APAC Community Call

Training Session #4

May 18, 2023

Agenda

- OHDSI News
  - Vocabulary Landscape Assessment Results
  - New Community Dashboard
  - DevCon 2023
- Training Session #4
  - A*STAR Data Vault: a research platform for the Singapore ecosystem and beyond by Mukkesh Kumar and Cindy Ho

Vocabulary Landscape Assessment Results

- Community survey on the current status of OMOP standardized vocabularies was conducted early 2023
- Results from the survey were announced this month, full report is available at [Landscape Assessment Report Final.docx]

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

3. Executive Summary

We received 183 responses from community members and information about 60 data sources around the globe and talked to package maintainers and working groups. Overall, 87% of the community feels confident about Vocabularies' integrity.

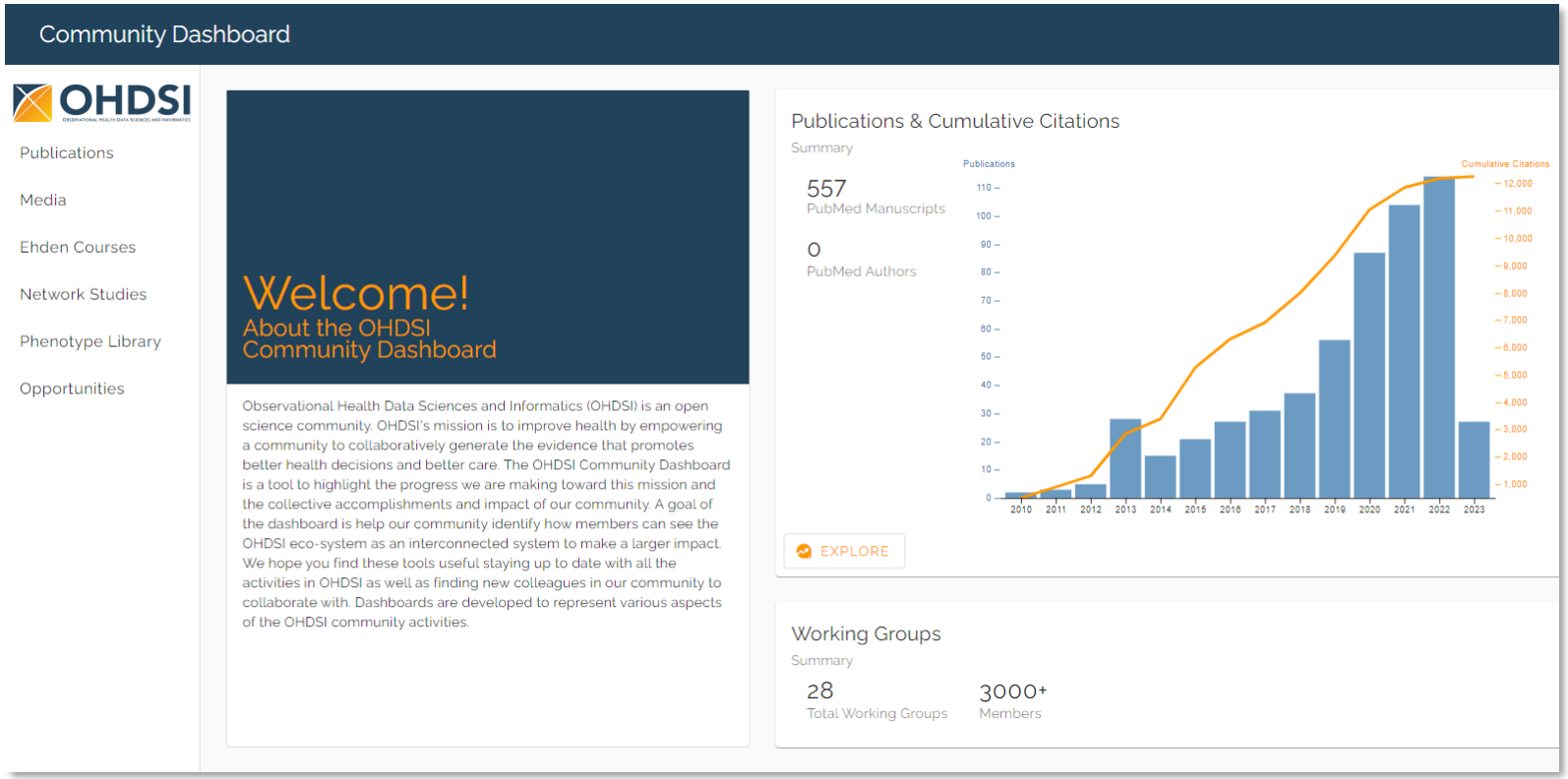
The most commonly used vocabularies were SNOMED, CPT4, HCPCS, LOINC, RxNorm, ATC, CVX, ICD family, and ICDO3. However, more efforts are needed to establish an external contribution pipeline for timely inclusion of source vocabularies and coding schemes into the Vocabularies.

Main errors in the Vocabularies content reported by the community include erroneous mappings, uphill mappings, unclear domain assignment and gaps in hierarchies.

The OHDSI community is interested in an ability to download a given version of the Vocabularies and project the impact of the changes across versions on research and ETL tasks. The common data and Vocabularies refresh in the community follows an annual or semi-annual cycle, but there is a high variation of the Vocabularies versions across the community, which may complicate network studies. The community also expressed a need for easily digestible documentation showcasing real-world Vocabularies' use cases, transparent roadmap, and impact of the releases and changes on ETL and research, and more details regarding specific vocabularies development and quality assurance.

New Community Dashboard

- New community dashboard now LIVE at https://dash.ohdsi.org/!
  - Includes tracking sites for publications, YouTube videos, EHDEN Academy courses, network studies, the phenotype explorer, and opportunities



DevCon 2023

- Recordings are now available at https://www.ohdsi.org/devcon2023/

Training Session #4

A*STAR Data Vault: a research platform for the Singapore ecosystem and beyond

By Mukkesh Kumar and Cindy Ho

Agency for Science, Technology and Research
SINGAPORE

A*STAR DATA VAULT: A RESEARCH PLATFORM FOR THE SINGAPORE ECOSYSTEM AND BEYOND

Mukkesh Kumar, Cindy Ho

Singapore

18th May 2023

CREATING GROWTH, ENHANCING LIVES

Mukkesh Kumar

Dr Mukkesh Kumar is the Head of Data Management Platform at A*STAR, leading the Clinical Data Management, Clinical Data Curation & Data Stewardship, Clinical Data Analytics & Reporting, CDMS/EDC Administration and Healthcare Software Development teams. Dr Mukkesh Kumar is a PhD alumnus of the NUS Saw Swee Hock School of Public Health, he has developed a predictive care framework for diabetes & maternal health, combining coalitional game theory concepts with machine learning. Working in close partnership with Singapore's Ministry of Health (MOH), Dr Mukkesh Kumar is developing the core OMOP data curation team at A*STAR to support MOH-TRUST Strategic Research Data Contributors.

Cindy Ho

Ms Cindy Ho is a Statistician at A*STAR, leading the data curation and data stewardship of population health studies. Ms Cindy Ho's abstract on 'GUSTO Data Vault: Working Towards OMOP Data Standardization' was showcased at OHDSI APAC Symposium 2022 in Taiwan. She has pioneered the OMOP data curation efforts for Growing Up in Singapore Towards healthy Outcomes (GUSTO) cohort study and is developing the OMOP Data Catalogue module to be integrated within GUSTO Data Vault platform for OMOP-based research. Ms Cindy Ho graduated with a major in Statistics from the NUS Department of Statistics, with domain expertise in cardiometabolic diseases.

Content

Introduction to GUSTO Data Vault	Page 4
OMOP Data Catalogue	Page 10
Singapore's National Level OMOP Data Curation Team	Page 15

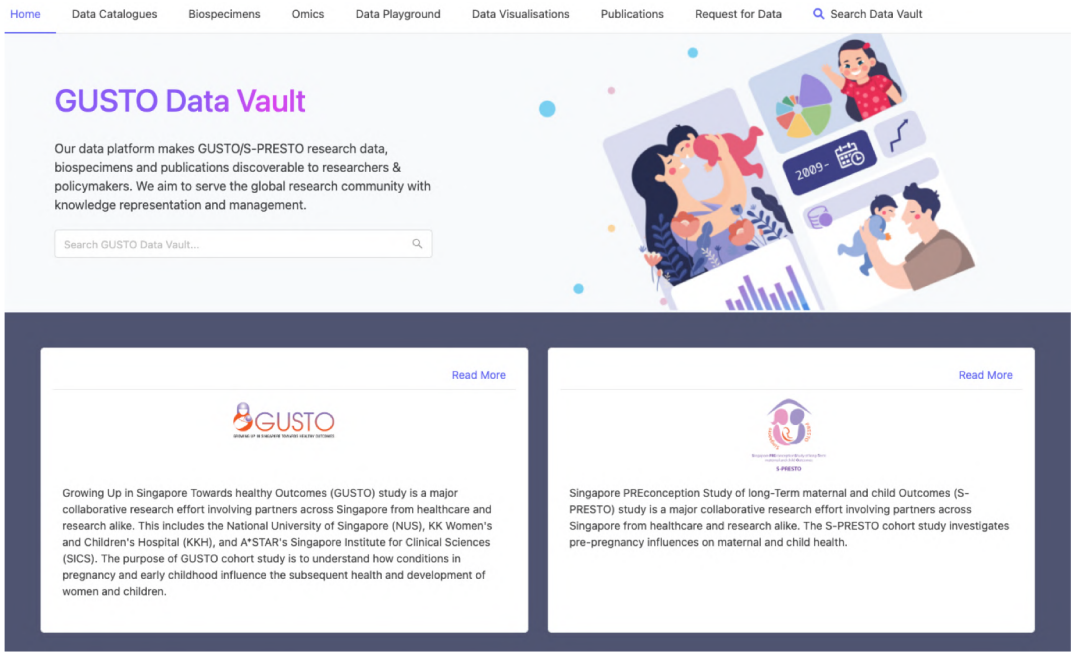
CREATING GROWTH, ENHANCING LIVES

Introduction to GUSTO Data Vault

A*STAR Data Vault

The GUSTO Data Vault platform have advanced data exploration capabilities for research data, biospecimens and publications asset management.

URL: https://gustodatavault.sg/



plotly

Singapore Institute for Clinical Sciences Enables More Open Sharing of Health Research with Interactive Data Vault Platform

Plotly shares success story of SICS' GUSTO Data Vault

CREATING GROWTH, ENHANCING LIVES

CREATING GROWTH, ENHANCING LIVES

plotly

Research Institute Visualizes Health Data Through Interactive Platform

Dash Enterprise enables Singapore Institute for Clinical Sciences (SICS) to deploy data catalogs, analytics, and published research.

After launching two cohort studies spanning over a decade's worth of maternal and child health data, SICS encountered roadblocks to getting that data into the hands of researchers: data complexity, data silos, and divergent skill sets.

This fueled their decision to move from using the Dash open-source framework to Dash Enterprise.

With Dash Enterprise, SICS was able to further streamline and accelerate their work to build and manage a data vault platform that would equip researchers and collaborators to develop their own hypotheses to gain insight into the long-term health impacts of conditions during pregnancy and early childhood.

Download this user story to learn how SICS used Python and Dash Enterprise to achieve the following:

- Collaboration and speed with control over user accesses, permissions, and app management
- 5,500 global annual visitors who spend an average of 20 minutes at a time on the platform, suggesting that the platform is promoting engagement with the cohort data
- Over 300 indexed publications and interest from leading research groups around the world

Plotly Success Story on GUSTO Data Vault

Plotly
@plotlygraphs

When data science teams and domain experts work together, scientific discovery accelerates. 🔬

Singapore Institute for Clinical Sciences (SICS) knows the power of collaboration.

4:04 am · 23 Nov 2022

plotly · Follow

plotly When data science teams and domain experts work together, scientific discovery accelerates. 🔬

Singapore Institute for Clinical Sciences (SICS) knows the power of collaboration. They needed to get data into the hands of researchers, but met data complexity, data silos, and divergent skill sets. SICS tapped on Dash Enterprise to build and manage a data vault for two cohort studies spanning over a decade's

5 likes

NOVEMBER 23, 2022

Log in to like or comment.

GUSTO Data Vault was Showcased at OHDSI APAC Symposium 2022, Taiwan

GUSTO Data Vault: Working Towards OMOP Data Standardization

PRESENTERS: Cindy Ho, Mukkesh Kumar

INTRO:

- Growing Up in Singapore Towards healthy Outcomes (GUSTO) aims to understand how conditions in pregnancy and early childhood influence the subsequent health and development of women and children.
- The GUSTO Data Vault platform have advanced data exploration capabilities for research data, biospecimens and publications asset management.

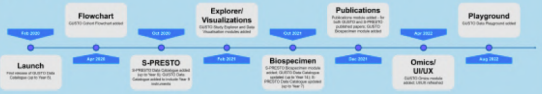
METHODS

- Data Vault (containerized web application with Docker) was built using PostgreSQL database and Django framework.
- Tools used: HTML, CSS, jQuery, Ajax, Python, Plotly Dash, Dashboard engine in Dash Enterprise.

RESULTS

- GUSTO Data Vault platform showcases data across Demographics, Women's Health, Children's Health, Metabolic, Neurodevelopment, Imaging and Omics domains.
- Researchers can browse and discover curated datasets and use the data visualizations to aid with hypothesis construction/discoveries.
- Applicants can submit their data access request to the GUSTO executive committee for approval, after exploring the Data Vault platform.
- Data Vault is refreshed with data updates about every 4 months.
- Data Vault sees around 5,500 unique visitors in a year, mostly from Singapore, with 20% based in countries across Asia, Australia, Europe, and North America.

Applying the **FAIR** (Findable, Accessible, Interoperable, Reusability) data principles, an *open interactive platform* (**GUSTO Data Vault**) was created using *open-source technologies* for population health studies.



Take a picture to visit GUSTO Data Vault

Take a picture to download the abstract

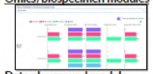
Our future work includes the implementation of OMOP Common Data Models (CDM) in Data Vault for expansion of real-world data access.

Snippets of Data Vault

Data catalogue/study explorer

Interactive data visualizations

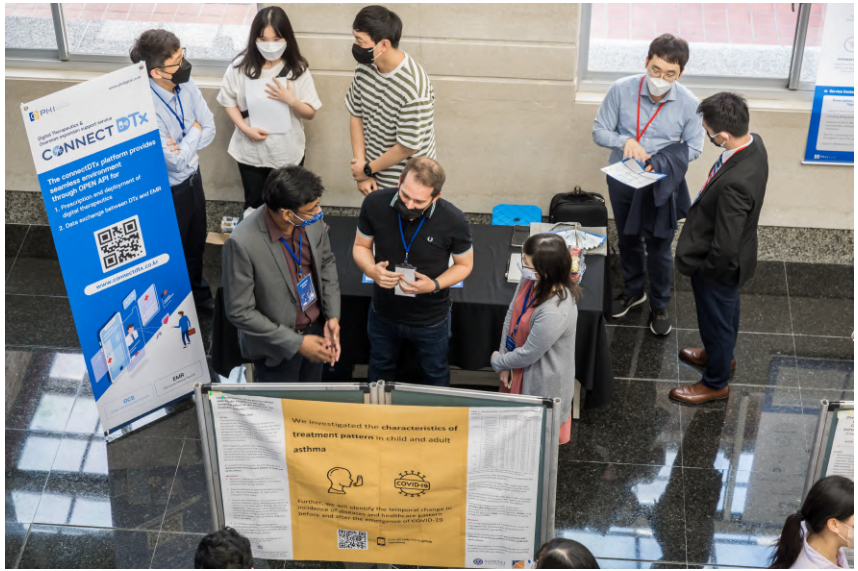
Omics/biospecimen modules



Data playground module

Publications

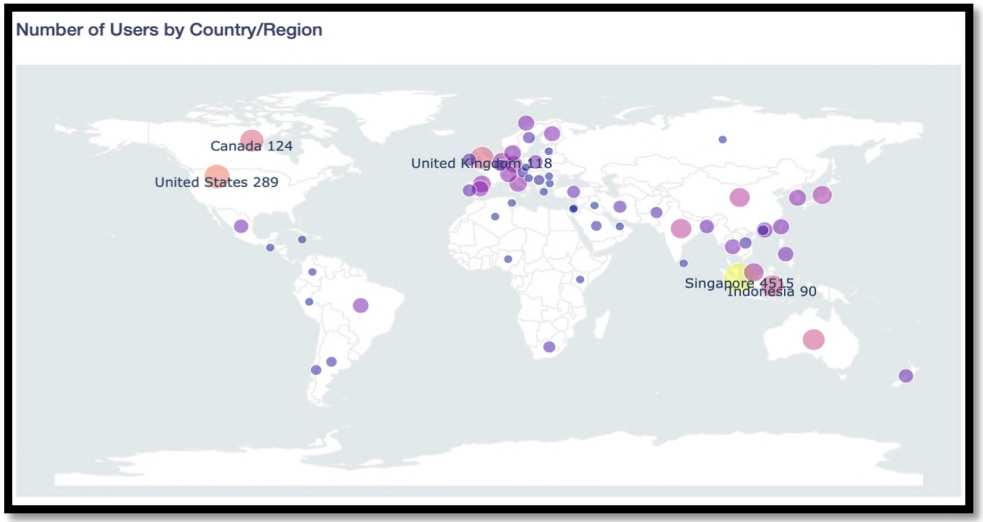
Cindy Ho, Li Ting Ang, Maisie Ng, Hang Png, Shuen Lin Tan, Estella Ye, Sunil Kumar Raja, Mengling Feng, Johan G Eriksson, Mukkesh Kumar



Global Impact of GUSTO Data Vault

Data Vault sees around 5,500 unique visitors in a year (56% of new users), mostly from Singapore, with 20% based in countries across Asia, Australia, Europe and North America.

GUSTO Data Vault is engaging research communities and furthering global research.

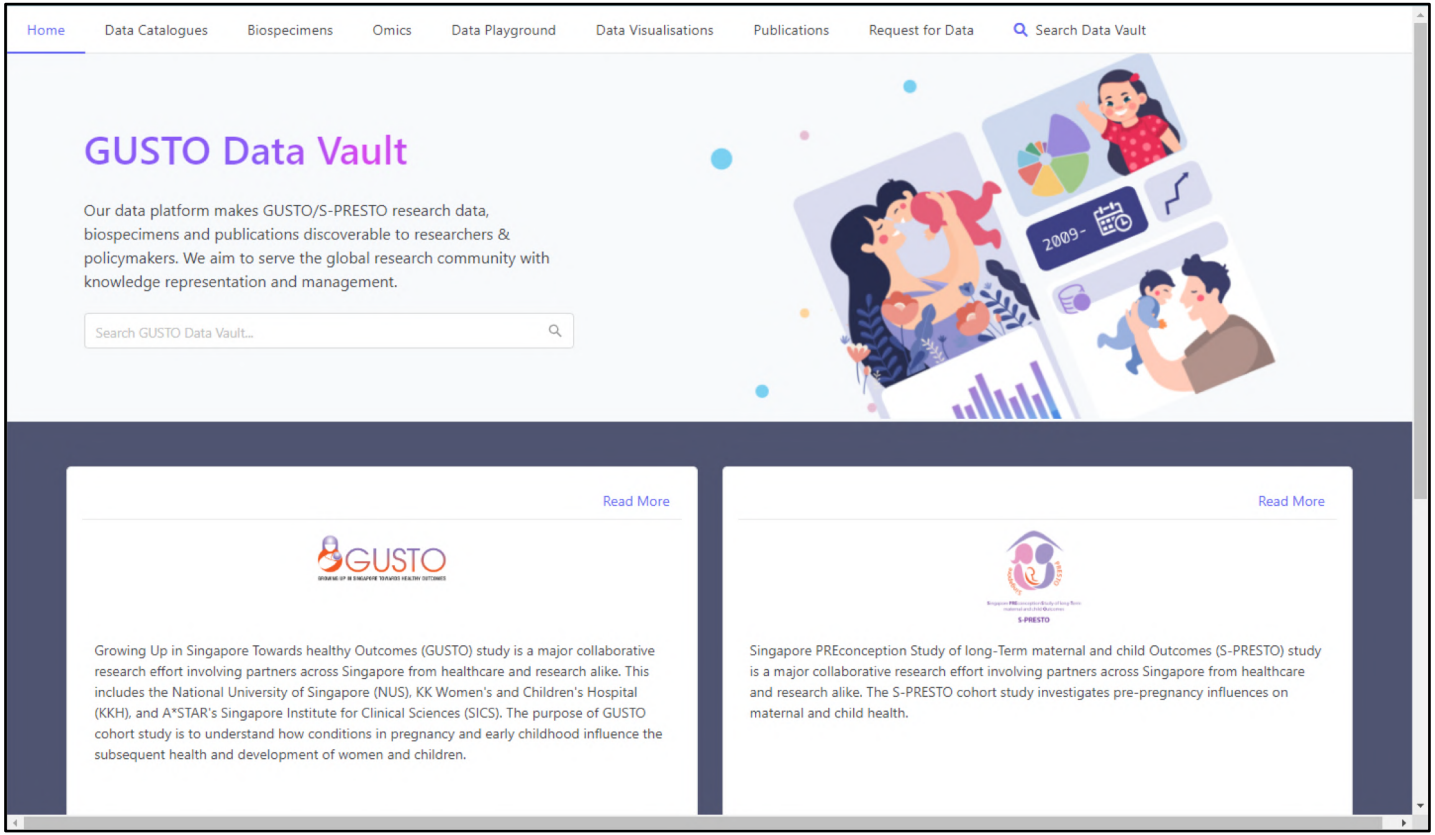


Number of Users by Country/Region - Top 10

COUNTRY/REGION	NUMBER OF USERS	%
Singapore	4515	80.38
United States	289	5.15
Canada	124	2.21
United Kingdom	118	2.1
Indonesia	90	1.6
Australia	73	1.3
India	44	0.78
China	36	0.64
Malaysia	29	0.52
Japan	26	0.46

CREATING GROWTH, ENHANCING LIVES

GUSTO Data Vault Demo



CREATING GROWTH, ENHANCING LIVES

OMOP Data Catalogue

GUSTO OMOP Data Conversion Journey

- WhiteRabbit, Rabbit-in-a-hat

- Athena

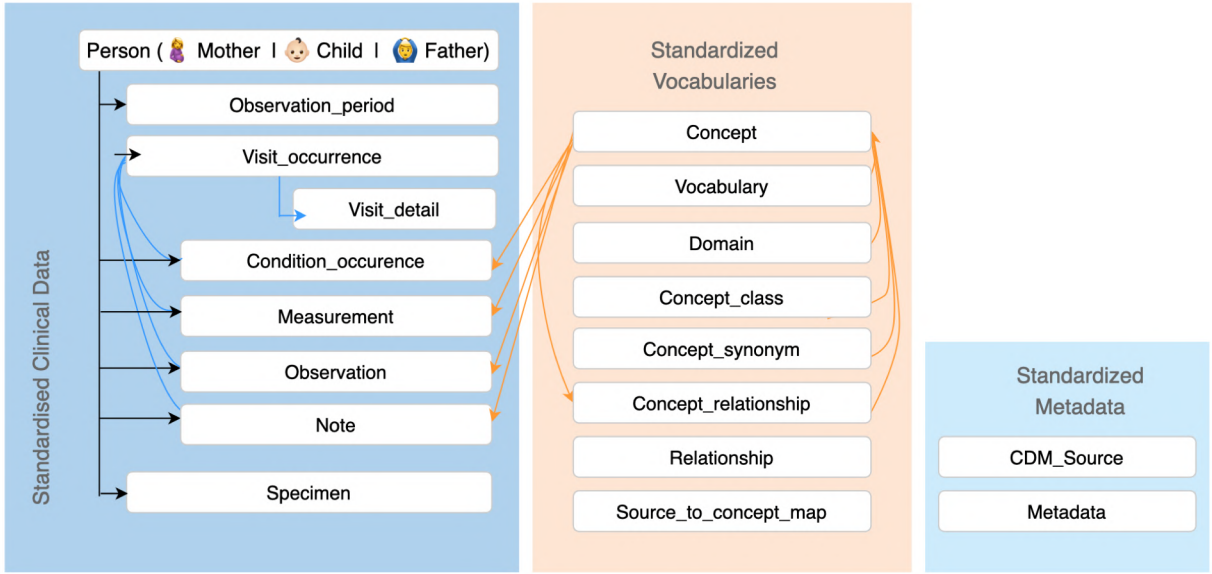
- Customised R scripts

- IQVIA OMOP Converter

CREATING GROWTH, ENHANCING LIVES

CDM Version & Fields/Tables

- CDM v5.4



GUSTO domains:

- *Demographics*
- *Metabolic*
- *Women's Health*
- *Child's Health*
- *Neurodevelopment/Mental Health*
- *Imaging*
- *Omics*

CREATING GROWTH, ENHANCING LIVES

Some Issues and Challenges Faced

- A lot of manual work
- Variables too specific/too broad
- Mapped to 2 concepts
  - *E.g. Has your child ever been diagnosed with bronchiolitis/bronchitis?*
- Specific categories for some variables
  - *E.g. How often have you done moderate exercise?*
    - Never, Once every 2-3 months, Once a month, Once a fortnight, 1-2 times per week, 3-6 times per week, Once a day, More than once a day

CREATING GROWTH, ENHANCING LIVES

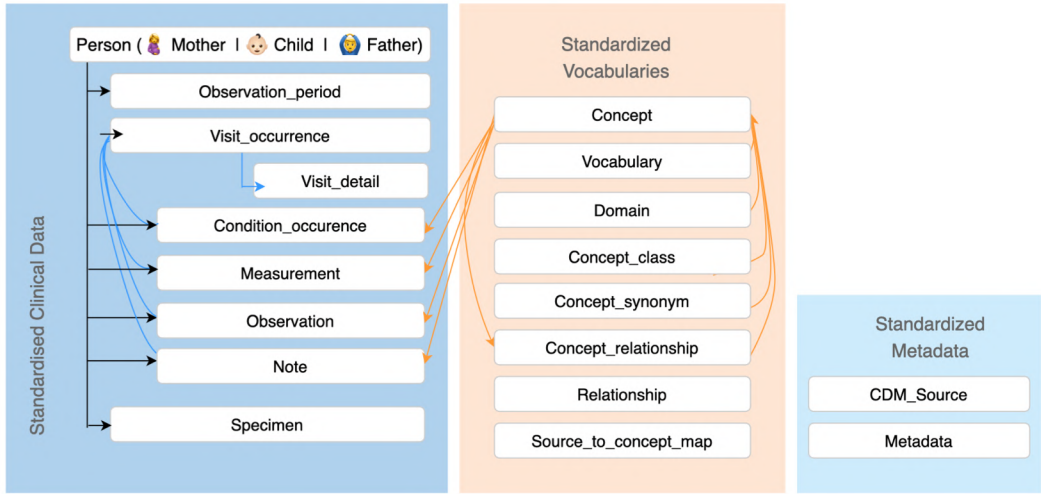
OMOP Data Catalogue Demo

Home | Data Catalogues | GUSTO OMOP Common Data Model | Biospecimens | Omics | Data Playground | Data Visualisations | Publications | Request for Data | Search Data Vault

GUSTO OMOP Common Data Model

- Home
- Clinical Data
  - Person
  - Condition
  - Observation
  - Measurement

The OMOP Data Catalogue showcases the GUSTO data assets which have been converted into Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) format. OMOP CDM is an open community data standard, designed to standardise the structure and content of observational data. The harmonisation of data using OMOP CDM enables systematic and collaborative OMOP-based research. The current CDM version is CDM v5.4, as depicted below for GUSTO tables and fields



CREATING GROWTH, ENHANCING LIVES

Singapore’s National Level OMOP Data Curation Team

National Level OMOP Mapping

Trusted Research and Real world-data Utilisation and Sharing Tech (“TRUST”) is Singapore’s national health-related data exchange platform that can facilitate safe and secure sharing and usage of anonymised health-related research and real-world data between public institutions, and between the public and private sectors to improve health outcomes.

TRUST has selected the OMOP-CDM model as the data standard for Clinical data.

A*STAR and MOH are collaborating with the aim of setting up a central data curation team to support TRUST Strategic Research Data Contributors (e.g. GUSTO).

TRUST

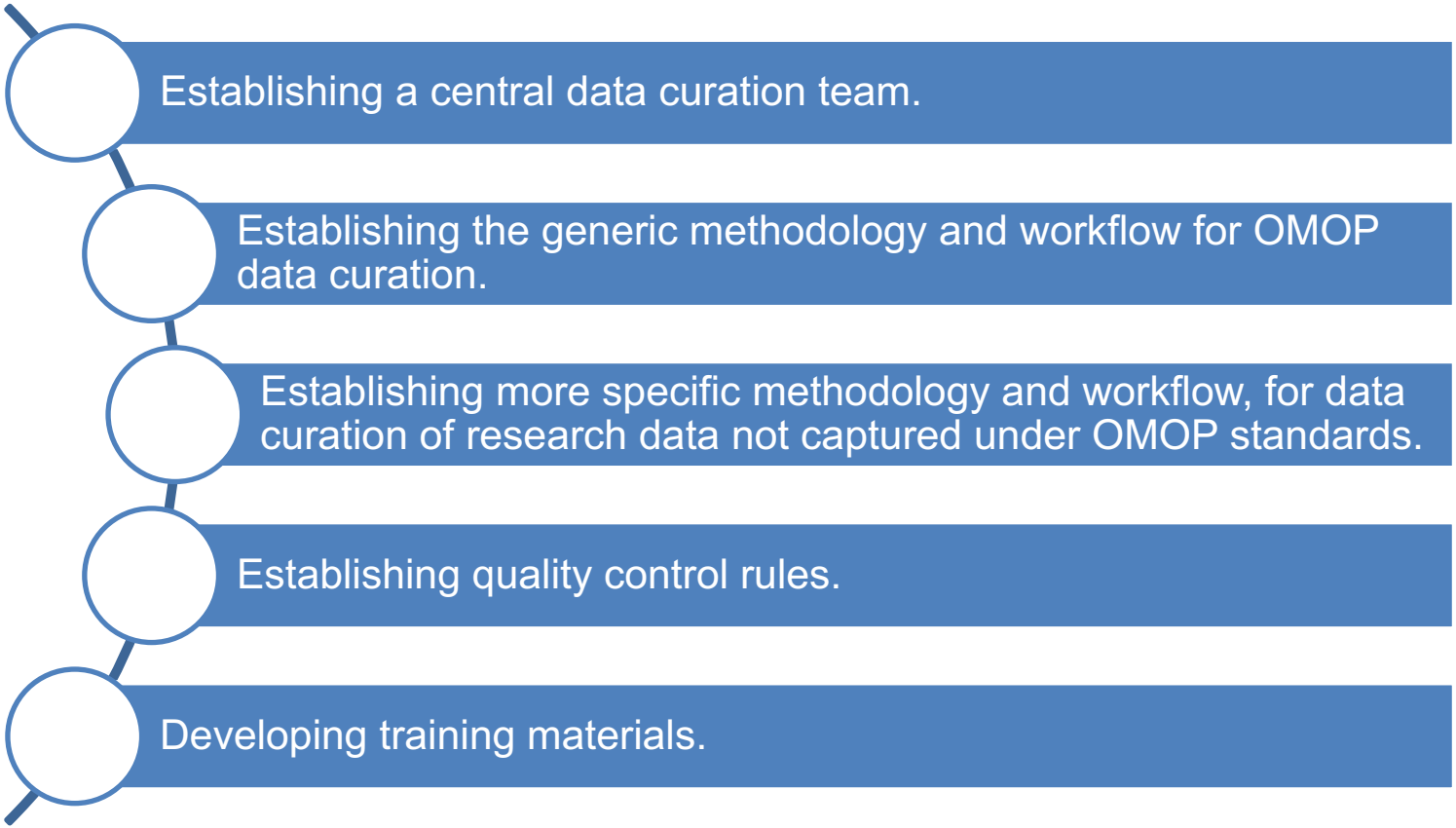
OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Agency for Science, Technology and Research
SINGAPORE

MINISTRY OF HEALTH
SINGAPORE

CREATING GROWTH, ENHANCING LIVES

National Level OMOP Data Curation Team - Project Milestones



CREATING GROWTH, ENHANCING LIVES

CREATING GROWTH, ENHANCING LIVES

Agency for Science, Technology and Research
SINGAPORE

THANK YOU

www.a-star.edu.sg