

Single Day Event

EDC & and Source-Agnostic SDTM Mapping
Automation via OAK Open Source

Josh Bernal, Genentech/Roche
(on behalf of Ram Ganapathy, Genentech/Roche)





Agenda

- **OAK at Roche**
- **Proposed R packages - {oak}, {mint}, and {raw.synthetic.data} - Scope & Deliverables**
- **OAK Open Source Vision**
- **Proposed Team Structures**
- **High-Level Roadmap**
- **OAK Workstreams**
- **How to learn more and contribute**



OAK at Roche - Overview & Progress

OAK 'Garden' aims at streamlining the clinical data flow from data collection to tabulation and the automation of the SDTM datasets, specifications and "define.xml".



The OAK 'garden' comprises a number of R packages that will form the framework for simplifying and automating SDTM electronic submission deliverables.



The core concept of OAK Garden is the reusable algorithms. SDTMv mappings are implemented as algorithms that transform the collected source (eCRF, non-CRF) data into the target tabulation data model.



OAK has been applied on 12 studies at Roche as of Q3 2023

There is a planned full R-pipeline submission (using OAK and other tools) by EOY

Driven by metadata & Roche/CDISC Data Standards, OAK Garden can:

Automate ~80% SDTM domains from 300 eCRFs/non-CRFs with 11k source-to-target mappings via ~25 Reusable Algorithms.



OAK at Roche - Algorithms



CORE CONCEPT: REUSABLE ALGORITHMS

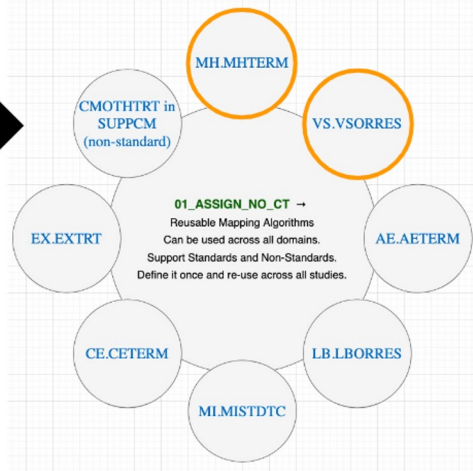
- SDTM Mappings are defined as algorithms that transforms the collected (CRF, nonCRF) source data into the target Tabulation data model. Mapping Algorithms are the back bone to the SDTM automation.

We have designed 25 unique mapping Algorithms to accommodate most (80%) of the TA standards & Non-CRF data models

Key Points:

- Algorithms can be re-used across Domains
- Algorithms can be **pre-specified for Standards**
- Users can reuse/add algorithms for non-standards or new data types
 - Both Standards and Non-Standards can therefore be supported**
- Programming language agnostic - this concept do not rely on a specific programming language for implementation. We have implemented them as R functions.

Mapping Algorithm	Description
01_ASSIGN_NO_CT	One to One mapping with no controlled terms.
02_ASSIGN_CT_ST	One to One mapping with controlled terms.
05_HARDCODE_CT	Hard code the target based on the source with controlled terms.
06_HARDCODE_NO_CT	Hard code the target based on the source without controlled terminology.
09_IF_THEN_ELSE	Conditionally check a condition and apply a mapping





OAK at Roche —>

Becomes...OAK Open Source



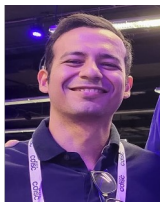
OAK Open Source - Leadership Team



Charles
Shadle
CDISC /
COSA*



Sam Hume
CDISC /
COSA



Omar Garcia
CDISC



Yogesh
Gupta
Pfizer



Rammprasad
Ganapathy
Roche/Genentech



Bhaskar Ponugoti
Merck



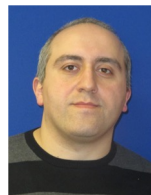
Joshua Bernal
Roche/Genentech



Lisa Hourteloot
Pfizer



Aditya Parankusham
GSK



Edgar Manukyan
Roche/Genentech



Susheel Arkala
Vertex



Sandra VanPelt
Pfizer



Preetesh Parikh
Pfizer



{mint} – SDTM Mapping Metadata Prep

Scope

Prepares the study SDTM mappings metadata (i.e. specifications) for any source and data standard in the format OAK expects. Use CDISC Library to develop this New R package. This can be extended to any EDC and Sponsor MDR as long the input matches to the ODM format.

Input:

- Standards Metadata in ODM format. Standards metadata will include the Algorithms associated with the SDTM mappings.
- Study metadata in ODM format

Output

Study SDTM mappings metadata in the format OAK expects.

Not in Scope:

How the data is extracted from Sponsor's MDR and Sponsor's EDC is not included. Every sponsor may need to build this extraction mechanism.

{metacore} – This feature could be implemented in the {metacore} package. Yet to be decided.



{raw.synthetic.data} – Generate raw test data

Scope:

An EDC/Source standard Agnostic and ODM compliant solution. Some generic functions from the Roche version of the {raw.synthetic.data} package can be exported to the open-source version.

Input:

- Study metadata in ODM Format (eCRF, Visit definitions)

Output:

Patient data in ODM Format

Not in Scope:

How the data is extracted from Sponsor's EDC in ODM format is not included. Every sponsor may need to build this extraction mechanism.



{oak} – Data Transformation Engine

Scope:

Data Transformation Engine that gets SDTM mappings metadata input from the New mint R package and also works with CDISC/Sponsor MDR Library & ODM clinical data format to enable Metadata driven automation. Some generic functions from the Roche version of the {oak} package can be repurposed to the open source version. This package will have the code for the Algorithms and generic SDTM functions.

Input:

- Study SDTM mappings Metadata which is an output of the new mint R package.
- Patient data in ODM Format

Output:

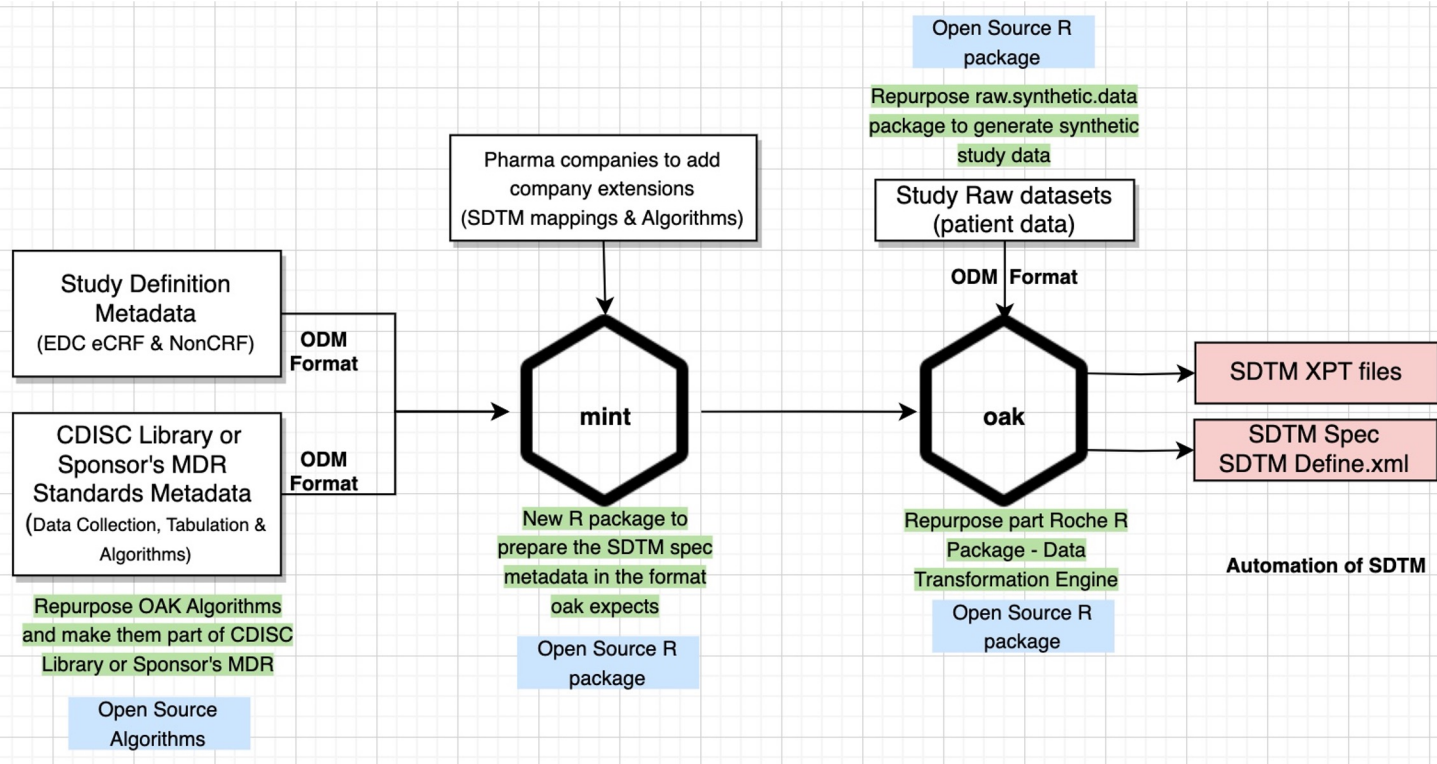
SDTM datasets

Not in Scope:

How the data is extracted from Sponsor's EDC in ODM format is not included. Every sponsor may need to build this extraction mechanism.



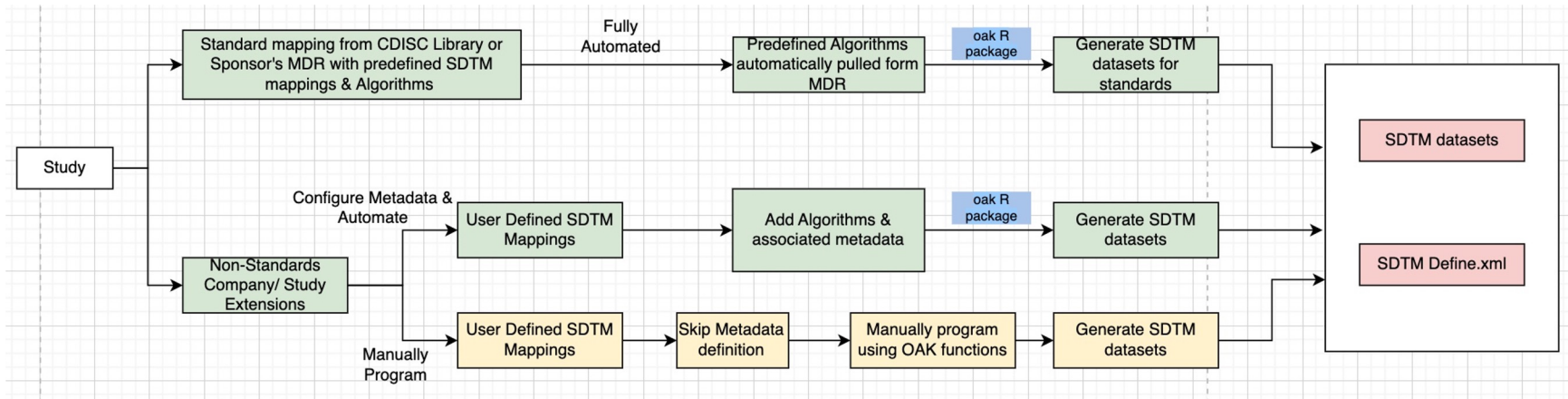
OAK Open-source Vision - SDTM Automation



- ❖ Metadata Driven SDTM Automation Solution
- ❖ EDC/Source and Data standards agnostic Solution (ODM Compliant*)
- ❖ Can be extended to include Sponsor extensions



OAK Open-source Vision - Study Setup Overview



- ❖ Fully Automated standard mappings from CDISC Library or Sponsor's MDR*
- ❖ Algorithms can be extended to automate study extensions (non-standards) for simple SDTM mappings
- ❖ Option to manually program complex SDTM mappings
- ❖ A SDTM domain can have some variables as fully automated, and some configured & automated and some are manually programmed.

*MDR =
Metadata
Repository



OAK Open-source Vision - Scope

- ❖ Pick domains like CM, MH, VS and DM for the PoC.
- ❖ Add algorithms and associated metadata to CDISC Library for CDASH standards.

(similar to what the Roche team did in Roche's MDR)

- ❖ R packages Development
 - **{mint}** – Prepare the study SDTM mapping metadata in the format OAK expects.
 - **{oak}** – Data Transformation Engine
 - **{raw.synthetic.data}** – Generate synthetic raw data in ODM format



OAK Open-source Vision - Summary

- ❖ Develop an Open-source Metadata driven R based SDTM automation solution.
- ❖ We will use CDASH aligned collection for POC, but vision is to extend this solution for any source standard (not just CDASH)
- ❖ Follow ODM standards & remain an EDC-agnostic solution.
- ❖ Completely leverage OAK Algorithms, CDISC Library.
- ❖ Provide a framework for automation when CDASH/Sponsor standards are extended to meet study specific nuances.
- ❖ Explore ways to use oak when ODM doesn't exist, example, Real World Data (RWD), external clinical trials, claims data, public consortia data

OAK Teams Structure

Leadership Team



- Define project scope
- Establish strategy
- Create roadmap
- Monitor progress
- Foster collaboration and effective communication

LT

Metadata Curation Team



- Develop & maintain algorithms' metadata standards
- Add algorithms metadata to MDR and CDISC Library
- Create user documentation to aid companies using approach with own metadata

MCT

OAK Development Team



- Lead effort to develop & maintain 3 OS R packages: {mint}, {oak}, & {raw.synthetic.data}
- Develop & maintain code contribution guide
- Produce unit test cases
- Prepare documentation & vignettes
- Designate sprints for further development

ODT

OAK Community Team



- Help to code / test R functions designed by the ODT
- Ensure packages meet required specifications
- Drive compatibility across different environments / platforms

OCT



OAK Workstreams – Q1 - 2024 Deliverables





How to Learn More & Contribute

❖ To become a member of one of the OAK teams:

Reach out to Omar Garcia Calderon <ogcalderon@cdisc.org>

❖ OAK Public Wiki

<https://wiki.cdisc.org/display/oakgarden>

❖ OAK GitHub

<https://github.com/pharmaverse/oak>

<https://github.com/pharmaverse/mint>

<https://github.com/pharmaverse/raw.synthetic.data>

❖ OAK Open Source KOM
Materials:

[Link](#)

❖ OAK Training Materials:

[Link](#)



The Global Healthcare Data Science Community

Contact Channels

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

Social Media

- [@phusetwitta](https://twitter.com/phusetwitta)
- [/phusebook](https://facebook.com/phusebook)
- [/phusetube](https://youtube.com/phusetube)
- [/company/phuse](https://linkedin.com/company/phuse)