



OAK Productivity Experiment

– How We Made SDTM Programming

5 x Faster



Amy Franklin, Roche, amy.franklin@roche.com

Botany 



**Oak Garden - Metadata-driven
SDTM automation in R**
Nov 2023 | PHUSE EU Connect 2023

Amy Franklin, Senior Data Scientist
Magdalena Krochmal, Senior Data Scientist

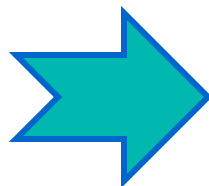
 **EU Connect 2023**

Botany 



**MINT+ Automation and
Digitization of SDTM specification**
Nov 2023 | PHUSE EU Connect 2023

Magdalena Krochmal, Senior Data Scientist
Adam Forys, Principal Data Scientist







**OAK Productivity Experiment
– How We Made SDTM Programming
5x Faster**

 Amy Franklin, Roche, amy.franklin@roche.com
Magdalena Krochmal, Roche, magdalena.krochmal@roche.com

Strasbourg, 11 Nov 2024

 **EU 2024**
The Clinical Data Science Conference
10-13 November
Strasbourg Convention Centre



Strasbourg Cathedral

Table of contents

1 OAK Garden Overview

2 DTS (SAS based Tool) and OAK (R based Tool)

3 **OAK Productivity Experiment - Setup, Results, Summary**

4 OAK Automation Metrics

5 Key Takeaways

6 Open-Source {sdtn.oak}



Oak Garden



OAK Garden aims to automate the study SDTMv deliverables



Semantic model
in GDSR

Part of
GDSR



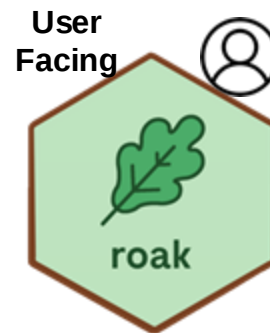
Content upload
to GDSR
(Java program)



SDTMv Spec
Creator
(Web App)

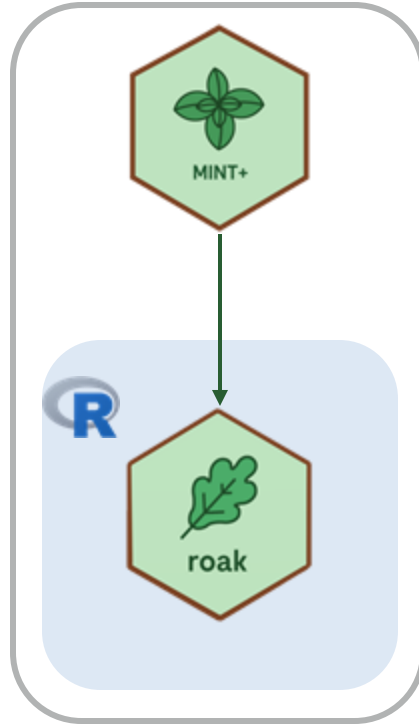


SDTMv Spec
Repository
(DocumentDB)



Data
transformation
engine
(R package)

Oak Garden - Automation of SDTM specs and Datasets



Oak Garden

Web Application:

Automates creation of :

- ❖ Study SDTMv spec
- ❖ Study Controlled Terminology
- ❖ Study VLM
- ❖ SDTM Define.xml

R package:

Automates creation of:

- ❖ SDTM(v) datasets
- ❖ SDTM(v) to SDTM
- ❖ XPT file creation

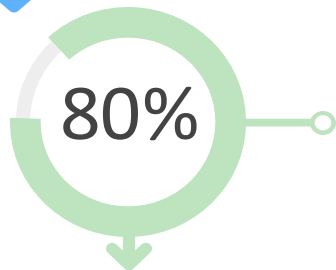


Oak Garden



Driven by **metadata & Global Data Standards**,

The OAK Garden can:



Automate ~80% of study SDTM(v)
domains with the concept of
Reusable Algorithms (in a typical Roche study)



In a typical Roche study:
Expect ~**20% programming and manual work** to
enable full automation

OAK Garden - SAS based DTS vs R based OAK - Quick Introduction



SDTMv in DTS (SAS)



Laborious study SDTMv specification creation and study DTS programming.

No automation. SAS based DTS templates can provide some efficiency.



No Digitalization of Business Process (SDTMv spec)



SDTMv in OAK (R)



Metadata-driven automated SDTMv specification creation using MINT+

Automation based on adherence to Global Data Standards.

~80% automation of SDTM domains with ~22 Reusable Algorithms



Digitalization of Business Process (machine readable SDTMv spec)



Current state OAK (Q1-2025)

- Validated, stable versions of MINT+ and {roak} are available to the study teams.
- More than 50 study teams** have started using OAK for regulatory reporting **including one submission.**
- Experiments performed to **measure the impact on productivity** to help us understand how new ways of working improve the study team's efficiency.

What is the actual productivity gain with OAK?

Productivity Experiment



Study Details - Glofitamab in Mantle Cell Lymphoma (GO43878)



Protocol Title - A PHASE III, OPEN-LABEL, MULTICENTER, RANDOMIZED STUDY EVALUATING GLOFITAMAB AS A SINGLE AGENT VERSUS INVESTIGATOR'S CHOICE IN PATIENTS WITH RELAPSED/REFRACTORY MANTLE CELL LYMPHOMA

Data collections Standards used in study



87.74% - (1424) ● Standard ⓘ

12.26% - (199) ● Nonstandard ⓘ

85% - Average % of CRF Standards adherence across studies in Onc - Oncology
87.74% - Standards Adherence of GO43878 study

Study profile matches with a typical P3 study.

GO43878 study breakdown:

- Mapped in SDTMv IGv 3.4
- SDTMv domains programmed - **40**
- Standard eCRFs - **91.92%(91)**
- Non-standard eCRFs - **8.08%(8)**
- Non-Standard Raw variables - **12.26% (199)**

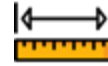
Experiment Details



Setup



OAK Programming vs. DTS (SAS-based)



What we measured



Average Tool Experience



Time taken for SDTMv spec creation
in MINT+



Time taken for SDTMv programming

OAK Productivity Experiment - Summary



Results: OAK is far quicker than the SAS based DTS tool.

Developer's skillset: Effort required to program widely varies from developer to developer based on their familiarity with the Therapeutic area, Molecule, Early phase vs Late phase, etc and their technical expertise with Git, R, SAS.

~3-5 times productivity gain - Depending on the developer's skill set, the OAK team expects ~3-5 times productivity gain to create SDTMv in OAK when working with the wider pool of resources.



We expect further productivity increase:

- As the SDTM programmers gain more experience in OAK.
- Results of this readout are based on Q1-2024 release of {roak} & MINT+. Subsequent releases feature substantial improvements in the package and process.

OAK Automation - Metrics from 56 studies (Jan 2025)



Average values from 56 studies which have completed SDTMv specs in MINT+.

Data Tabulation Metrics

Standards SDTMv mappings automated	82% (100535 std mappings)
Non-standards	18% (21642 non-standards)
Non-standard SDTMv mappings added by the user	18647 (Rest are yet to be mapped)

Automated with standard
Algorithms in MINT+



97%

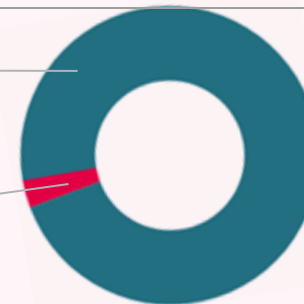
(18199 mappings)

Manually programmed in R



3%

(452 mappings)



Across Standards & Non-standards



99.6% Automated mappings
(excluding any automation failures)



0.4% Manually programmed mappings

OAK Automation - Metrics Insights



More insights to metrics



6 studies **automated all non-standard mappings** using standard algorithms in MINT+.



9 studies had **less than 10 non-standard mappings** flagged as manually programmed.

Considering both standards and non-standards,
OAK is already automating 99% of the mappings.

In subsequent releases, we are very close to eliminating any programming effort for SDTM(v) creation.

Our Future State
Could Be



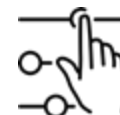
Complete the
SDTM(v) specs



QC the SDTM(v)
output
(99% automated)



Review log
files



Address any automation failures,
program domains (TDDs, RELREC),
variables (ex. ARM) that cannot be
automated



OAK Releases Since the Experiment



Some of the features that will further enhance OAK's productivity, including new ways of working:

MINT+

- **UI improvements** - Updates to the MINT+ UI based on user feedback

{roak}

- **Package Improvements** - Bug fixes to improve automation and reduce manual programming efforts for users.
- **Code Generation** - Our goal is to reduce the programming time to nearly zero by enabling code generation for SDTMv share and activity. Users should be able to complete the specification in MINT+ and see the SDTMv output in Harbour with minimal to zero programming.

Future OAK Releases



Some of the features that will further enhance OAK's productivity, including new ways of working:



MINT+

- **Non-standard SDTM Mapping suggestions** - Generate a report containing suggestions for non-standard SDTMv mappings that a user can review and add to MINT+, reducing search time.

{roak}

- **Package Improvements** - Bug fixes to improve automation and reduce manual programming efforts for users.
- **Configuration files automation** - Generate the first version of the configuration files automatically.
- **raw.synthetic.data** - Roll out {raw.synthetic.data} package to be used for test data generation.

Key Takeaways for Study Teams & SDTM Programmers



Request fewer resources to program SDTMv in OAK

Determine the resources required based on the standards in your study. Studies with higher standards adherence require fewer resources for SDTM mapping.



Focus on understanding SDTM concepts

Understanding of SDTM concepts and SDTMv mapping helps to achieve more efficiency. Focus on learning SDTM mapping.

Keep Standards Adherence high

Ensure that the eCRF and NonCRF data collected in your study adhere to the global data standards outlined. Higher standards lead to greater efficiency, enabling OAK to automate the majority of the mappings.



Keep Non-standard Data Collection Consistent

Ensure that non-standard data is collected consistently across studies within the molecule. Make sure similar data is collected in the same manner to facilitate uniform mapping.



{sdm.oak}

- SDTM Programming in R

Pharmaverse / CDISC COSA



Core Concept - Reusable Algorithms



Reusable

Algorithms can be re-used across multiple SDTM domains



Programming language-agnostic

{sdm.oak} team implemented them in R



Automation-ready

Algorithms can be prespecified for data collection standards in metadata repository



{sdtm.oak} - Modular approach



- Provide a framework to modular programming of SDTM in R.
- Very similar to **{admiral}** style (easier for programmers).

For example, a mapping of **CM.CMTRT** from raw data source **MD1-MDRAW** using **assign_no_ct** mapping algorithm which is programmed as **assign_no_ct()** function.

```
1 # Create a CM domain. The first step in creating CM domain is to create the topic variable
2 cm <- MD1 |>
3   # Derive topic variable
4   assign_no_ct(
5     raw_dataset = MD1, # Not required as pipe will send the piped object to the first argument.
6     raw_variable = MDRAW,
7     target_sdtm_var = CMTRT
8   )
9
```

{sdtm.oak} Path to Open-source SDTM Automation



Modular programming

```
cm_raw_data <- read_raw_data_csv(filename = "~/inst/cm_domain/cm_raw_data.csv") |>
# Derive oak_id_vars
derive_oak_id_vars()

# Create CM domain. The first step in creating CM domain is to create the topic variable

cm <- cm_raw_data |>
# Derive topic variable
assign_ct{
  raw_dataset = MD1, # This is added for pseudocode. Not required as pipe will send it
  raw_variable = MDRAW,
  target_sdtm_var = CMRTT
} |>
# Derive qualifier CHDOSU
# Use merge and add the qualifier to the topic variable
assign_ct{
  raw_dataset = MD1,
  raw_variable = DOSU,
  target_sdtm_var = CHDOSU,
  study_ct = study_ct,
  target_sdtm_variable_codelist_code = "C71620",
  merge_to_topic_by = c(oak_id_vars,
    topic_var_source = MDRAW
  )
} |>
# Derive qualifier CHDOSFRM and merge it with the target dataset
assign_ct{
  raw_dataset = MD1,
  raw_variable = MDRFRM,
  target_sdtm_var = CHDOSFRM,
  target_sdtm_variable_codelist_code = "C66720",
  study_ct = study_ct,
  merge_to_topic_by = c(oak_id_vars,
    topic_var_source = MDRAW
  )
} |>
# DERIVE CHROUTE
assign_ct{
  raw_dataset = MD1,
  raw_variable = MDRFRM,
  target_sdtm_var = MDRTE,
  target_sdtm_variable_codelist_code = "C66720",
  merge_to_topic_by = c(oak_id_vars,
    topic_var_source = MDRAW
  )
} |>
```

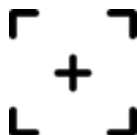


Create SDTM spec & automate

Prepare SDTM Spec in the format OAK expects

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

```
vs <- create_domain(
  domain = "vs",
  spec = "study_sdtm_spec.csv",
  ct = "study_ct.csv")
```



Initial focus is to develop {sdtm.oak} and pave the way for SDTM programming in R.



sdtm.oak   CRAN **0.1.0** Stars **26** Contributors **9**

An Electronic Data Capture system and Data Standard agnostic solution that enables to develop SDTM datasets in R



AVAILABLE

{sdtm.oak} v 0.1.0 on CRAN!

Try it!
Give feedback!
Contribute!

Doing now what patients need next