TRUST

IMPROVING HEALTH OUTCOMES THROUGH TRUSTED DATA EXCHANGE

trustplatform.sg

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

2024 OHDSI APAC Symposium

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

- OMOP journey

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

"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)

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).

TRUST
IMPROVING HEALTH OUTCOMES THROUGH TRUSTED DATA EXCHANGE

Currently, TRUST offers access to ~40 anonymised datasets across the following domains

List updated as of August 2024

Population
- Birth & Death
- Passes & Permits
- Citizenship & Residency

Social
- Economic
- Housing
- Education*

*subject to MOE approval

Health
- Polyclinic
- Patient Management
- Surgery
- Laboratory
- Accident & Emergency
- Case & Visit
- Drugs
- Diabetes
- Comorbidity
- Beds
- Radiology
- Diagnosis

Lifestyle / Preventive Care
- iQuit Smoking
- Leisure Time Activity
- Sleep Tracking
- Steps Challenge
- Screening
- Immunisation

Health Finance
- Community Assistance
- Billing & Subsidy
- Medical Claims

Disease Registry
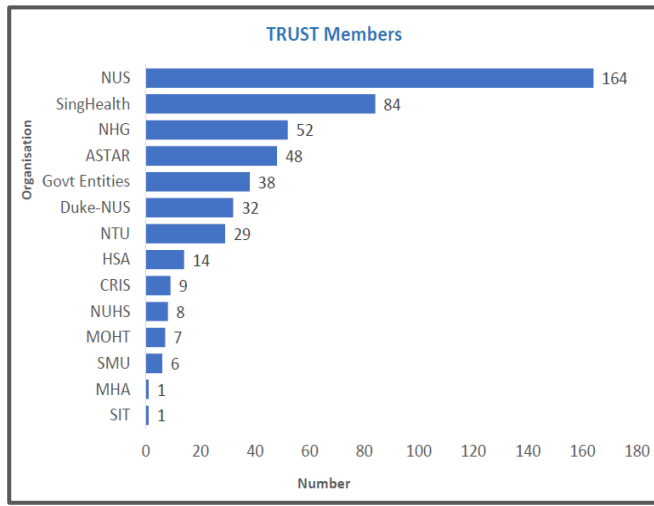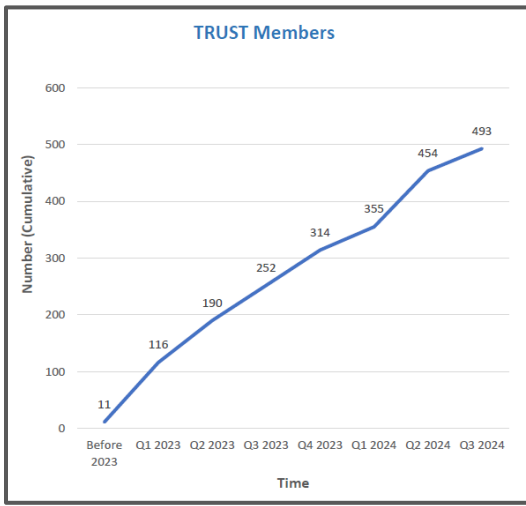- Myocardial Infarction
- Stroke
- Cancer
- Renal

Growth in usership + use cases *(data up to 30 August 2024)*

TRUST Membership

Registered Members on TRUST website for access to data catalogue, training materials & data request application form

Expanded by 39% since Q1 2024 with NUS (33%) comprising the majority

TRUST Datasets

TRUST Datasets available for request by TRUST Members

7 new datasets available in Q2 2024, mostly from ICA
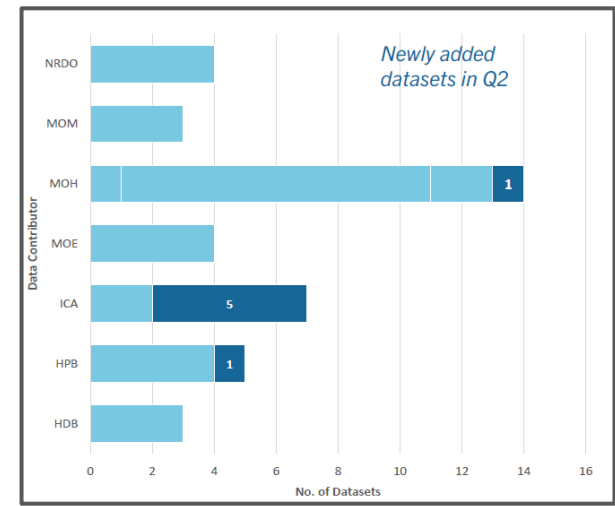
TRUST Use Cases

Approved TRUST Data Requests by TRUST DAC

22 approved use cases, with majority of use cases from NUS
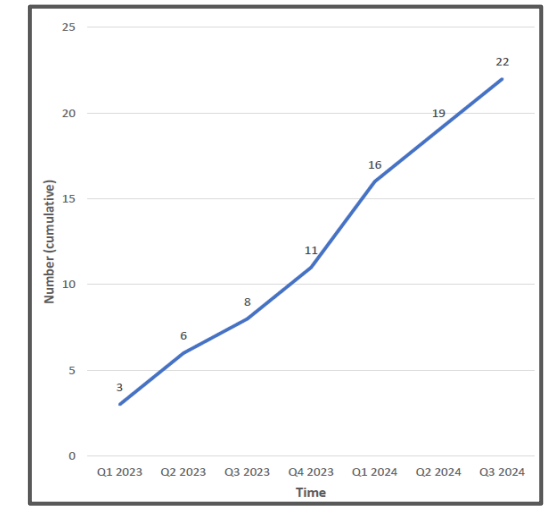
TRUST Usership

TRUST Users' access to data analytics portal based on approved TRUST Data Requests by TRUST DAC

Expanded by 65% since Q1 2024 with NUS (37%) users as the majority
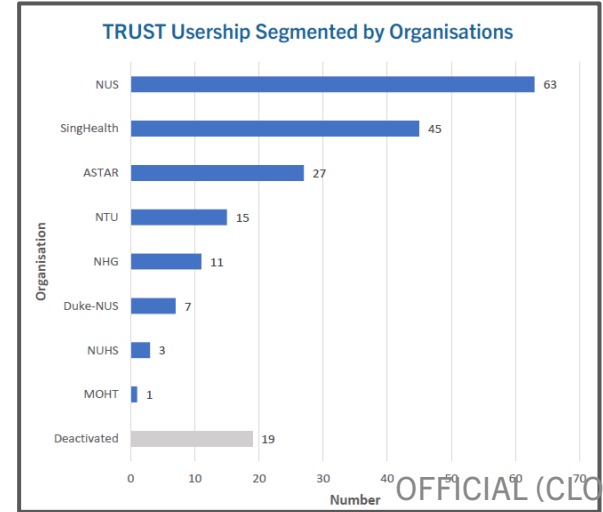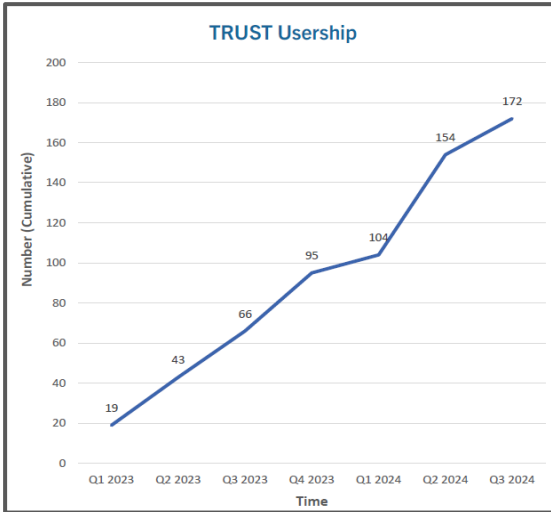
TRUST Use Cases (breakdown)

Approved TRUST Data Requests by TRUST DAC

More use cases in areas of Population Health (n=6) & Health Economics (n=8)
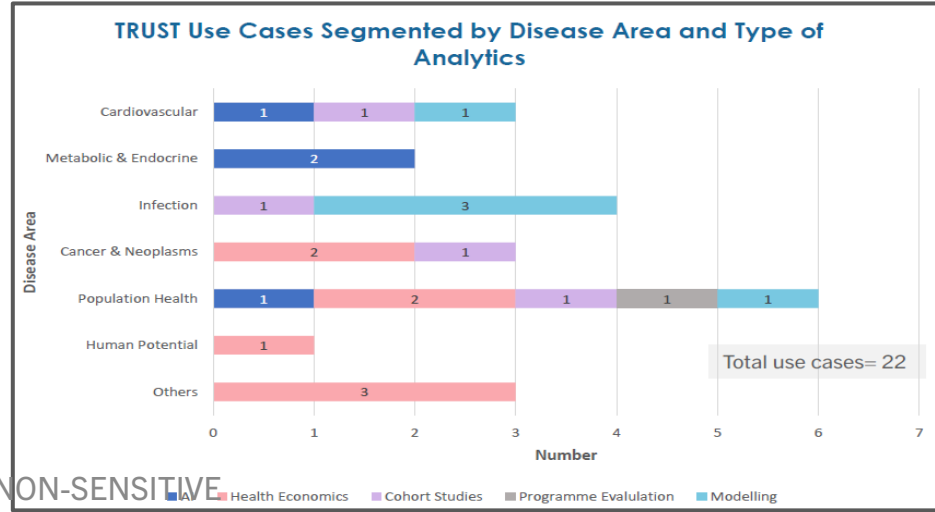
New opportunities

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)

TRUST portal as launch point for user to access various features and functions
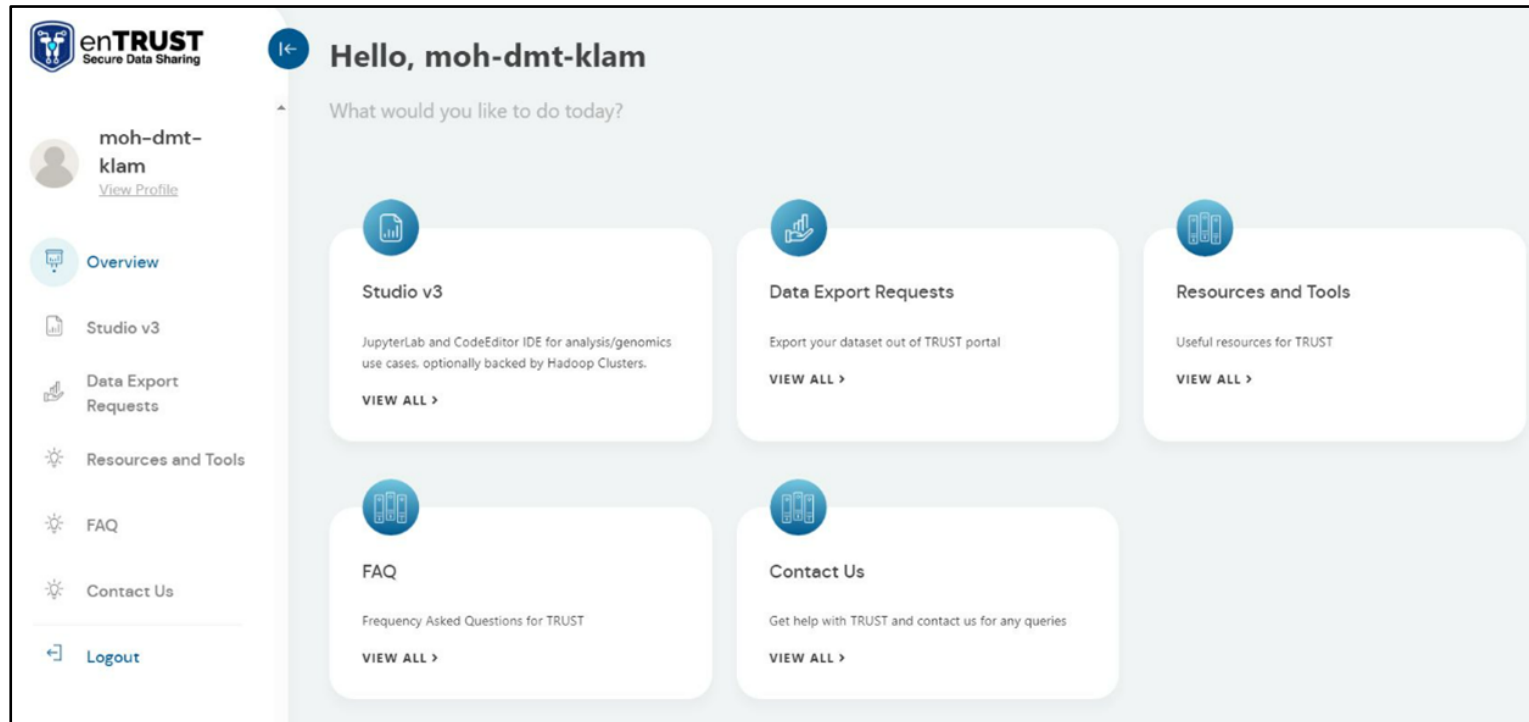
Key Features and Functions

- 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

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

SAFE PURPOSE	SAFE PEOPLE	SAFE SETTINGS	SAFE DATA	SAFE OUTPUT
All data requests will be reviewed by TRUST Data Access Committee to ensure that purpose of use fulfils public interest and social value.	TRUST users must have appropriate credentials for access to TRUST and the approved data for research.	TRUST is hosted in a secure environment with government-standard security measures.	All data accessed on TRUST are anonymised to government standards to reduce re-identification risks.	Only verified aggregate data and insights with low re-identification risk can be output.

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

Data Onboarding › Data Preparation › Data Exploration / Research

Safe Purpose

Data Access Committee Review and Approval

Research and Real World Data

Safe Data

⚙ Tokenisation of Direct Identifier

+

⚙ Government Recognised Anonymisation Standards

Safe Settings

Data Linkage Of Requested Data

Fusion & Slicing

Secure Virtual Desktops

Sector specific tools e.g. OMOP, genomic analysis

Safe User

Researchers

Safe Output

Vetted output of aggregated insights

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)

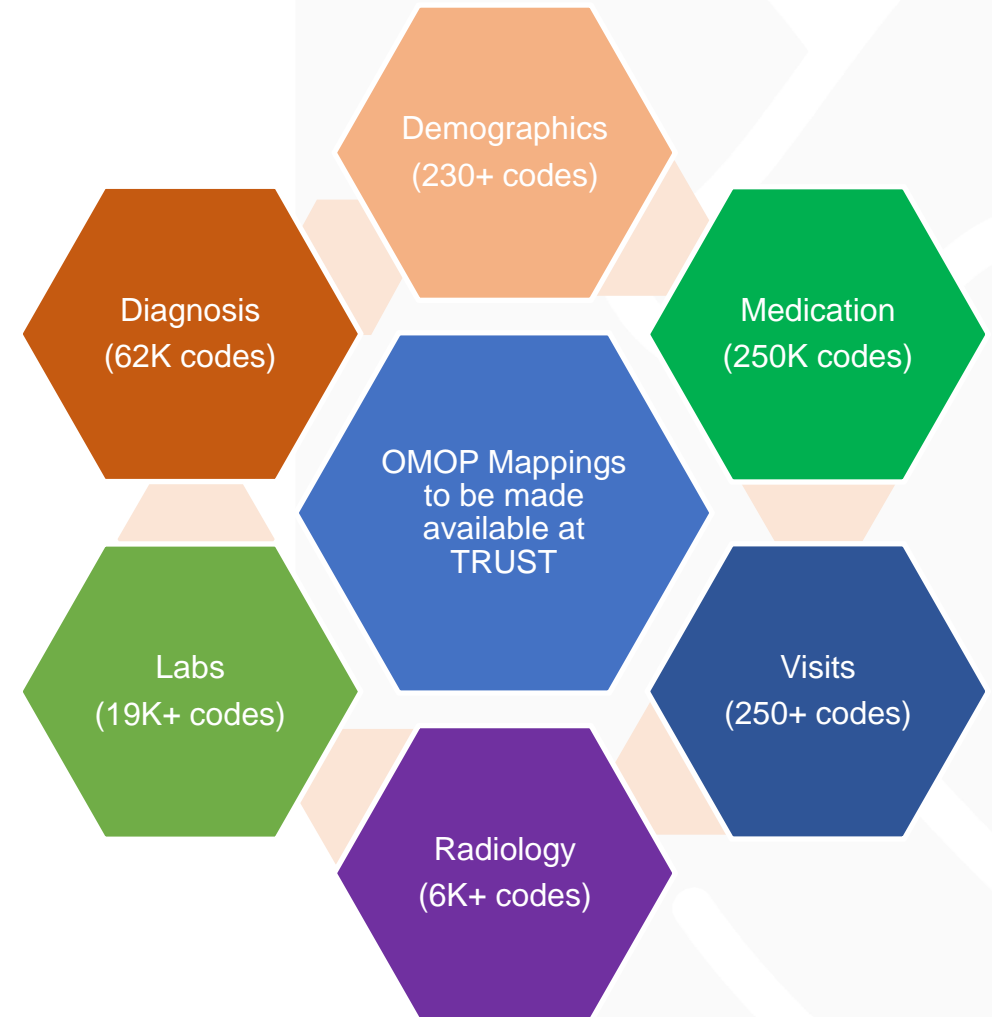Mr Rajakanth Raman
ED, Rainbow Across Borders
(Layperson Domain)

OMOP journey

TRUST OMOP Approach

All-of-Singapore EMR mapping efforts since 2021:

- 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.
- Mapping codes to be made available centrally in the TRUST platform to enable whole-of-country data harmonisation.
- Achieved 98% of DQD of mapping to ensure accuracy and relevance.

Demographics (230+ codes)

Medication (250K codes)

Diagnosis (62K codes)

OMOP Mappings to be made available at TRUST

Visits (250+ codes)

Labs (19K+ codes)

Radiology (6K+ codes)

Key challenges

- Scaling efforts
  - Bulk of data must be **manually mapped** and reviewed to OMOP standard concept codes ensuring data quality standards score remain intact.

- Clinical domain knowledge
  - **Lack of domain knowledge** among the data engineering team members to perform OMOP mapping. Regular or **continuous involvement of a clinician** is required until satisfactory DQD scores are achieved for the pertinent clinical domains.

- Data quality of source
  - Much effort have been put in to **improve the quality of data** before OMOP mapping, such as remove special characters, corrected typos, remove duplicates etc.,

Plans ahead - National Data Curation Team (NDCT) Initiative Overview

1. Formation of NDCT:
   - TRUST to establish NDCT in Singapore's A*STAR Bioinformatics Institute (BII) to
   - Facilitate OMOP mapping knowledge sharing
   - Grow a core team of OMOP expertise
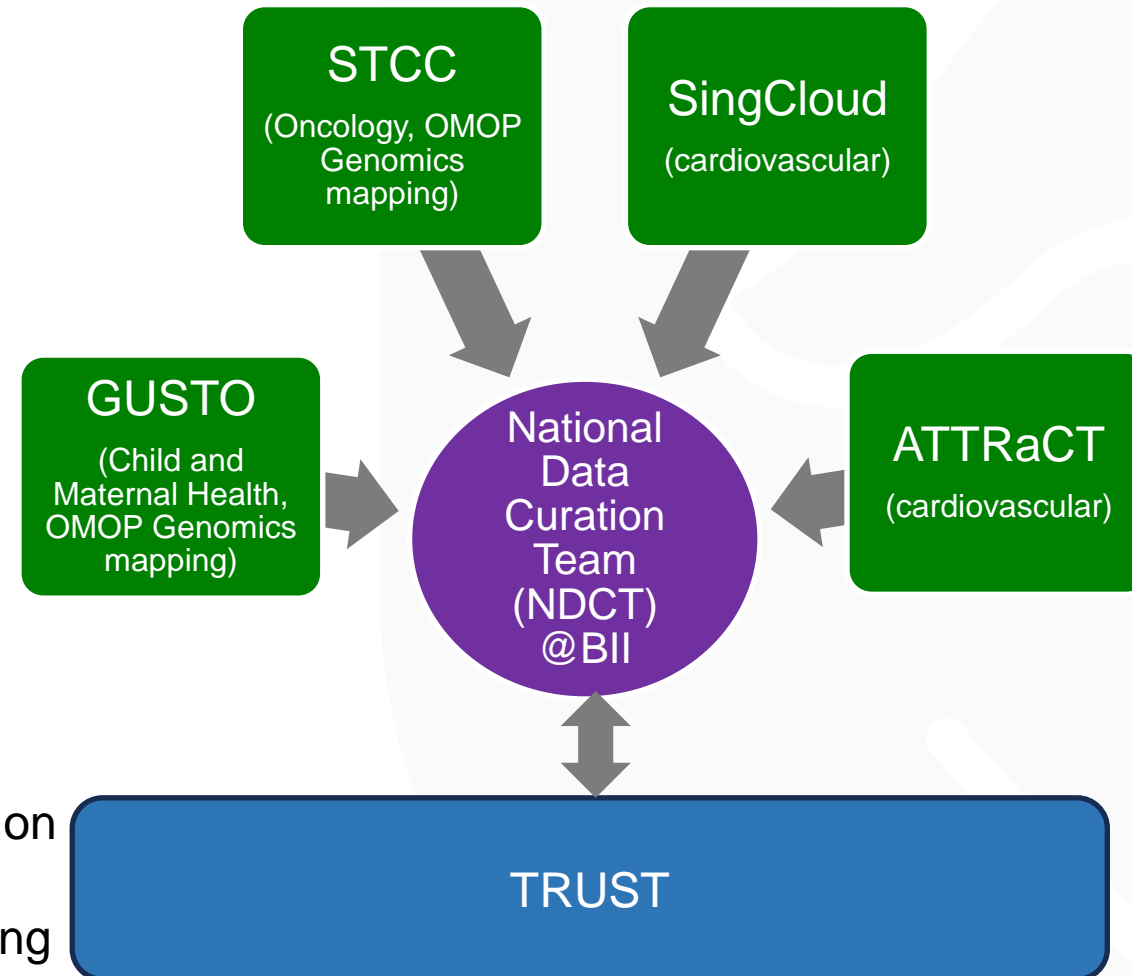   - Improve data curation efficiency

2. Initial Focus Areas:
   - Clinical domains identified for OMOP mapping:
   - Oncology
   - Child and Maternal Health
   - Cardiovascular
   - Genomics data mapping

3. Approach:
   - NDCT will be collaborating with clinicians to provide guidance on OMOP mapping codes, harmonisation process and share knowledge with research programmes to support OMOP mapping
   - TRUST to store all mapping codes centrally and establish secure API for OMOP code sharing, to be completed in Q3 2025

STCC
(Oncology, OMOP Genomics mapping)

SingCloud
(cardiovascular)

GUSTO
(Child and Maternal Health, OMOP Genomics mapping)

National Data Curation Team (NDCT) @BII

ATTRaCT
(cardiovascular)

TRUST

Thank you
Questions?