
# APAC Community Call

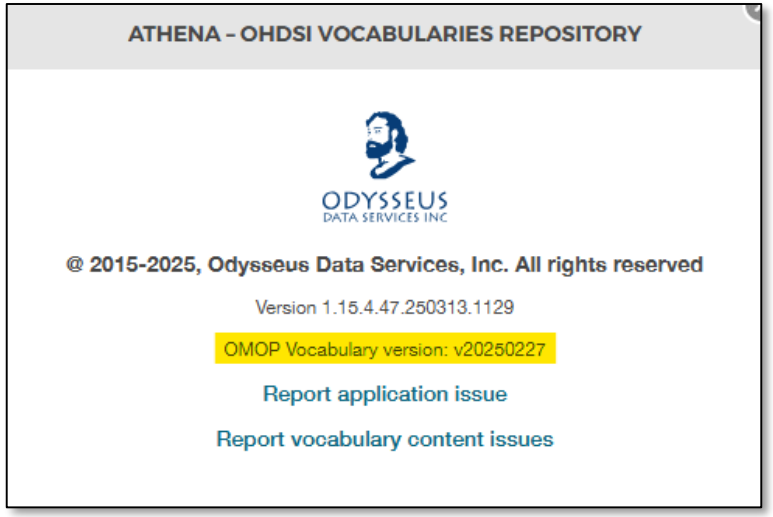
May 14, 2025

# Agenda

- OHDSI Global News
- OHDSI APAC News
- Localization of Book of OHDSI (Japanese version)
- Enabling OHDSI APAC Partnership through TRUST

# OHDSI Global News

- Latest update of OMOP Standardized Vocabularies has been released
- Adds new drugs, procedures and lab tests, improved mappings for conditions, and expanded hierarchies
- Full release notes available at https://0github.com/OHDSI/Vocabulary-v5.0/releases/tag/v20250227_1740652703.000000



# OHDSI Global News

**OHDSI Global Symposium Set for October 7-9 at Hyatt Regency Hotel in NJ, USA**

SAVE THE DATE: Collaborator Showcase Submissions will be due July 1!

Agenda and registration will be shared when available

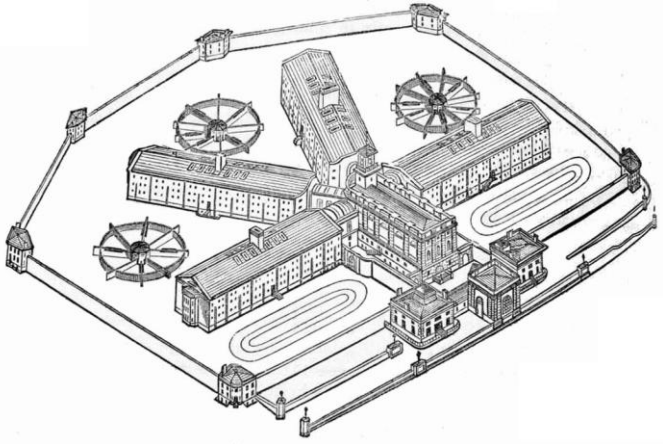
# OHDSI Global News

**Europe Symposium registration is open!**

**July 5-7 at "Old Prison" in Hasselt, Belgium**

Registrations will close after June 16

Event landing page: https://www.ohdsi-europe.org/symposium-event
Registration page: https://forms.gle/STSJZ7KH6hAJKv9H9



OHDSI EUROPE SYMPOSIUM 2025

Join the Network: Advancing Federated Real-World Evidence in Europe

**Symposium Agenda – July 7, 2025**

| Time | Topic | |
|---|---|---|
| 8:00 - 9:00 | *Registration & Coffee* | |
| 9:00 - 9:10 | Welcome to the European OHDSI Journey | |
| 9:10 - 9:30 | Journey of OHDSI: Where have we been and where can we go together? | |
| 9:30 - 11:00 | Impact of Leveraging OMOP CDM for Scalable and Reliable Evidence Generation Showcased by the National Nodes | |
| 11:00 - 11:30 | *Coffee Break* | |
| 11:30 - 12:45 | Collaborator Showcase: Rapid Fire Presentations | |
| 12:45 - 13:45 | *Lunch* | |
| 13:45 - 16:00 | OHDSI Collaborator Showcase | Early Investigator Mentor Meeting (14:00 - 15:00) |
| 16:00 - 17:10 | Bridging Policy and Practice: OHDSI's Role in Implementing the European Health Data Space *(Panel debate)* | |
| 17:10 - 17:30 | Closing remarks | |

# OHDSI Global News

## Join OHDSI Global's Guideline-driven Evidence Generation Initiative!

For more information on the initiative and selected studies, please visit:

https://ohdsi.org/clinical-guideline-evidence-opportunities-2025/

### Community Presents Clinical Guideline Evidence Opportunities for 2025; Learn More and Share Your Potential Interest in Collaboration

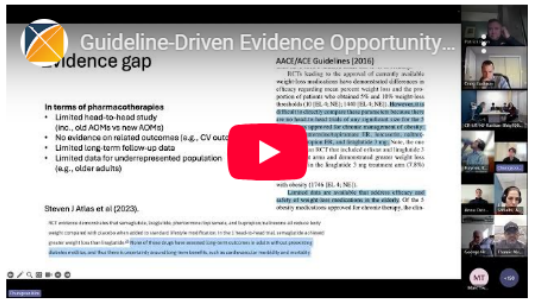
Clinical guidelines not only offer treatment recommendations for healthcare providers but also highlight evidence gaps that could shape critical questions for both clinicians and patients. The OHDSI community aimed to identify these gaps and explore how they could be addressed through network studies across the OHDSI Evidence Network.

Throughout January, collaborators around the world highlighted such gaps in a forum thread and joined a community call to provide a brief description of the gap ad why OHDSI is positioned to generate reliable and informative real-world evidence. Please check out the videos below or read about these evidence opportunities, and then fill out the brief form below to share your interest in joining one or multiple studies.

Join a Guideline Evidence Generation Network Study

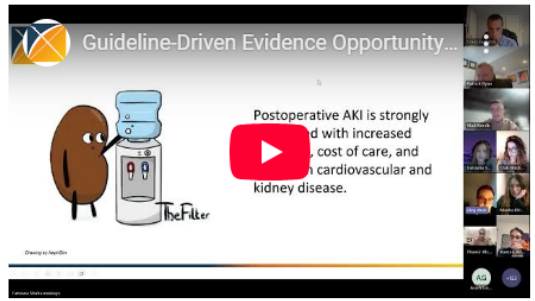
## Video Presentations

**Obesity Management**



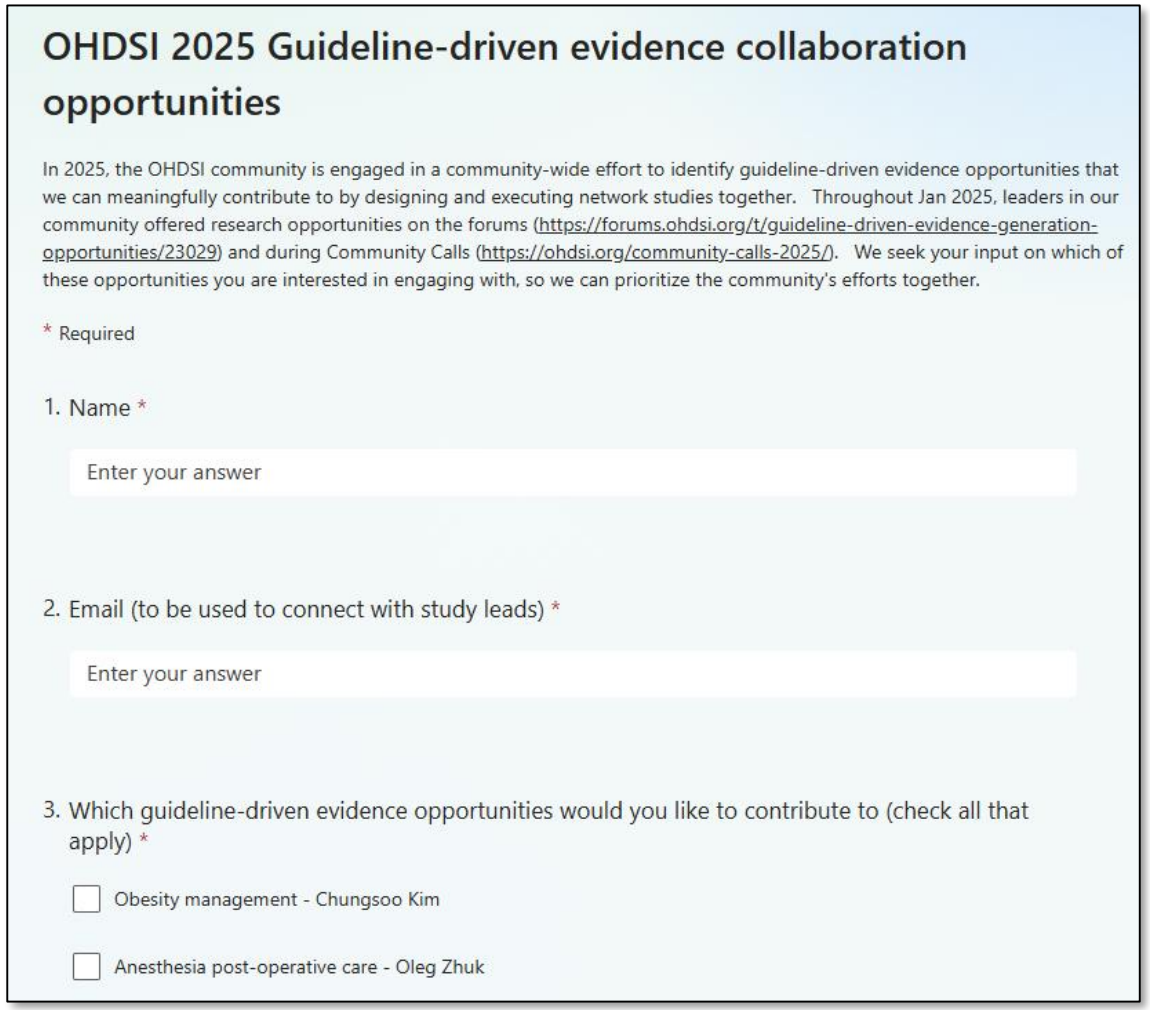

Presenter: Chungsoo Kim

**Anesthesia Post-Operative Care**




Presenter: Oleg Zhuk

# OHDSI Global News



## OHDSI 2025 Guideline-driven evidence collaboration opportunities

In 2025, the OHDSI community is engaged in a community-wide effort to identify guideline-driven evidence opportunities that we can meaningfully contribute to by designing and executing network studies together. Throughout Jan 2025, leaders in our community offered research opportunities on the forums (https://forums.ohdsi.org/t/guideline-driven-evidence-generation-opportunities/23029) and during Community Calls (https://ohdsi.org/community-calls-2025/). We seek your input on which of these opportunities you are interested in engaging with, so we can prioritize the community's efforts together.

* Required

1. Name *

Enter your answer

2. Email (to be used to connect with study leads) *

Enter your answer

3. Which guideline-driven evidence opportunities would you like to contribute to (check all that apply) *

- [ ] Obesity management - Chungsoo Kim
- [ ] Anesthesia post-operative care - Oleg Zhuk

https://forms.office.com/pages/responsepage.aspx?id=lAAPoyCRq0q6TOVQkCOy1dASOXn2aCVJpwpY4vxZsSdURVhXMk1LTzRFMUMzM1czMUNWV0xXQzA5MyQlQCN0PWcu

# OHDSI APAC News

## 🚀 2025 APAC Studies Kick-off!

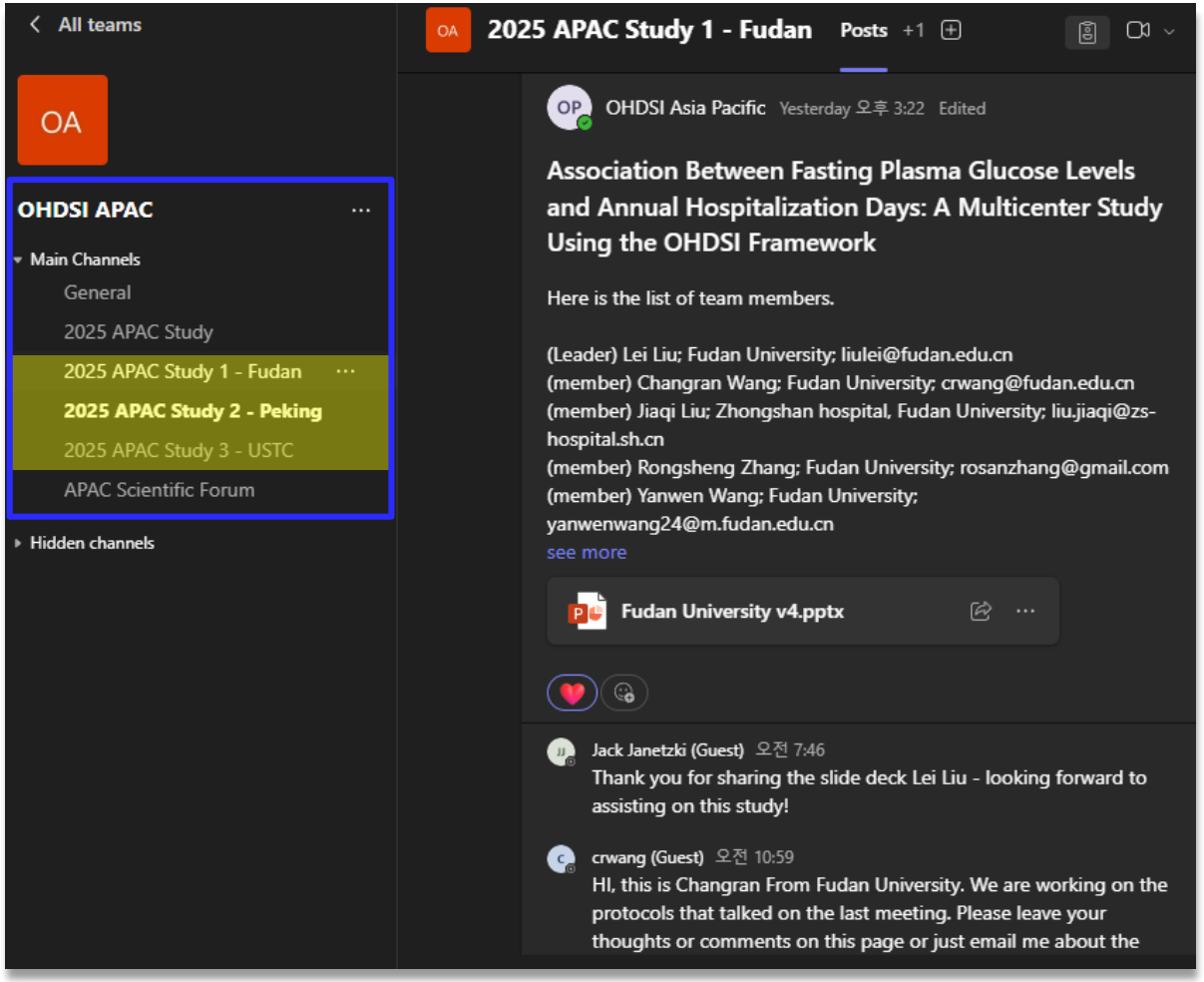
MS Teams channels for each of three studies has been created.

👉 [2025 APAC Study - Fudan]

👉 [2025 APAC Study - Peking]

👉 [2025 APAC Study - USTC]

Warmly welcome your participation and contribution in each channel!



# OHDSI APAC News

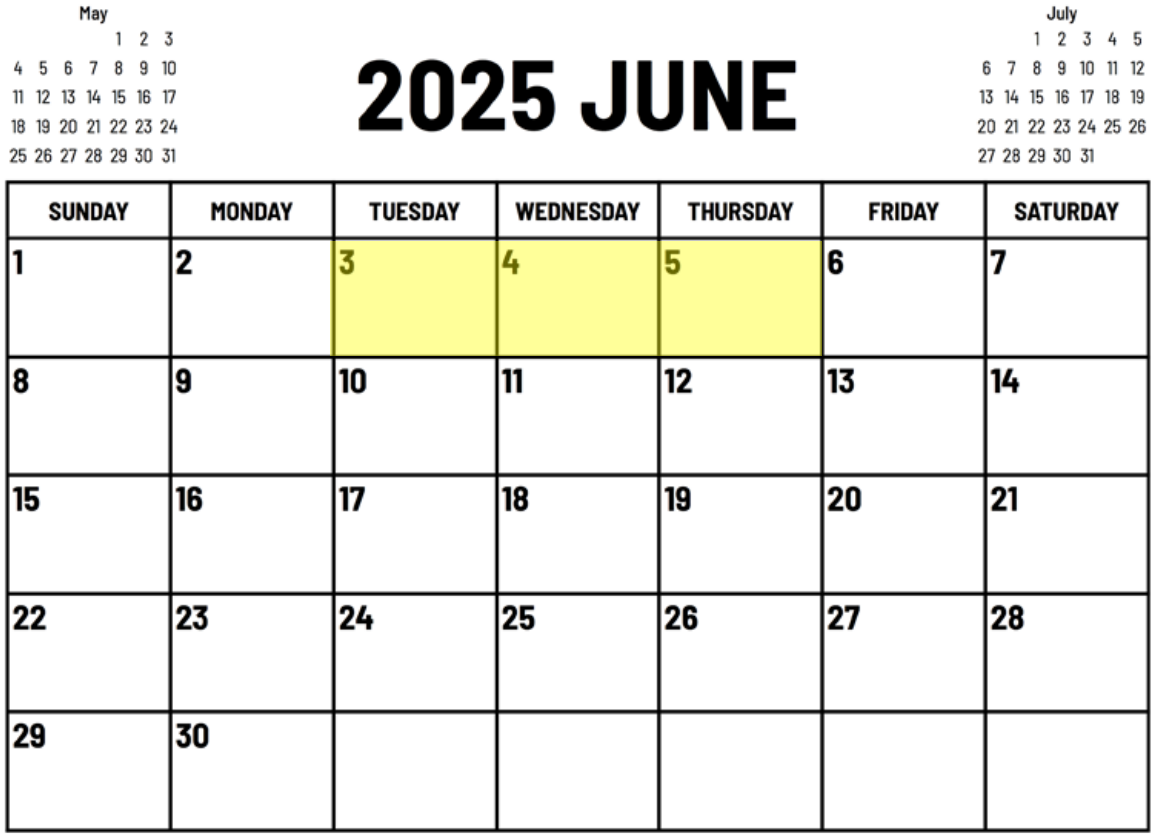
**First draft Protocol Meetings** (GMT):

**Fudan**: Tuesday, May 3rd

**Peking**: Wednesday, May 4th

**USTC**: Thursday, May 5th

🕗 All meetings begin at 9:00 or 8:00 am GMT

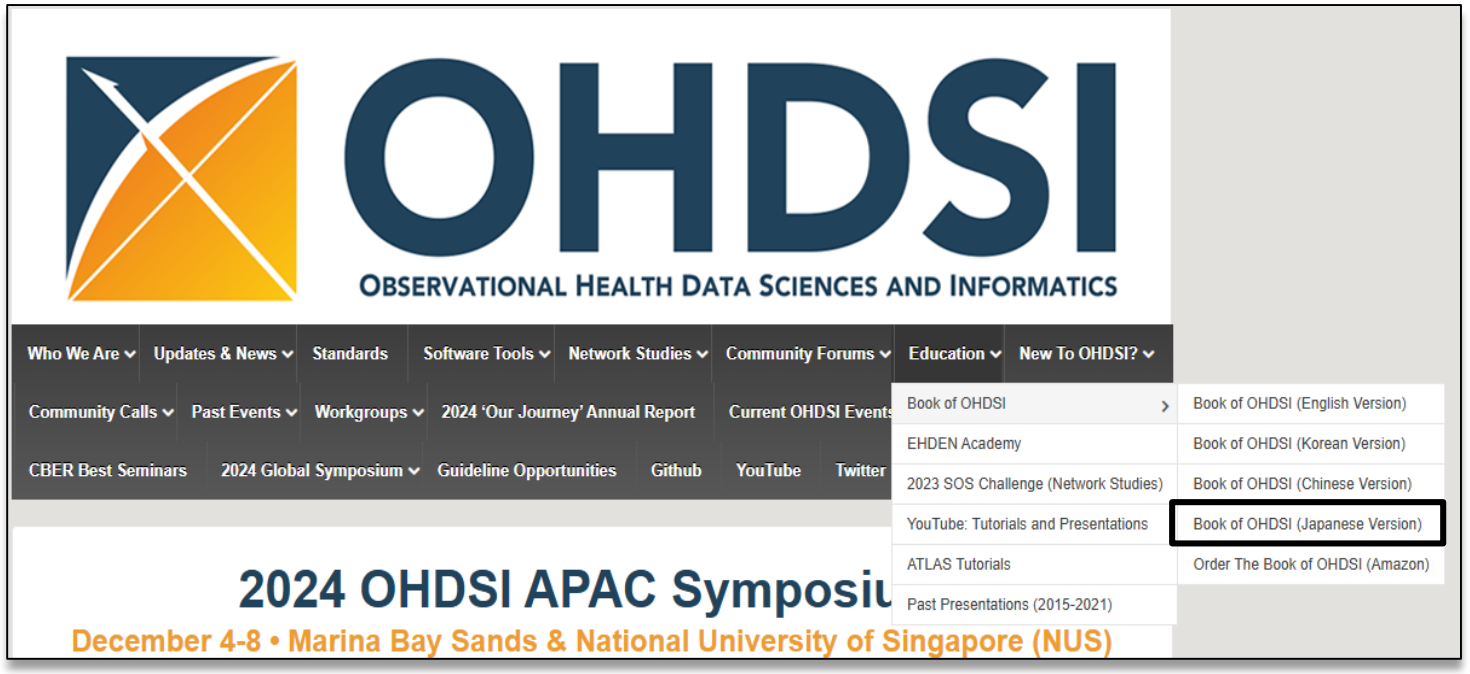


# OHDSI APAC News

**Congratulations!**

**The Book of OHDSI Japan has been published!!**

https://ohdsi.github.io/TheBookOfOhdsiInJapanese/



# OHDSI APAC News

- First APAC newsletter of 2025 has been released
- Entirely prepared by Keiko Asao from Japan – thank you for your contribution!

https://mailchi.mp/ohdsi/apac-2025-q1-newsletter

▶ ***Spotlight on Japan: ATLAS Training Workshop Held in Tokyo***

- On February 25, 2025, a hands-on ATLAS training workshop was successfully held at the Tokyo Metropolitan Industrial Trade Center.
- Sixteen participants joined the 4-hour session to learn about OMOP-CDM fundamentals and gain practical experience in cohort creation and data analysis using ATLAS.
- The event was hosted by FedAna and OHDSI Japan, with operations led by Kappa Medical. The training aimed to cultivate future ATLAS users in Japan and expand local capacity for real-world evidence generation.


# TRUST

IMPROVING HEALTH OUTCOMES THROUGH TRUSTED DATA EXCHANGE

trustplatform.sg

***"Trusted Research and Real world-data Utilisation and Sharing Tech"***

OHDSI APAC Community Call
May 2025

**Jointly developed by:**

MINISTRY OF HEALTH SINGAPORE

MINISTRY OF DIGITAL DEVELOPMENT AND INFORMATION

synapxe

GOVTECH SINGAPORE

# Agenda

- What is TRUST?
- How we enable safe and efficient data sharing and analytics
- How to get started
- OMOP journey

TRUST

IMPROVING HEALTH OUTCOMES THROUGH TRUSTED DATA EXCHANGE

# What is TRUST

‘Trusted Research and Real world-data Utilisation and Sharing Tech’

# TRUST is a data framework and analytics platform to enable anonymised health analytics by researchers from public and private sectors

One stop to request, access and analyse data

TRUST

IMPROVING HEALTH OUTCOMES THROUGH TRUSTED DATA EXCHANGE

Secure cloud analytical environment within Government Commercial Cloud

**Real-World Data**
in Vault or Govt agencies' repository

Anonymised and transfer

Real-World Data

Strategic Research Data

Anonymised and transfer

**Strategic Research Data**
in their own repositories or data aggregators e.g. A*STAR BiomedDAR

Anonymised linked datasets for analysis

ICA · Inland Revenue Authority of Singapore · Housing & Development Board · Health Promotion Board · Ministry of Health Singapore · Ministry of Education Singapore

TTRaCT · CADENCE · GUSTO · PRECISE · SingCLOUD · STCC

*"Trusted Research and Real world-data Utilisation and Sharing Tech" Platform*

# How TRUST addresses key data challenges faced by researchers

**UNCLEAR DATA ACCESS RULES**

↓

**CLARIFY PERMISSIBILTY OF USE; OPEN UP ACCESS**

- Data permissibility rules and governance for key datasets has been clarified
- Streamlined **pre-agreements** with data custodians and users
- Establish **central Data Access Committee** for streamlined and efficient data approval

**VARIED DATA SECURITY & INFRASTRUCTURE**

↓

**NATIONAL DATA-EXCHANGE PLATFORM**

- Established **secure environment on Government Commercial Cloud** for data linkage, access and analysis
- Established **Trusted Third Party to enable linkages across datasets** and anonymisation tool according to MOH anonymisation standards.

**LACK DATA STANDARDS**

↓

**DATA CATALOGUE & INTEROPERABILITY**

- Adopt **internationally recognised data standard** (e.g. OMOP)
- A central data curation team has been set up and OMOP mapping work is ongoing

# Enabling high value health-data analytics research

## Evaluate social determinants of health to improve cardiovascular health outcomes

CADENCE
Cardiovascular Disease National Collaborative Enterprise

Generate new insights into determinants that influence cardiovascular health and equity. Guide better designed interventions for impactful and sustainable cardiovascular outcomes, through analysis of clinical-lifestyle-social data.

## Understand COVID-19 genomic risk factors in disease severity to guide future intervention strategies

Agency for Science, Technology and Research
SINGAPORE

PREPARE

Assess the prevalence and allele frequencies of host genetic variants determining the susceptibility and severity of SARS-CoV-2 infections as well as in vaccine effectiveness. Provide insights to future measures and policies to safeguard those at higher risk of infections.

## Unlock value of population cohorts to gain deeper insights to Asian precision medicine

Multi Ethnic Cohort (MEC)

SingHEART

SEED
SINGAPORE EPIDEMIOLOGY OF EYE DISEASE

HELIOS
Health for Life in Singapore

PRECISE

Enable next phase of PRECISE/SG100k Precision Medicine studies in diseases such as cardiovascular, metabolic, neurological, psychiatric, ophthalmologic, as well as rare diseases. Enable improved risk prediction, risk assessment and interventions through precision population health approaches.

## Study long-term risks of diseases and overall healthcare cost impact on Gestational Diabetes Mellitus (GDM) mothers to guide appropriate care

GUSTO
GROWING UP IN SINGAPORE TOWARDS HEALTHY OUTCOMES

Evaluate mother-child pairs with a history of GDM and their increased risk for diseases, including developing mental disorders. Findings will allow health-care providers to formulate strategies to appropriately follow-up, screen and treat GDM mothers and their children.

# Orchestrate and enable safe and expeditious health data analytics in the HHP research ecosystem

## Promulgated data sharing principles and best practices


## Established pre-agreements with 12 Public Research Organisations to enable expeditious data access (as at 22 April 2024)

A*STAR Agency for Science, Technology and Research
Duke-NUS Medical School
CRIS
MOHT MOH Office for Healthcare Transformation
Nanyang Technological University Singapore
National Healthcare Group
NUHS National University Health System
NUS National University of Singapore
SIT Singapore Institute of Technology
SMU Singapore Management University
SUSS Singapore University of Social Sciences
SingHealth

## Support broader types of analytics e.g. genomics & low/no code analysis

lifebit

## Initiated national level effort to harmonise data standard

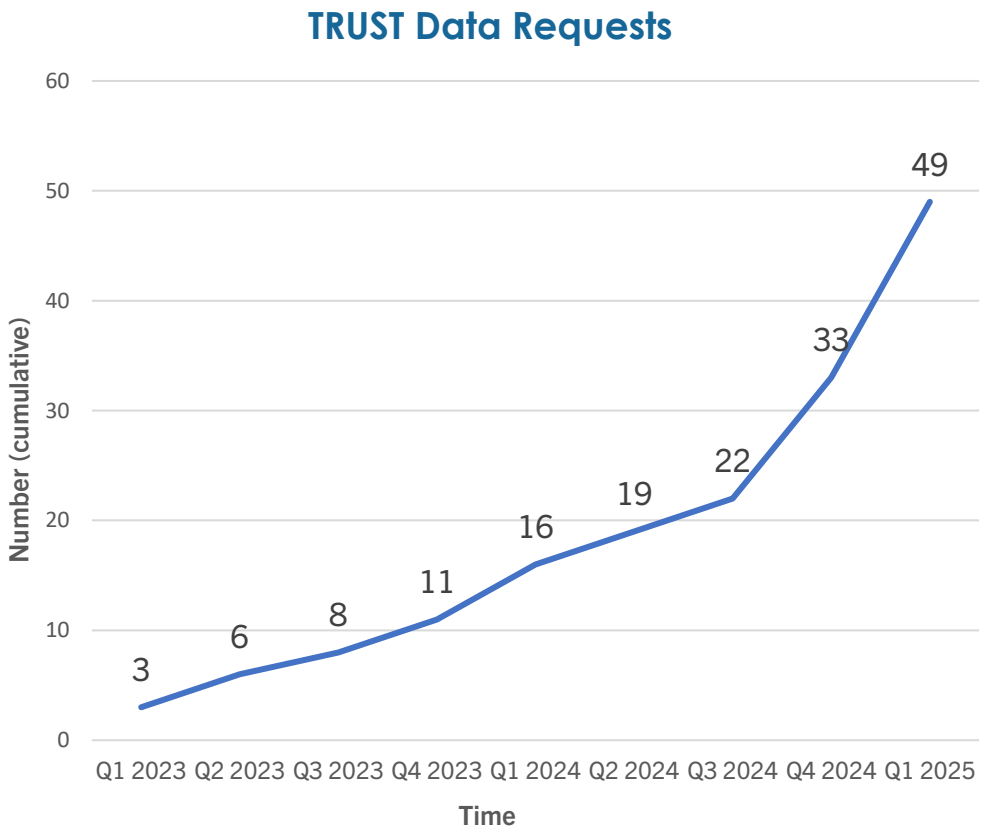
- **Key clinical data domains[1] have been mapped to** OMOP CDM[2]
- **Established partnerships to setup national curation team** on OMOP mapping

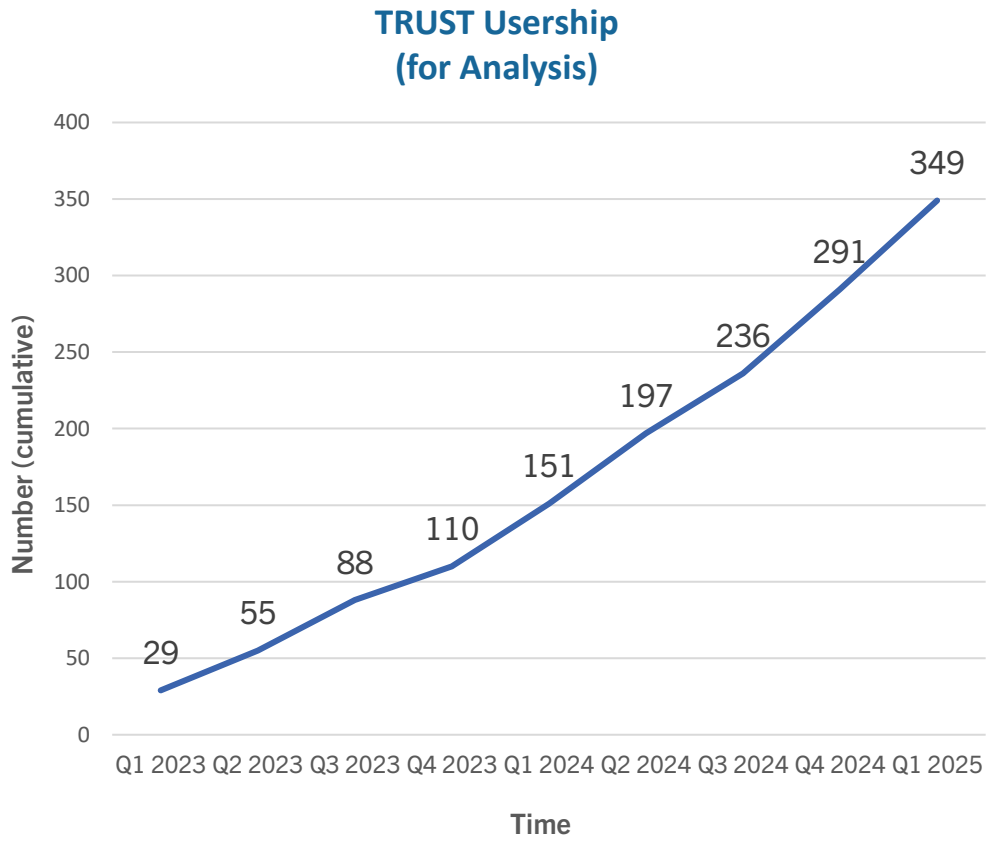
[1]*Domains (from NEHR) include Demographics, Diagnosis, Medication, Visits, Labs, Radiology, Procedures.*

[2]*Observational Medical Outcomes Partnership Common Data Model (OMOP CDM).*

# Gaining traction and growth in usership + use cases (data up to end March 2025)

## Close to 50 projects supported and 350 users



# New opportunities

## Increasing Impact

- Enable unstructured data (e.g., free text clinical notes, retinal images) and broaden data types (e.g. geospatial)
- Support strategic industry partners

- Future-proof with Privacy Preserving Tech (e.g. federated analysis)
- Enhance interoperability with other Trusted Research Environments local and internationally

## Increasing Interoperability

## Enhancing experience

- Scaling and automation (e.g. output checking)
- Enable self-serve (e.g. data exploration & visualisation)
- Develop the TRUST Academy, comprising training curriculum for both Users and Data Contributors (e.g. governance, best practices, data science)

# How do we enable safe and efficient data sharing and analytics

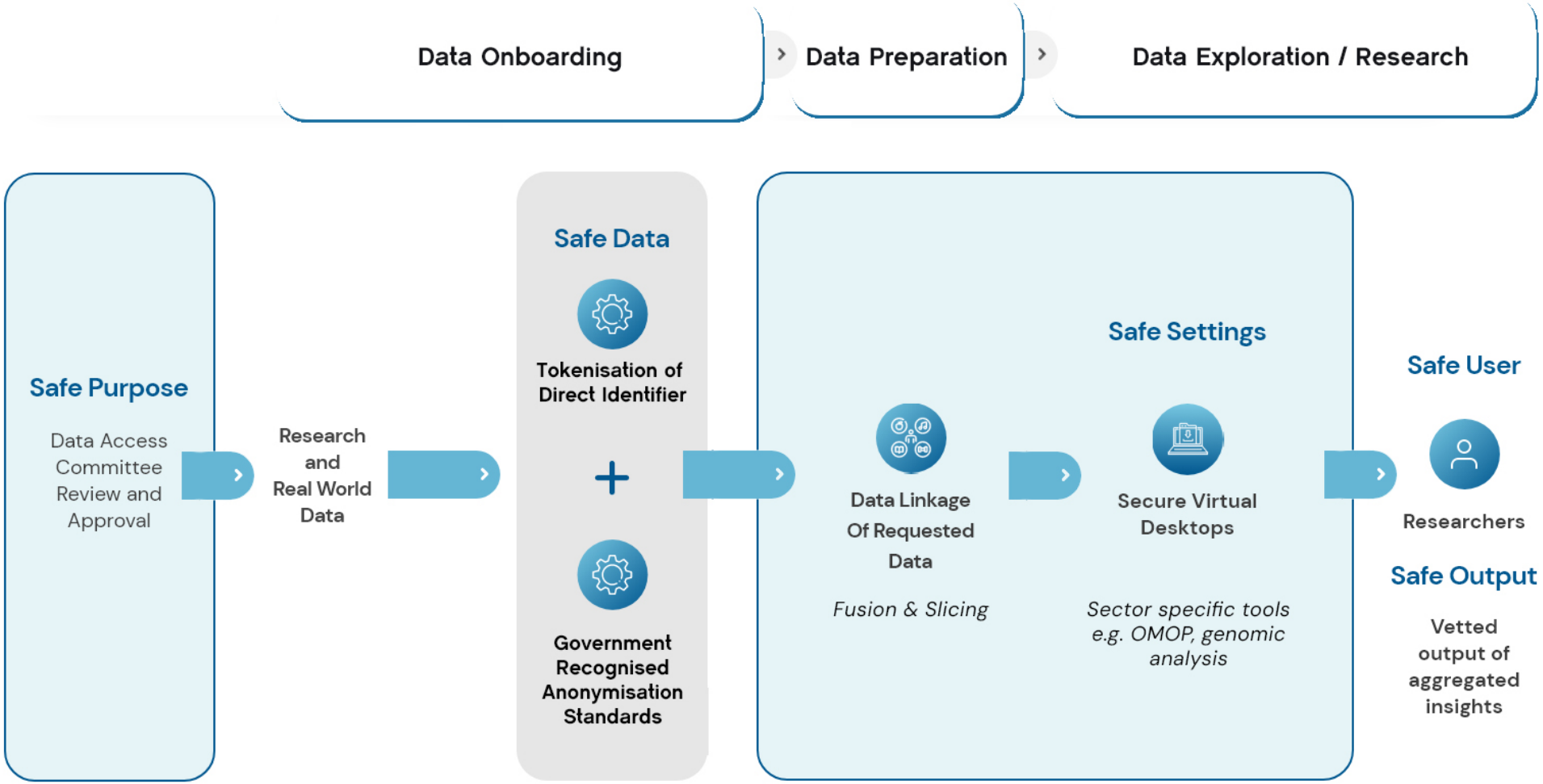
# TRUST's core features are built on the 5 Safes Framework, ensuring safe data access


Deploy synergistic policy and technical solutions across the data lifecycle

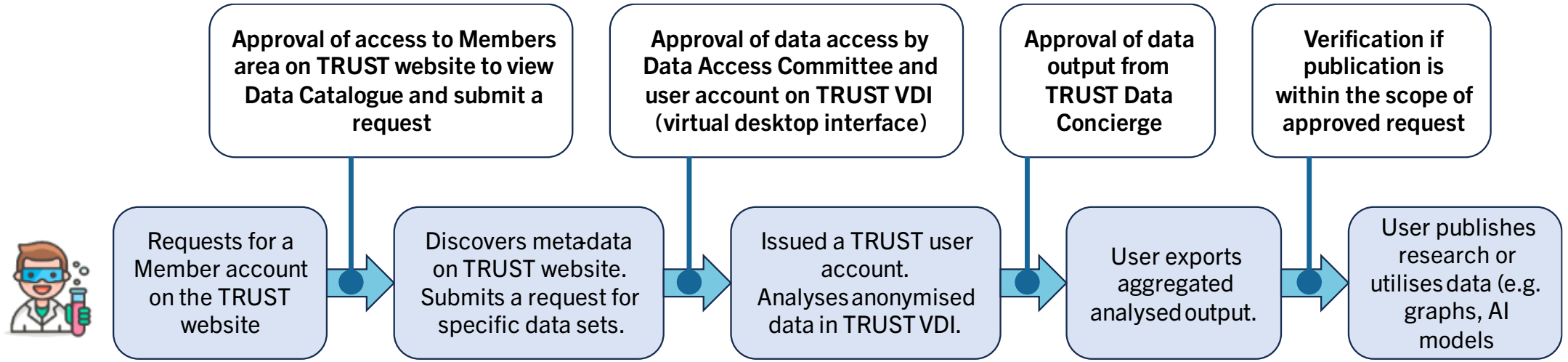
Balance privacy & public interest with safe use of data

Improved health outcomes & better care delivery

# TRUST adopts the Five Safes Framework



# Ensuring Data User's expeditious and safe access




a. Approval of data output by TRUST Data Concierge.

- All output must comply with TRUST's output policy, which states that **these must be aggregated/de-identified, generated based on data from at least 5 individuals.**

b. Verification that public release of analysis is within scope of approved request.

- Researchers who wish to publish insights generated from their research on TRUST are required to **submit their publications to TRUST DAC secretariat for pre-publication review.**

# TRUST Data Access Committee (DAC)

## Government Data Reps

Mr Philip Ong (Chair)
DS (Development), MOH

Mr Lai Kai Bin
DD, GDD, SNG/MDDI

Ms Lim Yi Ding
D, DOS TC

## Public Research Organisation / Healthcare Cluster Data Reps

A/Prof Ngiam Kee Yuan
GCTO, NUHS

A/Prof Yeo Khung Keong
Dy GCMIO (Research), SHS

A/Prof Tan Cher Heng
GCRO, NHG

Prof Chng Wee Joo
Vice President (Biomedical Science Research), NUS

Dr Sebastian Maurer-Stroh
Executive Director, BII, A*STAR

Prof John Chambers
Prof, CVD Epi, NTU
CSO, PRECISE

Prof. Roger Vaughan
D, CQM & CSSD, Duke-NUS

## Domain Experts

Prof Julian Savulescu
D, Centre for Biomedical Ethics
(Ethics Domain)

Prof Simon Chesterman
Vice Provost (Educational Innovation), NUS (Legal Domain)

Ms Ai Ling Sim-Devadas
DD (Advocacy & Engagement), LKCSOM, NTU (Layperson Domain)

Mr Rajakanth Raman
ED, Rainbow Across Borders
(Layperson Domain)

# How to Get Started

# How to get started

For researchers from our Public Research Organisations*, you will be able to access TRUST if you are:

- An employee of an institution that has signed the *Data Request Agreement* with TRUST; and
- A bona fide researcher (verification will take place through *Pubmed ref*, *ORCID ID*, CV or institution profile page); and
- Have a verified institution email account

Complete data request received by every last Friday of the month will be assessed and reviewed by TRUST DAC within 4-6 weeks. Data would be provisioned on TRUST for access 6-10 weeks# post approval.

*Data provisioning could be delayed by complexity of the data requested and the data availability from the data contributor

# Onboarding process post-DAC approval

TRUST Operations will reach out to initiate the onboarding process.

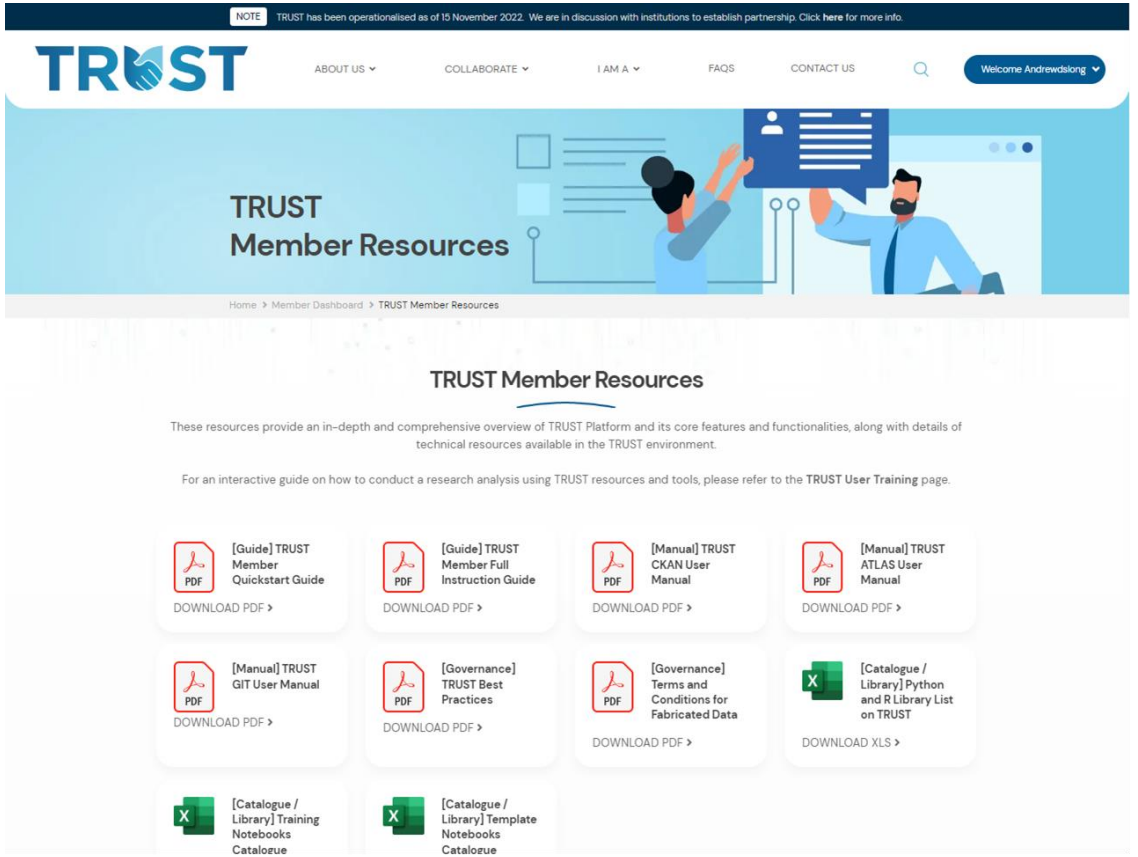
1. User Onboarding Meeting - the onboarding meeting provides an overview of the next steps and what users can expect during the onboarding process.
   The agenda will cover:
   - TRUST Workspace & trustplatform.sg member account
   - BYOD data - TRUST Data Anonymisation & Conveyance
   - User Support
   - Availability of online user resources & monthly Foundational Training
   - Tools & Libraries
   - Projected timeline on completion of onboarding process.

2. Guides
   - TRUST will provide documentation to support users e.g. TRUST Member Quick Start Guide

# User Onboarding, Training and Support

- Since launched, TRUST has made available **40 datasets**, approved nearly **50 data requests** and supported close to **350 users** on their research analytics.
- We are actively engaging researchers to understand their research questions and to support their data needs.
- Users are supported with an onboarding programme by the TRUST team, augmented with additional resources available through the TRUST portal
    - User guides and Onboarding sessions
    - Step-by-step video tutorials
    - Community / peer forum*

*to be launched in 2025




A *TRUST Data Concierge* team supports users throughout their journey

> *"We are extremely grateful towards the TRUST Support team for their generous support and great responsiveness and guidance, thank you!"*
> **- Dr Chen Wenjia, SSHSPH, NUS (first batch of TRUST users)**

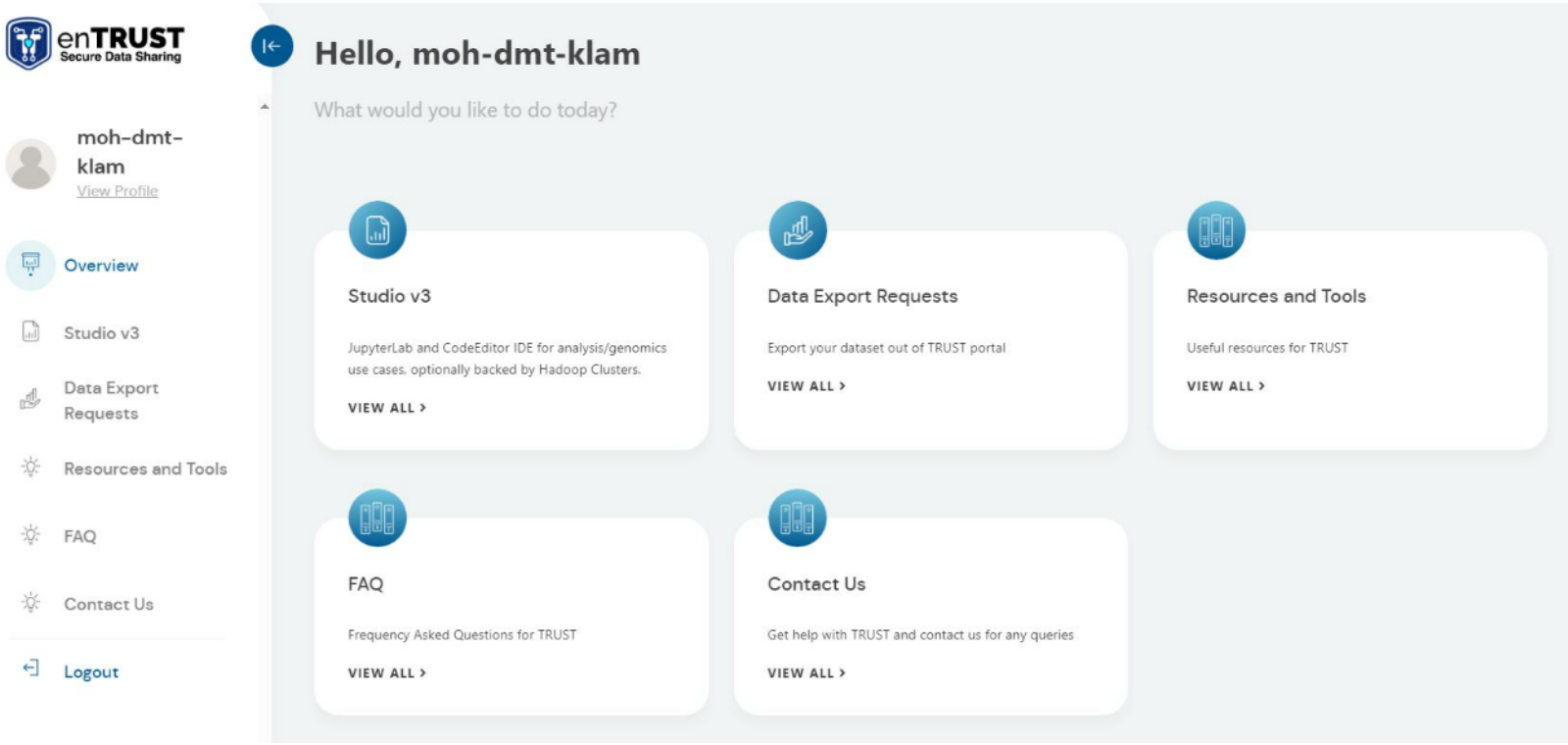
# TRUST portal as launch point for user to access various features and functions

- Key Features and Functions

- R, Python and Spark access via Sagemaker Studio, CodeEditor IDE

- Low/no code data exploration and analytical tools via Lifebit Platform (coming soon Q4 2024)

- Requests within air-gap environments:
  - General requests/enquiries from within the portal
  - Export requests of analytical insights



# Recommended data science competency for TRUST users

To carry out data analysis on TRUST, it is highly recommended that users should have a base level of competency in the following areas.

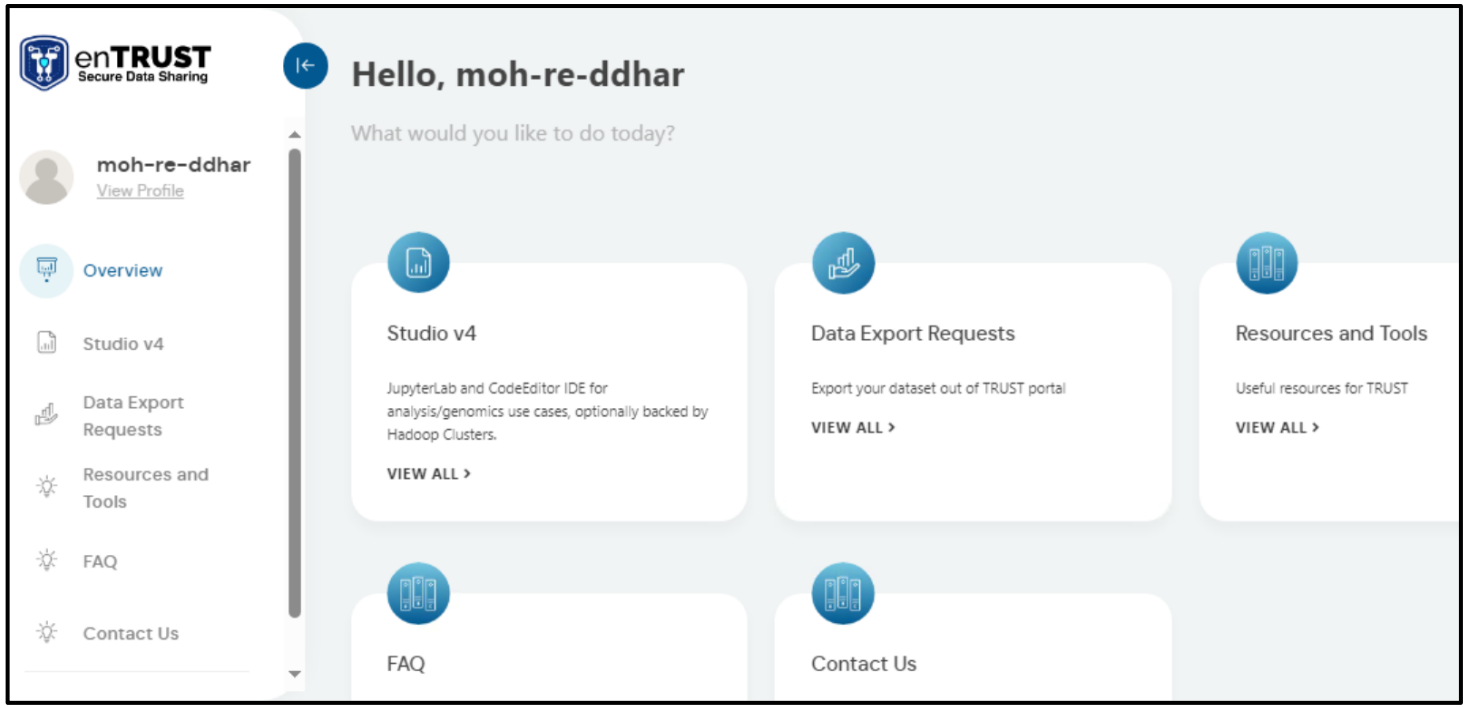
| | Recommended Competency | Required For |
|---|---|---|
| Strongly Recommended | Some familiarity with Python and / or R syntax | For TRUST data research |
| | Some experience with handling Data Science notebook interface | For TRUST data research |
| | Some knowledge of how to work with Linux command line | For data discovery and transferring files from S3 bucket to notebook instance |
| | Some knowledge of using Cloud storage | For provisioning resources and understanding the cost |
| Optional | Some knowledge of using Cloud storage | For big data research analysis |
| | Some familiarity with Pyspark and / or SparkR syntax | For big data research analysis |

# enTRUST Portal Overview Page

User Journey 1 – How to Conduct Research Analysis on enTRUST Platform?

| File Type | Training Topic | Version | Duration | Filesize |
|---|---|---|---|---|
| ▶ | enTRUST Overview | 3.0 | 2min 50sec | 70MB |

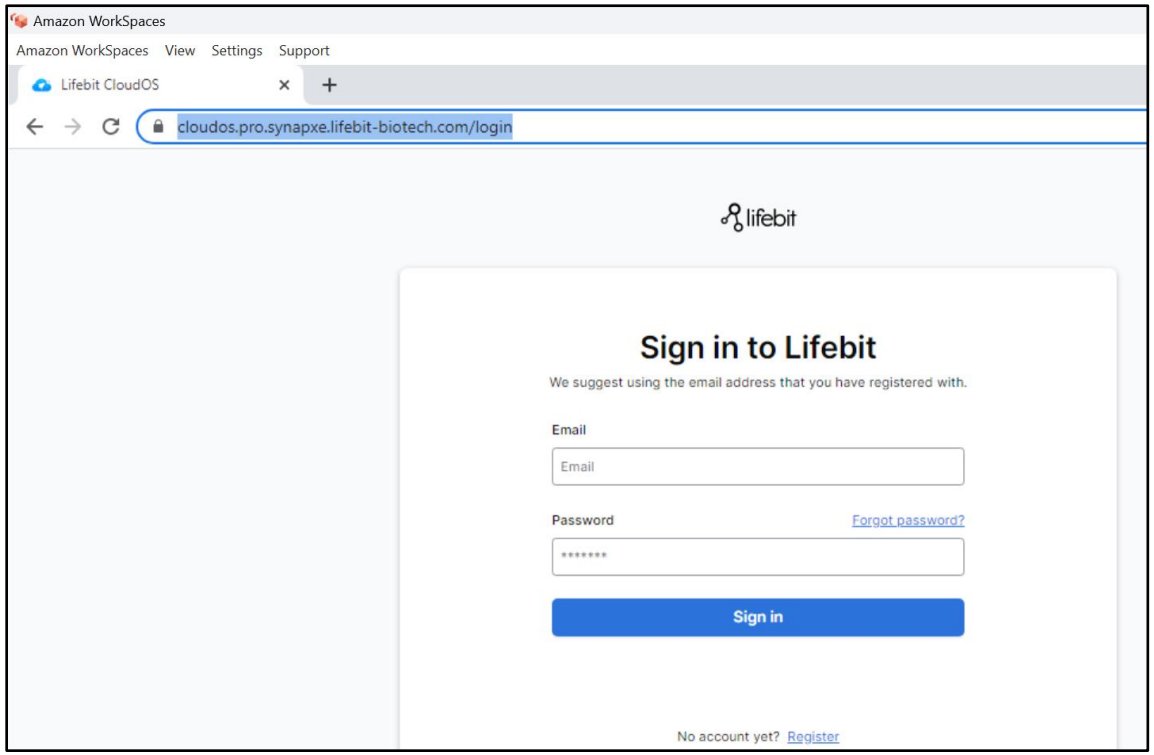
## enTRUST Overview

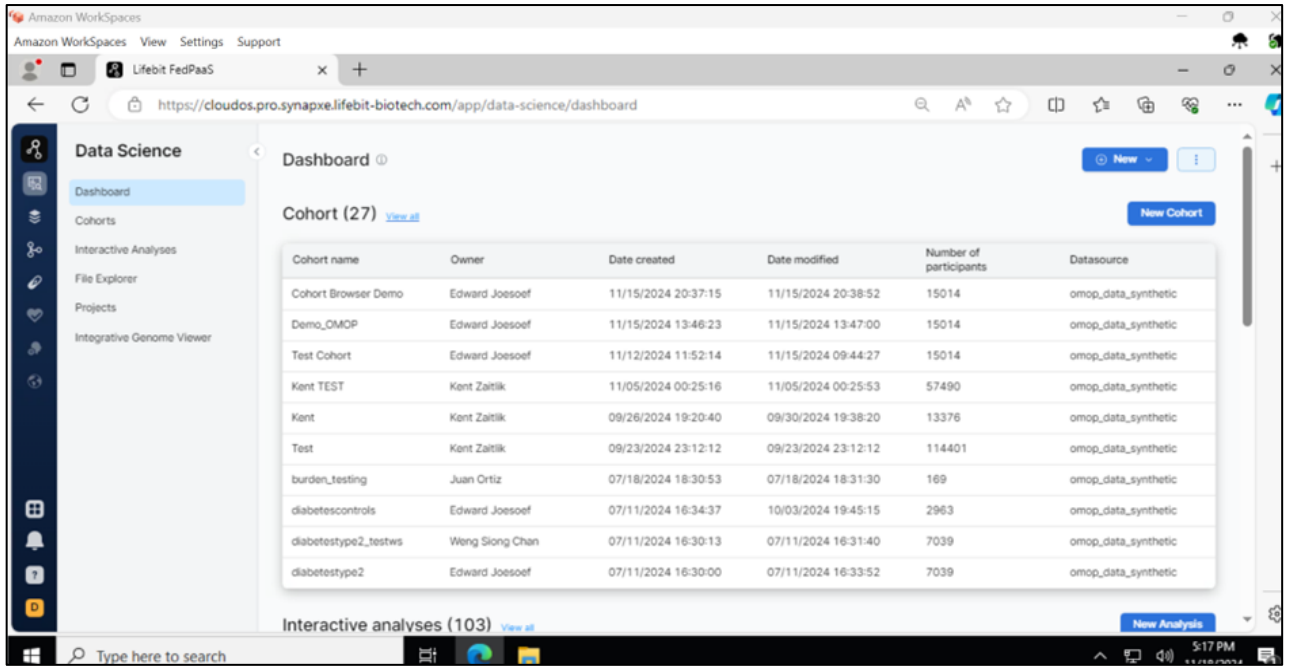


# Lifebit Overview

| Lifebit | | | |
|---|---|---|---|
| **File Type** | **Topic** | **Duration** | **Filesize** |
| ▶ | Lifebit Overview | 2min 15sec | 29MB |

Lifebit Overview



# Charging Framework

## Key considerations

1. **Takes into account users' diverse research needs**, offering flexibility to deploy the required computing resources.
2. **Drives responsible use of TRUST resources** while enabling flexibility, scalability and customisation for users to decide on the computing resources for their research.
3. **Transparent approach** where charges are listed down clearly, based on usage & base operational fees.

**TRUST** will cover the core infrastructure development, comprising:

- Manpower to operate TRUST
- Platform Development
- Part of TRUST Foundational Services Fees

**Users** will cover:

- Compute, storage & usage costs: Amazon Web Services (AWS) cloud charges on actual usage
- Part of TRUST Foundational Services Fees

# Components of TRUST Charging Framework fees borne by Users

## i. Base Package

Workspace cost

a) Users can select a **preconfigured 160 hours** of a selected instance.

## ii. Value Added Services

Comprises:

a) Big data processing tools e.g. EMR* (available now)
b) Sector tools e.g. Lifebit (available in Q4 FY2024)

*Amazon EMR (previously called Amazon **Elastic MapReduce**) is a managed cluster platform that simplifies running big data frameworks on AWS to process and analyse vast amounts of data.

## iii. Foundational Services Fees

Comprises platform implementation, security and data processing related costs*

- Platform implementation – cloud costs incurred to prepare the TRUST platform for use
- Security – cyber security costs against cyber threats and security breaches
- Data processing related costs – refers to the processing of TRUST data or BYOD (e.g. storage, management, processing, cleaning and transfer).

Timeline for charging out Foundational Services Fees:

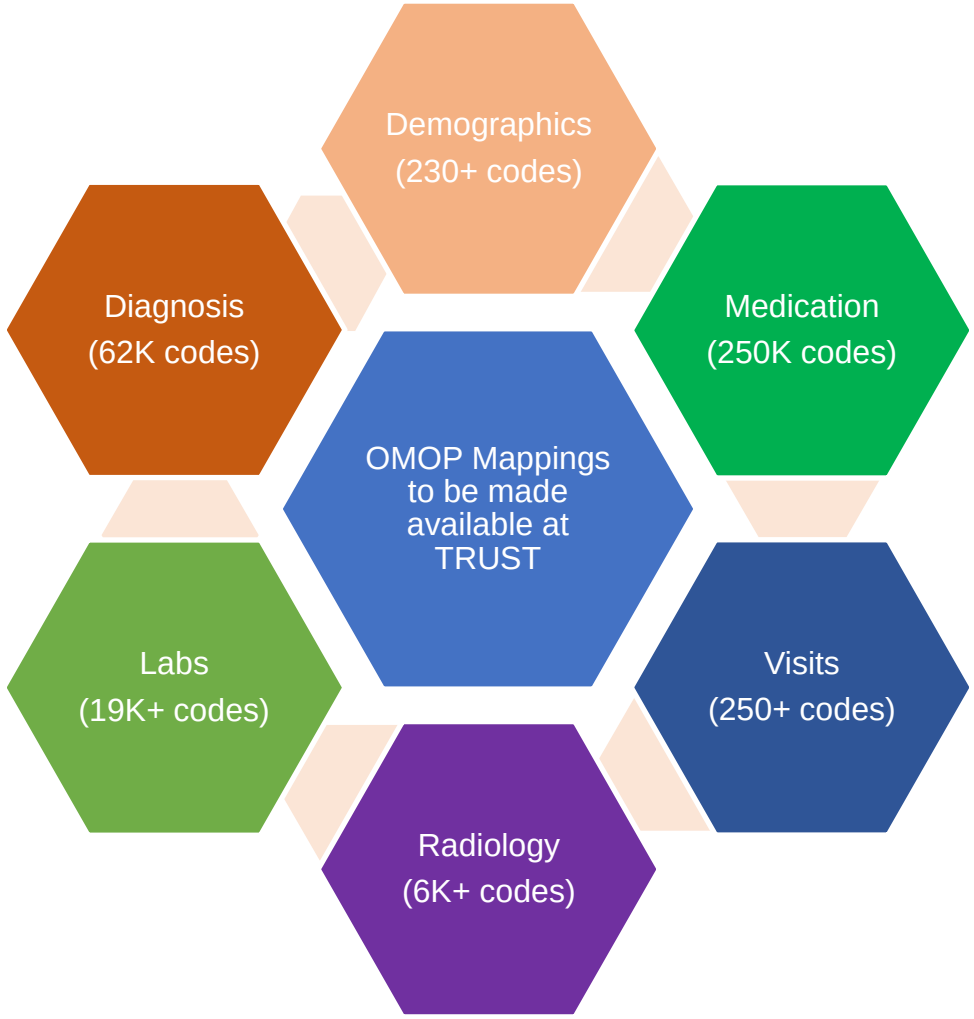
- Phase 1 (FY 2024 – FY 2025) – 20%
- Phase 2 (FY 2026 – FY 2029) – 50%
- Phase 3 (FY 2030 onwards) – Full fees apply

# OMOP journey

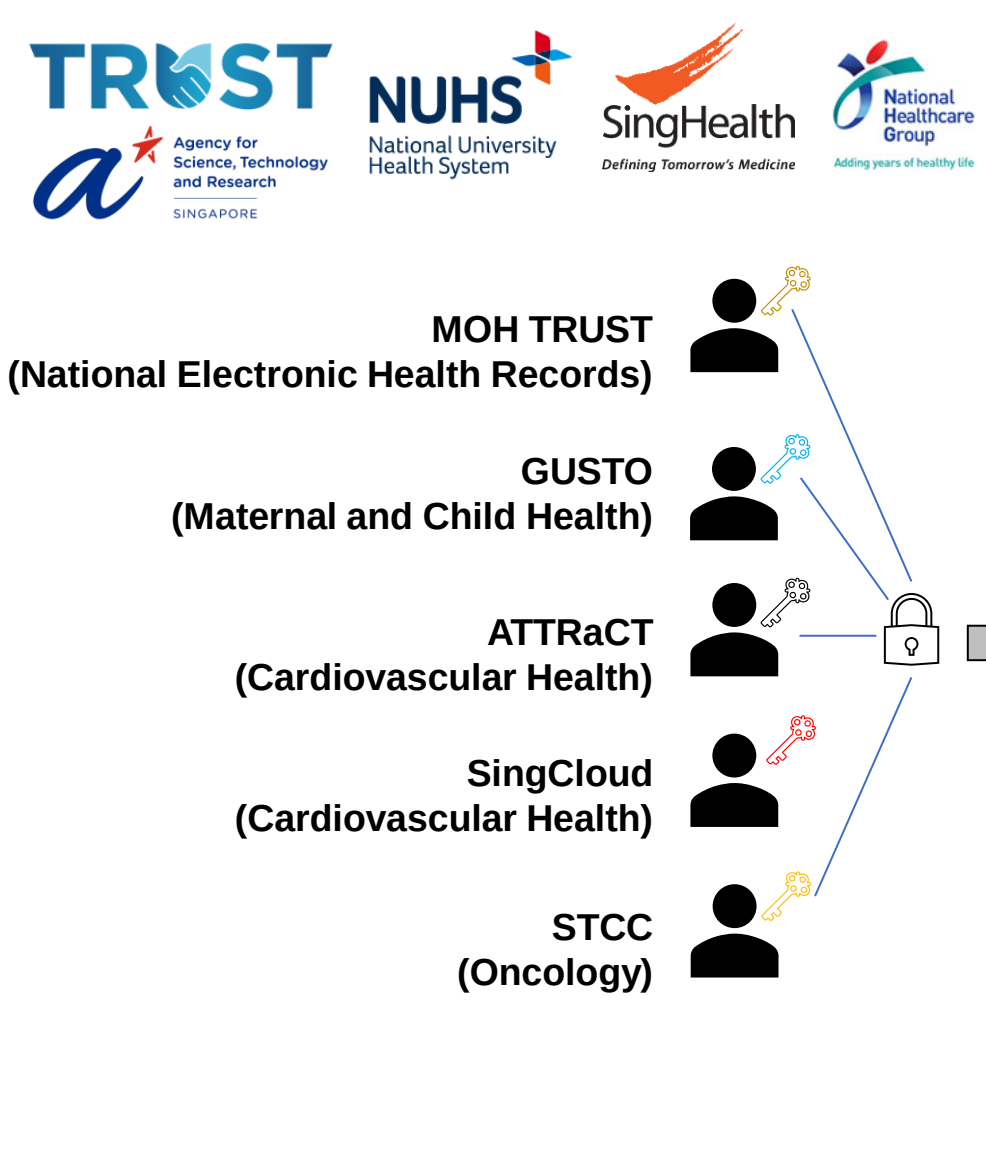
# TRUST OMOP Approach

All-of-Singapore **National Electronic Health Records** mapping efforts:

- Mapping and harmonisation of demographics, diagnoses, medications, visits, laboratory tests, and radiology codes completed.
- Standardised ICD/SNOMED/SDD codes to OMOP terminology across all healthcare institutions.
- Achieved 98% of DQD of mapping to ensure accuracy and relevance.
- Mapping codes now available centrally to enable whole-of-country data harmonization (see next slide)



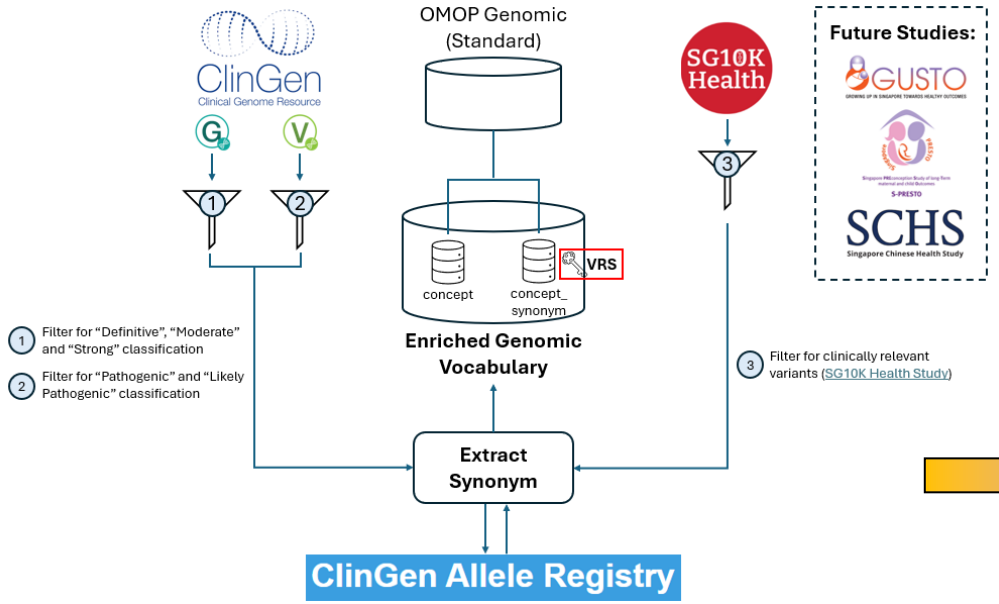
# Singapore's Ecosystem Harmonisation of OMOP CDM Master Mapping Table



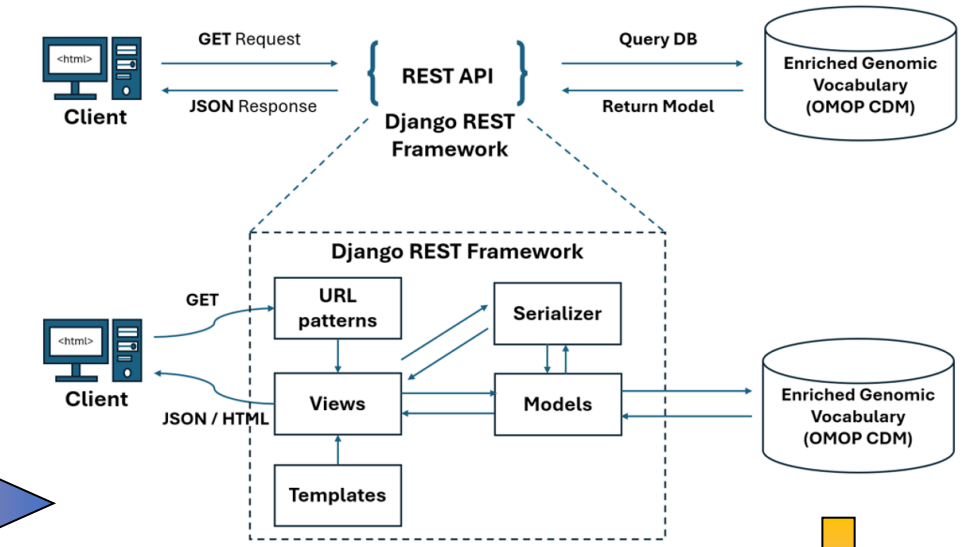

First release of the harmonized OMOP master mapping table (h1.0) and consolidated master mapping table across sites (v1.0) was in March 2025.

Significant milestone on national level OMOP mapping harmonisation in Singapore.

# Embarked on Singapore's OMOP Genomics CDM Strategy

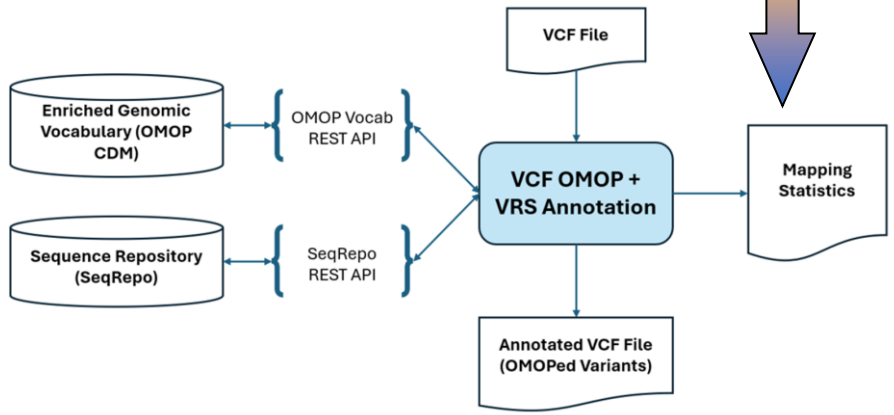



*A workflow to enrich genomic vocabulary from public sources or local genomic cohorts*




*Genomic vocabulary query*

The framework provides a solution for genomic data mapping in OMOP CDM for local and international collaborations.




*Annotation of genomics VCF file with OMOP Concept ID and VRS computed identifier*

# Potential Research Study Topics with OHDSI Community

- Use of GLP-1 receptor agonists and adverse risk outcomes (building upon the findings from OHDSI APAC Symposium 2024, Singapore).

- AI-based prediction of post-surgical risk complications in patients planning for an elective surgery (building upon the findings from IMAGINE AI 2024 Datathon, Singapore).

# Thank you

Questions?


# The Book of OHDSI in Japanese

OHDSI APAC Community Call
05/15/2025

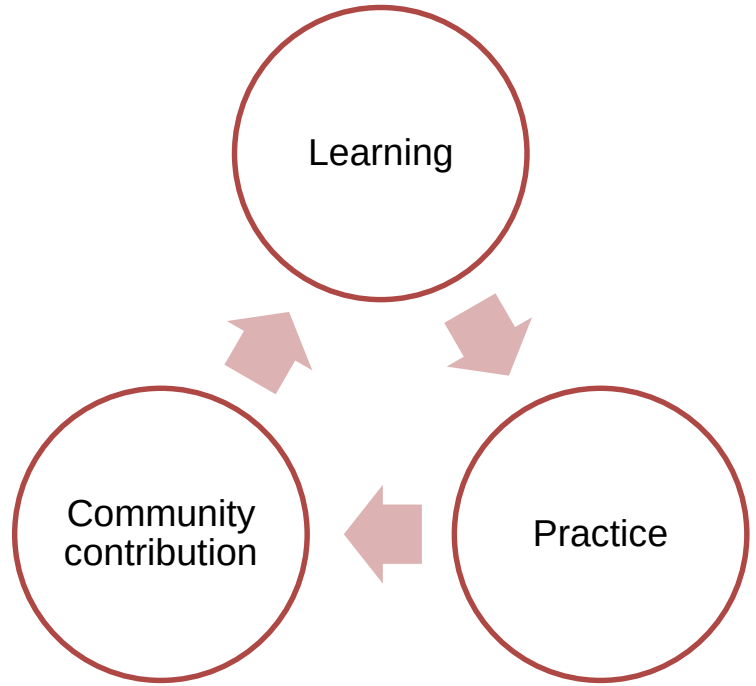
Keiko Asao

# Contributors to the Japanese Edition

- Supervisors: Shigemi Matsumoto and Tatsuo Hiramatsu
- Work: Keiko Asao, Tomoko Kobayashi, Motoi Miura.
- Machine translation program: Martijn Schuemie
- Supported by: Department of Real World Data R&D
  Graduate School of Medicine, Kyoto University

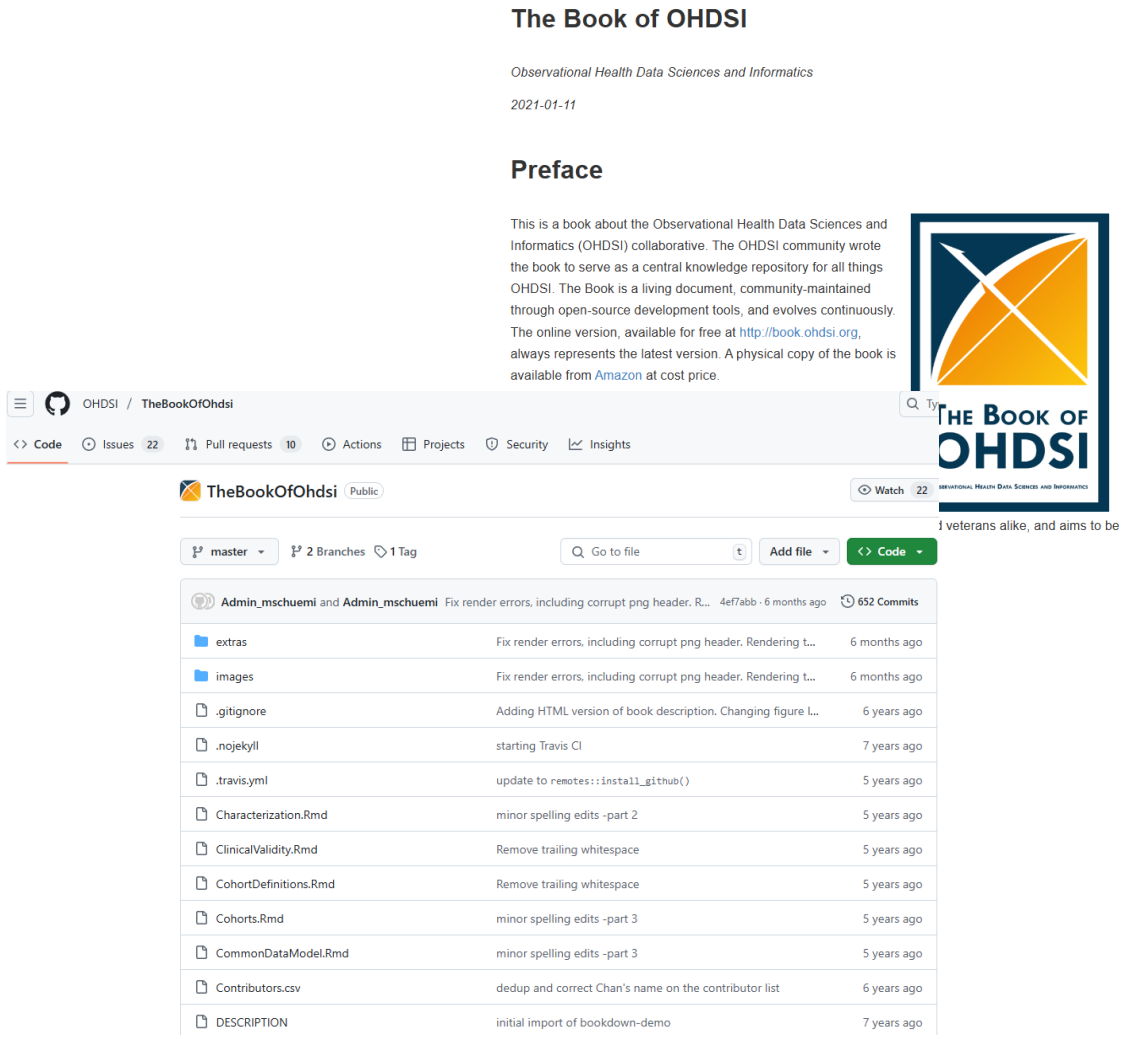
# What is The Book of OHDSI?

- OHDSI's official text book
- A practical guide to research using OMOP CDM and OHDSI tools
- Suitable for a wide range of readers, from beginner to advanced
- Co-authored by more than 50 OHDSI community members
- Written in R Markdown and published free of charge on GitHub (Creative Commons Zero)



# Overview: The Book of OHDSI in Japanese

- Translation project to promote understanding of OHDSI/OMOP CDM
- Original (English): approx. 115,000 words, 464 pages



# R Markdown

index.Rmd | WhereToBegin.Rmd

Knit on Save | Knit

Source | Visual

```
# Where to Begin {#WhereToBegin}

*Chapter leads: Hamed Abedtash & Kristin Kostka*

> "A journey of a thousand miles begins with a single step." - Lao Tzu

The OHDSI community represents a mosaic of stakeholders across academia, industry and
benefits a range of individuals and organizations, including patients, providers, and
systems, industry, and government agencies. This benefit is achieved by improving bot
analytics as well as the usefulness of healthcare data to these stakeholders. We beli
which benefits greatly from disruptive thinking. We actively seek and encourage fresh
work. \index{community}

## Join the Journey
Everyone is welcome to actively participate in OHDSI, whether you are a patient, a he
someone who simply believes in our cause. OHDSI maintains an inclusive membership mod
requires no membership fee. Collaboration is as simple as raising a hand to be includ
count. Involvement is entirely at-will. A collaborator can have any level of contribu
from someone who attends weekly community calls to leading network studies or OHDSI w
have to be data holders to be considered active members of the community. The OHDSI c
researchers, health care providers and patients & consumers alike. A record of collab
periodically updated on the OHDSI website. Membership is fostered via OHDSI community
chapters.\index{join the journey} \index{workgroups} \index{chapters}

```{r jointhejourney, fig.cap='Join the journey - How to become an OHDSI collaborator
fig.align='center'}
knitr::include_graphics("images/whereToBegin/joinTheJourney.png")
```


### OHDSI Forums
The OHDSI Forums[^forumUrl] is an online discussion site where OHDSI Community Collab
form of posted messages. The forums consist of a tree-like directory structure. The t
can be divided into categories for the relevant discussions. Under the categories are
```

## Chapter 2 Where to Begin

*Chapter leads: Hamed Abedtash & Kristin Kostka*

> "A journey of a thousand miles begins with a single step." - Lao Tzu

The OHDSI community represents a mosaic of stakeholders across academia, industry and government-entities. Our work benefits a range of individuals and organizations, including patients, providers, and researchers, as well as health care systems, industry, and government agencies. This benefit is achieved by improving both the quality of healthcare data analytics as well as the usefulness of healthcare data to these stakeholders. We believe observational research is a field which benefits greatly from disruptive thinking. We actively seek and encourage fresh methodological approaches in our work.

### 2.1 Join the Journey

Everyone is welcome to actively participate in OHDSI, whether you are a patient, a health professional, a researcher, or someone who simply believes in our cause. OHDSI maintains an inclusive membership model. To become an OHDSI collaborator requires no membership fee. Collaboration is as simple as raising a hand to be included in the yearly OHDSI membership count. Involvement is entirely at-will. A collaborator can have any level of contribution within the community, ranging from someone who attends weekly community calls to leading network studies or OHDSI working groups. Collaborators do not have to be data holders to be considered active members of the community. The OHDSI community aims to serve data holders, researchers, health care providers and patients & consumers alike. A record of collaborator profiles are maintained and periodically updated on the OHDSI website. Membership is fostered via OHDSI community calls, workgroups and regional chapters.

# Translation Approaches Considered for the Japanese Translation

| Approach | Advantages (Pro) | Issue (Con) |
|---|---|---|
| PDF/HTML-based translation (Korean and Chinese version method) | • Relatively easy to start<br>• Can be translated close to the display format | • Figure, formula, and code structures are hard to handle.<br>• Difficult to maintain the composition of the original work |
| GitHub + GPT-4o (French version method) | • Faithful to source file (rmd)<br>• Parallel processing of translation and structural adjustment | • Depends on output accuracy of ChatGPT<br>• Manual verification and correction required. |
| Using GitLocalize | • GitHub-based for easy translation management<br>• Easy to handle updates | • R Markdown support is limited<br>• Technical setup required |

Martijn, thank you!

https://github.com/OHDSI/TheBookOfOhdsiInFrench/blob/main/extras/TranslateBook.R#L80

# FYI: What is GitLocalize?

- Continuous localization tool for GitHub repositories
- Automatically links the original text to the translation, so you can see the extent to which the original text is affected by the update.
- Useful for frequent updates.


[Introducing GitLocalize: A Continuous Localization Tool for GitHub Repositories | GitLocalize blog](#) (in Japanese)

# Translation Process

## Step 1: Translation and post-editing

- GitHub + GPT-4o
- Create a terminology list to ensure consistency first
- Post-editing for R markdown format
- Post-editing for Japanese readability

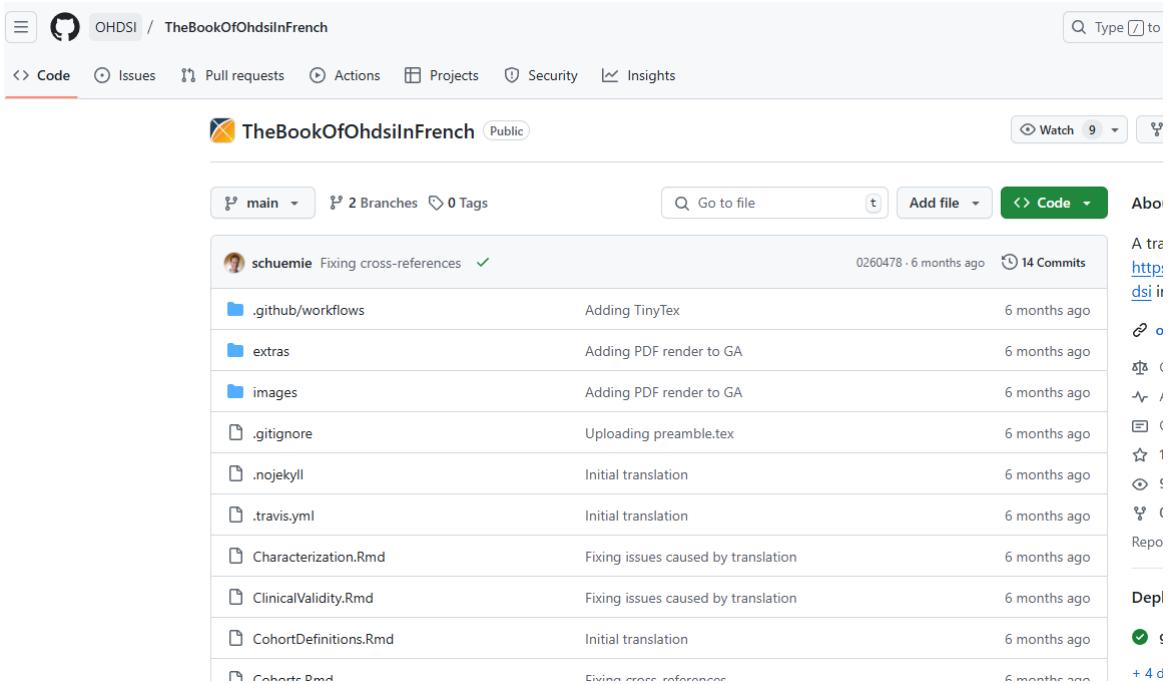
- (No translation for Figures)

## Step 2: Publication and corrections

- Published in HTML and PDF
- Going to accept revision and correction suggestions via GitHub and other methods (email, etc.)

# GitHub + GPT-4o

- [TheBookOfOhdsiInFrench/extras/TranslateBook.R at main - OHDSI/TheBookOfOhdsiInFrench]
- Martijn Schuemie created The Book of OHDSI translation prompt with GPT-4o while keeping the R programming part.




https://github.com/OHDSI/TheBookOfOhdsiInFrench/blob/main/extras/TranslateBook.R#L80

# Terminology List to Ensure Consistency

- How to translate OHDSI terminologies?
  - Listing of keywords and frequently used terms (about 200 words)
  - Regional community leaders′ reviews and inputs
- Translation should vary for the same terminology depending on context.
- Usage fluctuations even within the original text (occasionally).

- Examples of difficult translation

| | |
|---|---|
| Observation | Observations, observations (as domain) |
| Procedure | Procedure |
| Visit | Length of visit (Visit) |
| Encounter | Seeing a doctor |
| The Book of OHDSI | OHDSI Books |

# Post-editing For R Markdown File

## How it looks in HTML

本例で使用されるコホートは、既に ATLAS に作成されていると仮定します (Chapter (**ref?**)(Cohorts) で説明)。付録には、ターゲットコホート (Appendix (**ref?**)(AceInhibitorsMono), (**ref?**)(ThiazidesMono)) およびアウトカムコホート (Appendix (**ref?**)(Angioedema), (**ref?**)(Ami), (**ref?**)(ARBUse)) の完全な定義が提供されています。


The correction should conform to the style of the original text.

–翻訳作業中– OHDSIの書

本例で使用されるコホートは、既に ATLAS に作成されていると仮定します (Chapter 8 で説明)。付録には、ターゲットコホート (Appendix 19.2, 19.5) およびアウトカムコホート (Appendix 19.4, 19.3, 19.9) の完全な定義が提供されています。

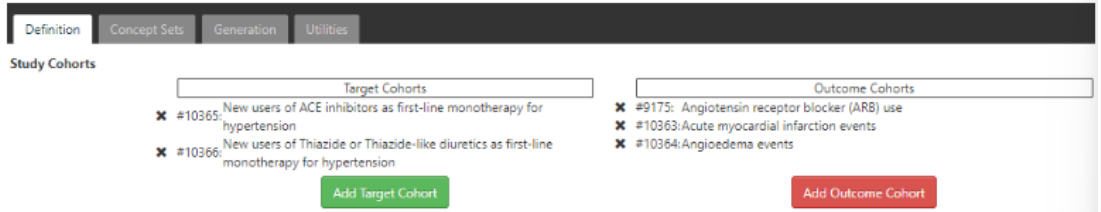



図 9.22: ターゲットおよびアウトカム定義の発生率。

## R markdown format

```
400
401 ### 設計
402
403 本例で使用されるコホートは、既に ATLAS
    に作成されていると仮定します (Chapter @ref(Cohorts)
    で説明)。付録には、ターゲットコホート (Appendix
    @ref(AceInhibitorsMono), @ref(ThiazidesMono))
    およびアウトカムコホート (Appendix @ref(Angioedema),
    @ref(Ami), @ref(ARBUse)) の完全な定義が提供されています。
404
405 ```{r atlasIncidenceCohortSelection,
    ig.cap='ターゲットおよびアウトカム定義の発生率。',echo=FALS
```

```
400
401 ### 設計
402
403 本例で使用されるコホートは、既に ATLAS
    に作成されていると仮定します (Chapter \@ref(Cohorts)
    で説明)。付録には、ターゲットコホート (Appendix
    \@ref(AceInhibitorsMono), \@ref(ThiazidesMono))
    およびアウトカムコホート (Appendix \@ref(Angioedema),
    \@ref(Ami), \@ref(ARBUse)) の完全な定義が提供されています。
404
405 ```{r atlasIncidenceCohortSelection,
```

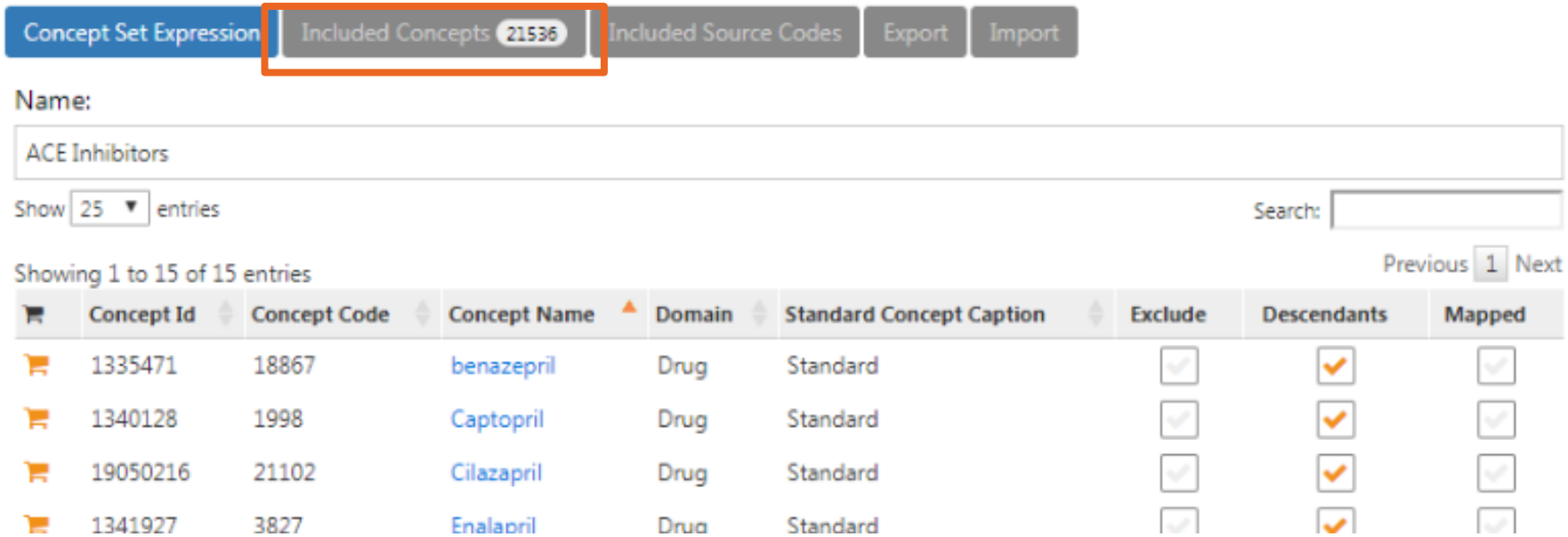
# Post-editing for Japanese Readability

- The contents are technically difficult.
  - Particularly the second half of Part III and Part IV (PheValuator, calibration of negative controls and effect estimates, positive controls, etc.).
  - Only when the content is understood can the translation be readable.

# Web Application Explanation

- As for the explanation of the web application, it is devised as the form of “English (Japanese)” so that it can be followed as the instructions of the application.

図 10.8 はコンセプトセット表現を示しています。対象とするすべてのACE阻害薬成分を選択し、その下位層すべてを含め、これらの成分を含むすべての薬剤を含めています。「Included concepts （包含されるコンセプト）」をクリックして、この表現に含まれている21,536のコンセプトすべてを確認することができ、「Included Source Codes（包含されるソースコード）」をクリックすると、様々なコーディングシステムに含まれるすべてのソースコードを探索することができます。



# Revisions and Corrections

- Intensively revised and corrected during the first three months from the publication.
- Please contact us at info@kappa-medical.com with any corrections you notice after reading the report.
- Ongoing revisions are described in the ReadMe on GitHub.

# Lessons Learned

- Team with members who are good at:
  - GitHub and R Markdown
  - Languages
  - OHDSI and OMOP contents
- Book structure in R Markdown
  - Edit to properly reflect chapter and section numbers, figure numbers, references, etc.
- PDF needs further adjustment
  - To use Japanese fonts, it was necessary to change the settings.
  - The index is shown in PDF only (not in HTML).
- Pitfalls of machine translation
  - Large translation omissions occur! (Note that paragraphs, table halves, etc., especially when one sentence is long.)
- Punctuation style was adjusted in Japanese version.
  - Example: Citations after the punctuation to ones before the punctuation.
  - Example: “Chapter X” was changed to “第X章” to make more natural in Japanese.

# Future Prospects: Making The Book of OHDSI Multilingual

- OHDSI global movement to update the Book with consideration for multilingualization.
- Including the localization into the scope when created the original may make the localization easier
  - Easy to understand, relatively short original text
  - Easy-to-translate structure (e.g., diagrams with instructions on the application for application descriptions)

# HTML and PDF available April 1, 2025

- https://ohdsi.github.io/TheBookOfOhdsiInJapanese/
- https://ohdsi.github.io/TheBookOfOhdsiInJapanese/TheBookOfOhdsiInJapanese.pdf