorithm	Description
assign_no_ct	One to One mapping with no controlled terms.
assign_ct	One to One mapping with controlled terms.
hardcode_ct	Hard code the target based on the source with controlled terms.
hardcode_no_ct	Hard code the target based on the source without controlled terminology.
condition_add	Conditionally check a condition and apply a mapping



assign_no_ct →
Reusable Mapping Algorithms
Can be used across all domains.
Support Standards and Non-Standards.
Define it once and re-use across all studies.

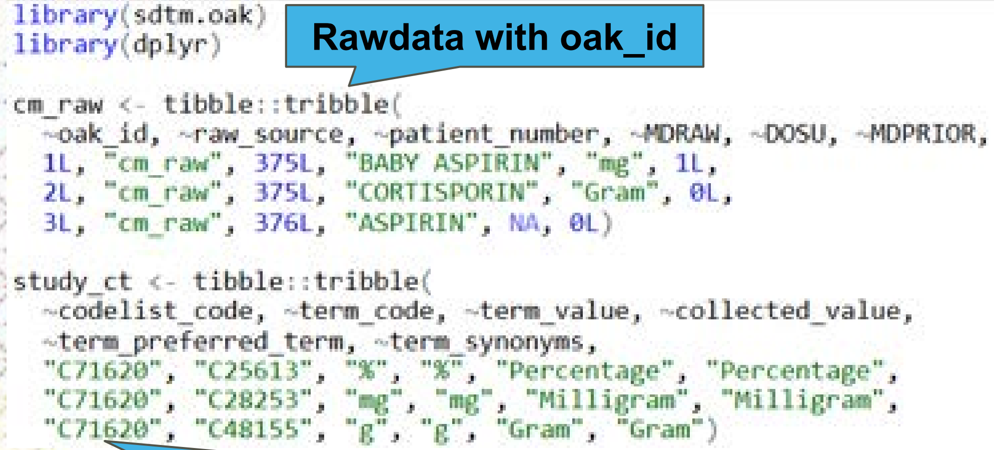
Read more about {sdtm.oak} algorithms/functions here

{sdtm.oak} – Programming Steps



Read in Raw datasets	•Process one raw dataset at a time •Similar raw data sources can be stacked or merged together.
Create oak_id_vars	•Unique identifiers for each record in the raw dataset •Used to merge the qualifiers to topic variables
Read in Controlled terminology	•User prepared Controlled terminology file •Can be in CSV or excel format
Map Topic	•Map Topic variable, usually one topic Variable at a time •Using the mapping Algorithms/ functions
Map Rest	•Map Rest of the variables that defines the specific topic variable •Qualifiers, Identifiers, timing, etc •Using the mapping Algorithms/ functions
Repeat Map Topic and Map Rest	•Repeat Map Topic and Map Rest for all topic variables in the source. •Map qualifiers that are common to all topic variables.
Create SDTM Derived variables	•Create derived variables like BLFL, study day
Add Labels and Attributes.	•SDTM Variable Labels and associated attributes.

Example on CM Domain



```
library(sdtm.oak)
library(dplyr)

cm_raw <- tibble::tribble(
  ~oak_id, ~raw_source, ~patient_number, ~MDRAW, ~DOSU, ~MDPRIOR,
  1L, "cm_raw", 375L, "BABY ASPIRIN", "mg", 1L,
  2L, "cm_raw", 375L, "CORTISPORIN", "Gram", 0L,
  3L, "cm_raw", 376L, "ASPIRIN", NA, 0L)

study_ct <- tibble::tribble(
  ~codelist_code, ~term_code, ~term_value, ~collected_value,
  ~term_preferred_term, ~term_synonyms,
  "C71620", "C25613", "%", "%", "Percentage", "Percentage",
  "C71620", "C28253", "mg", "mg", "Milligram", "Milligram",
  "C71620", "C48155", "g", "g", "Gram", "Gram")
```

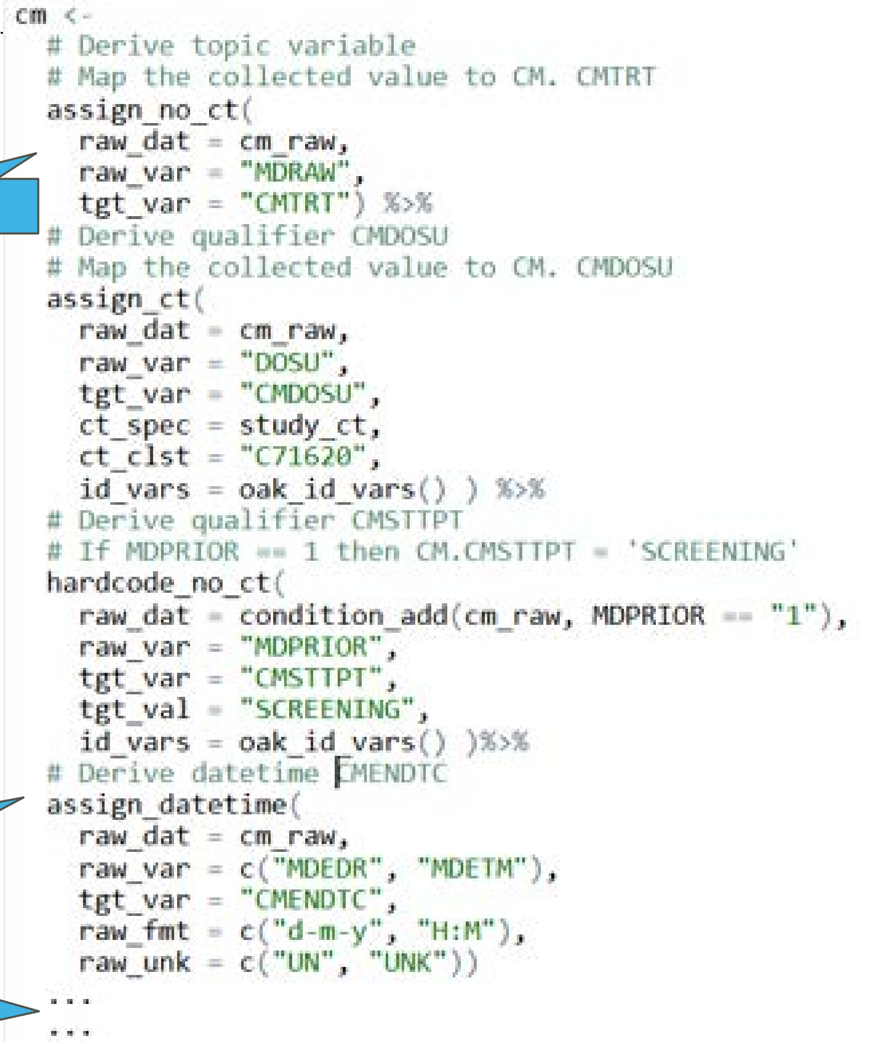
Rawdata with oak_id

Controlled terminology prepared

Map topic

Map rest

Create SDTM Derived variables
Add Labels and Attributes



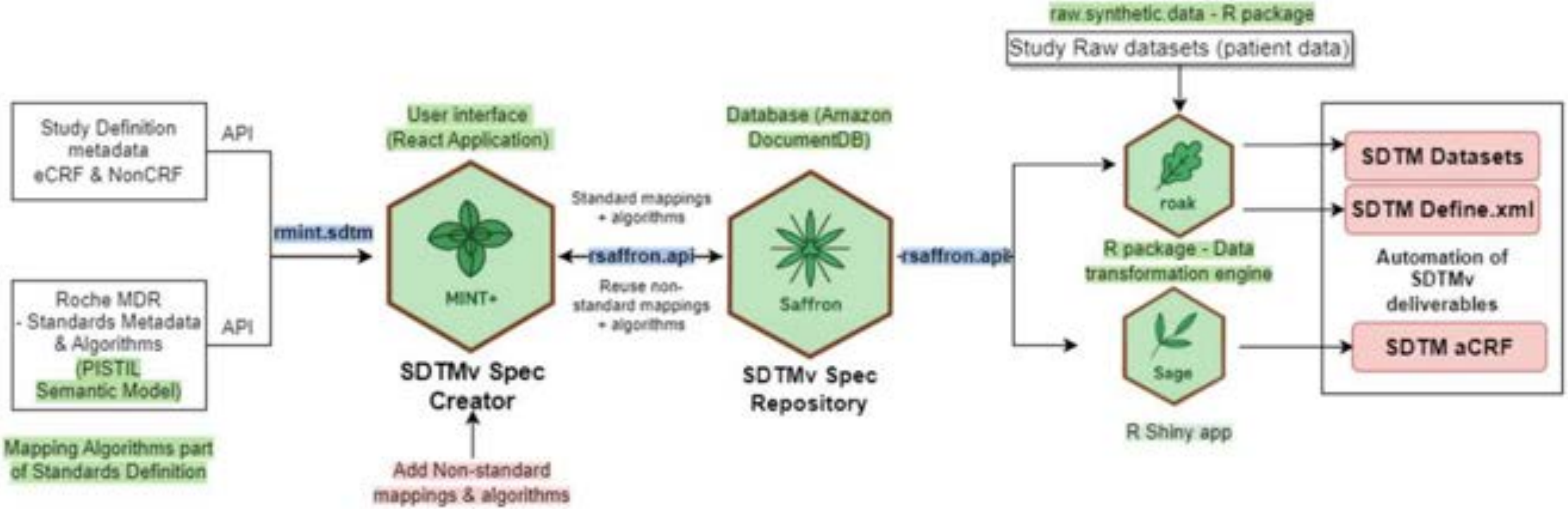
```
cm <-
  # Derive topic variable
  # Map the collected value to CM. CMTRT
  assign_no_ct(
    raw_dat = cm_raw,
    raw_var = "MDRAW",
    tgt_var = "CMTRT") %>%
  # Derive qualifier CMDOSU
  # Map the collected value to CM. CMDOSU
  assign_ct(
    raw_dat = cm_raw,
    raw_var = "DOSU",
    tgt_var = "CMDOSU",
    ct_spec = study_ct,
    ct_clst = "C71620",
    id_vars = oak_id_vars() ) %>%
  # Derive qualifier CMSTTPT
  # If MDPRIOR == 1 then CM.CMSTTPT = 'SCREENING'
  hardcode_no_ct(
    raw_dat = condition_add(cm_raw, MDPRIOR == "1"),
    raw_var = "MDPRIOR",
    tgt_var = "CMSTTPT",
    tgt_val = "SCREENING",
    id_vars = oak_id_vars() )%>%
  # Derive datetime CMENDTC
  assign_datetime(
    raw_dat = cm_raw,
    raw_var = c("MDEDR", "MDETM"),
    tgt_var = "CMENDTC",
    raw_fmt = c("d-m-y", "H:M"),
    raw_unk = c("UN", "UNK"))
  ...
  ...
```

Key Points of {sdtm.oak}

- Algorithms can **be re-used** across **multiple** SDTM domains.
- Programming **language agnostic**: This concept does not rely on a specific programming language for implementation.
- **Automation-ready**: Algorithms can be pre-specified for data collection standards in metadata repository (MDR).

OAK - Roche’s SDTM Automation Solution

Automate ~80% SDTM domains with ~22 Reusable Algorithms.



The Future Workflow for Automation

- **Prepare SDTM specification**: Users can define the raw data source, target SDTM domain, target SDTM variables, and algorithms used for automation. A template is still under development.
- **Prepare SDTM-controlled Terminology**: Users can define the SDTM-controlled terms applicable to the study. A template is still under development.
- **Automated process** to read the specification and make {sdtm.oak} function calls can create the code required to generate SDTM datasets or the datasets themselves.

More Information

{sdtm.oak} is the open source version of OAK. We have already made a few releases. If you are looking for more information, here are some resources.

- pharmaverse.github/sdtm.oak
- Introducing sdtm.oak
- List of Algorithms
- CDISC.OAK
- CRAN
- CDISC.SDTMIG

Admiral - ADaM in R Asset Library

To provide an open source, modularized toolbox that enables the pharmaceutical programming community to develop ADaM datasets in R.

Admiral “Family of Packages”

There are 3 types of admiral packages relevant for users:

- Core package—one open source package containing all core/common functions required to create ADaMs, usable by any company (e.g. general derivations for ADSL, OCCDS and BDS)
- TA (Therapeutic Area) package extensions—one open source package per TA with functions that are specific to endpoint algorithms and requirements for that particular TA (e.g. {admiralonco})
- Company package extensions—one closed source package covering specific needs and plug-ins for the company, such as access to metadata (e.g. {admiralroche})

The admiral core package is always a dependency of the other packages

Please watch this video to show more: [*https://www.youtube.com/watch?v=6i_hvns9TcM*](https://www.youtube.com/watch?v=6i_hvns9TcM)

Ecosystem

admiral	admiralonco	admiralophtha	admiralvaccine	admiralpeds	admiralmetabolic
GSK Roche	Roche GSK AMGEN Bristol Myers Squibb™	Roche NOVARTIS	Pfizer GSK Johnson&Johnson	Cytel Roche atorus Pfizer	novo nordisk® Roche

cdisc

Thank You!

Name: Jia Liu
Organization: Roche
Address: 1100 Long Dong Avenue
City, State ZIP: Shanghai 201203 Pudong New Area China
Phone: +86 13956158028
E-mail: jia.liu.jl12@roche.com