



A*STAR – Data4Life Session 1

GUSTO CDM and Data2Evidence Exploration

3 September 2026

Presented by:

A*STAR:

- Mukkesh P Kumar
- Cindy Ho
- Lynnette Tan

Data4Life:

- Tim Walz
- Peter Hoffmann



Contents

- **Introduction to GUSTO data**
- **Data Quality & Research Direction**
- **Challenges in CDM conversion**
- **Data2Evidence introduction + demo**

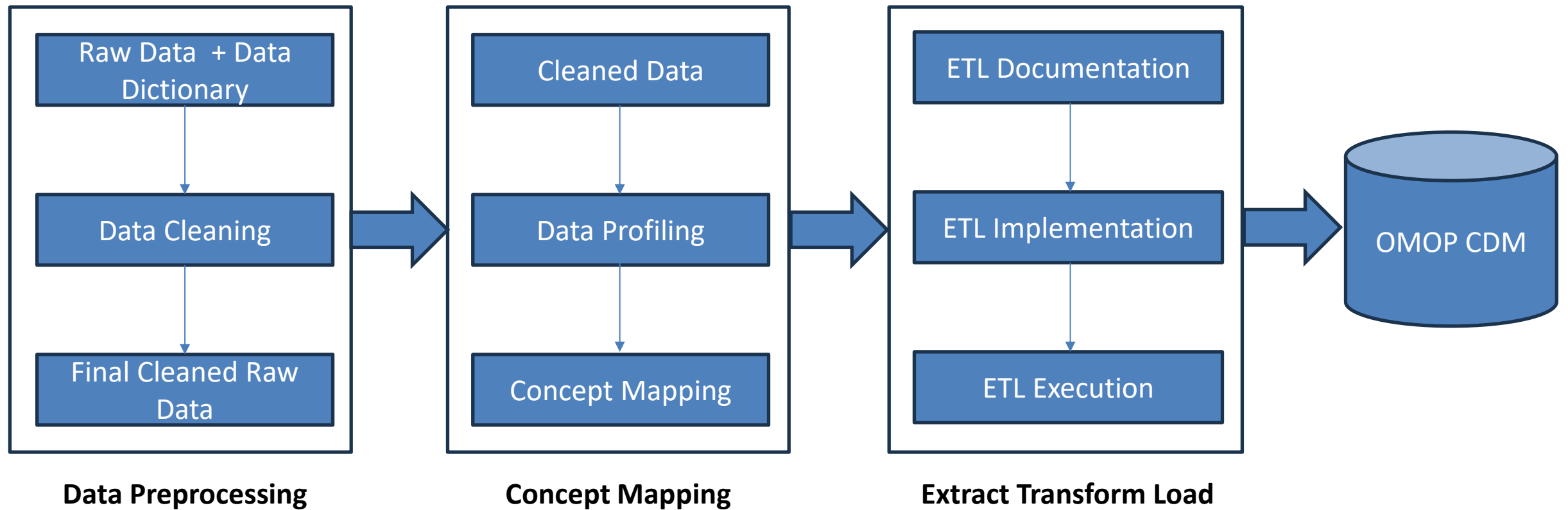


Brief Overview



The Journey

- Standard Journey: from raw data to OMOP CDM





Why OMOP CDM?

Short-term wins: To make use of **OHDSI tools** for **data analytics** on GUSTO dataset

Long-term goal: The end goal of data standardisation is **federated analytics** – enabling larger, more reliable studies

OHDSI tools



ATLAS

Cohort-level analytics



HADES

Large-scale analytics



Data Quality Dashboard

...and more!



- Strategic dataset – many partnerships
- Young mothers and children: Area of interest for Singapore, and potentially globally
- Collaborations are important



Introduction to GUSTO



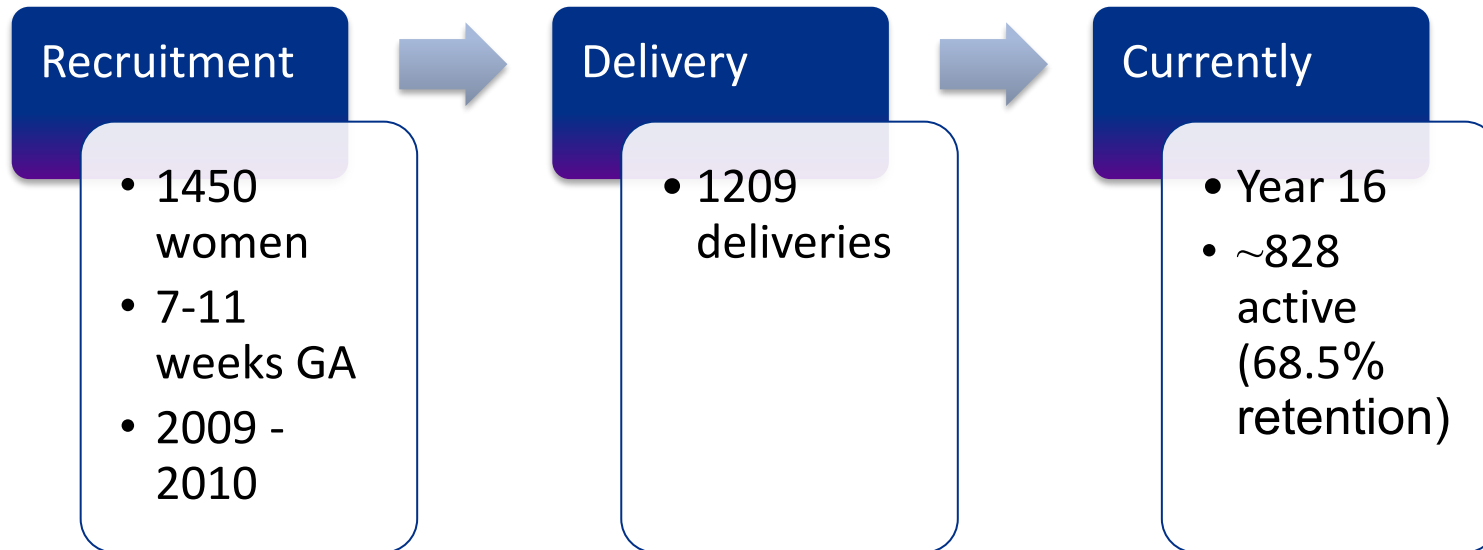
Introduction of GUSTO

- Growing Up in Singapore Towards healthy Outcomes (GUSTO)
- To understand how conditions in pregnancy and early childhood influence the subsequent health and development of women and children
- Empower the Singapore population to work towards a healthier next generation



Introduction of GUSTO

- Major collaborative birth cohort study – involves partners across healthcare and research in Singapore





GUSTO Data Collected

- Demographics / Social Relationships
- Women's Health
 - Cardiovascular: blood pressure
 - Medical history
- Children's Health
- Metabolic & Body Composition
 - Mother
 - Blood: OGTT (with derived GDM/T2D/IGT/IFG status), HbA1c, Lipids
 - Anthropometry
 - Child
 - Blood data: Fasting glucose, FBC
- Neurodevelopment
- Imaging
 - Fetal ultrasound, Brain MRI, Body MRI
- Omics



<https://gustodatavault.sg/>

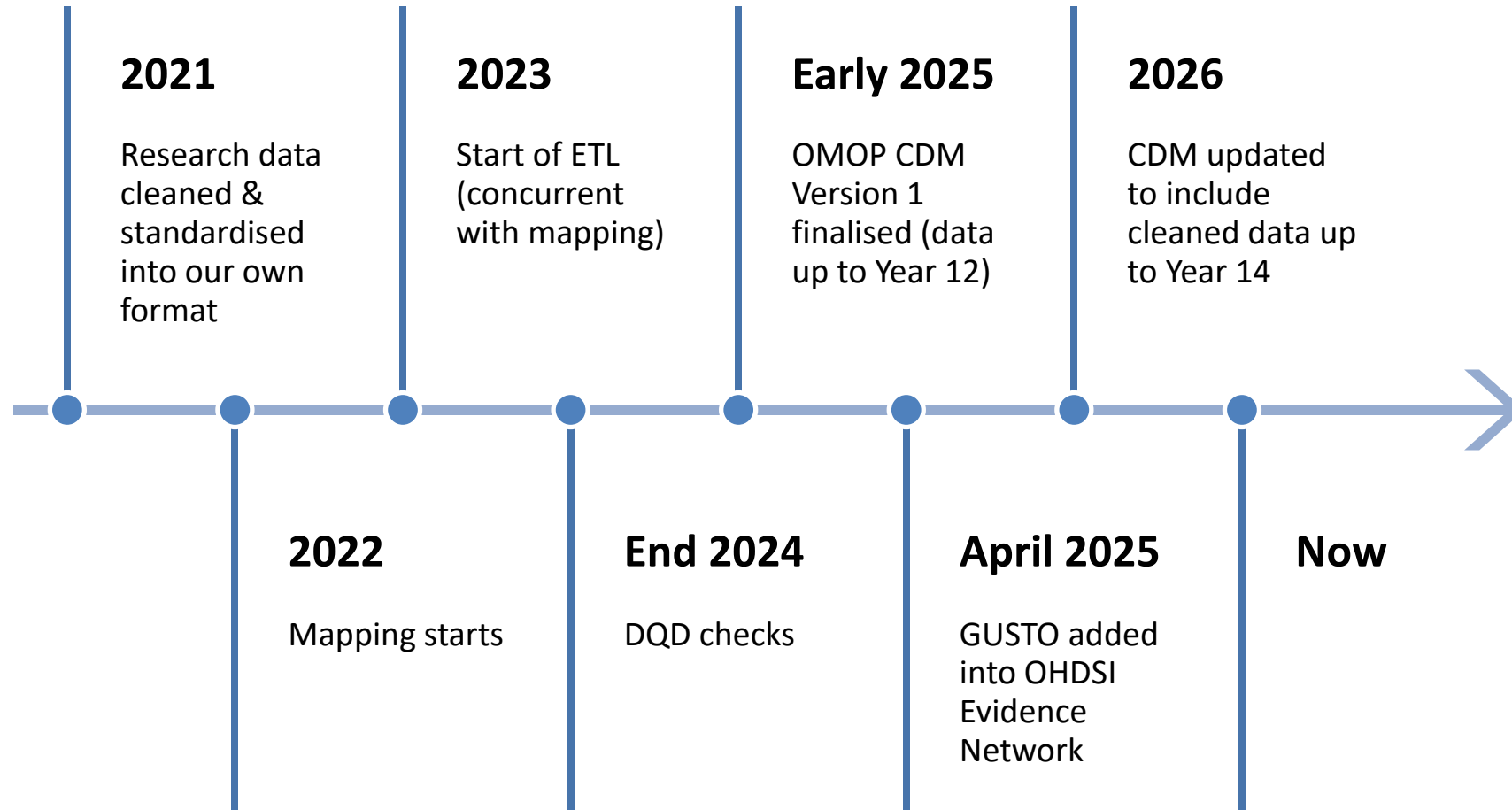


Why GUSTO data is valuable?

- Inter-generational
 - How mothers can affect the child's metabolic outcomes
- Multiracial
 - 56.0% Chinese, 25.4% Malay, 18.5% Indian (at delivery)
 - 57.4% Chinese, 26.0% Malay, 16.5% Indian (now)



Our OMOP Journey





Data Quality & Research Direction



Data Quality Interpretation

Dataset info

Data characterization

Data quality

GUSTO_V1_CACHE

Overview

	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	493	20	513	96%	291	0	291	100%	784	20	804	98%
Conformance	895	31	926	97%	137	8	145	94%	1,032	39	1,071	96%
Completeness	448	0	448	100%	17	0	17	100%	465	0	465	100%
Total	1,836	51	1,887	97%	445	8	453	98%	2,281	59	2,340	97%
1,469 out of 2,281 passed checks are not applicable, due to empty tables or fields. 42 out of 59 failed checks are SQL errors. Corrected pass percentage for NA and Errors: 98% (812/829).												

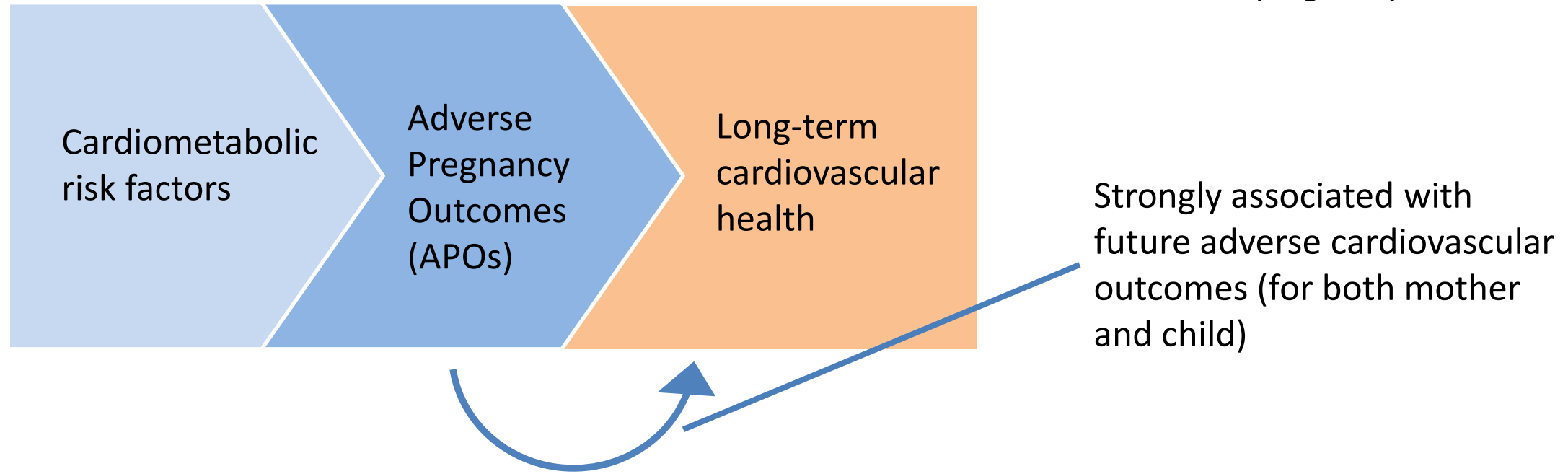


Research direction: Maternal cardiometabolic health

“...maternal cardiometabolic risk factors and APOs* also have long-term cardiovascular health implications...”

Source: Adverse pregnancy outcomes and long-term risk of chronic diseases: Evidence from the findings of the Tehran lipid and glucose study during a quarter of a century ([2026](#))

* *APOs - adverse pregnancy outcomes*





Current GUSTO Findings

Can you give me a short summary, in point form, of the maternal glycemia/GDM and also towards T2D progression findings for GUSTO mothers?

- **GDM prevalence varied by diagnostic criteria:** among ethnic Chinese GUSTO mothers, prevalence was **20.8% using WHO-1999 criteria** versus **12.0% using modified two-point IADPSG criteria**; the lack of 1-hour glucose data may have underestimated IADPSG-defined GDM (Loo et al., 2021).
- **Higher GDM risk** was associated with older maternal age, higher pre-pregnancy BMI, overweight/obesity, and previous GDM (Loo et al., 2021; Kumar et al., 2022).
- **Higher physical activity** was associated with lower 2-hour glucose and lower odds of GDM, particularly among overweight or obese mothers; sedentary behaviour showed no association (Padmapriya et al., 2017).
- **GDM strongly predicted later metabolic disease:** at 4–6 years after delivery, mothers with GDM had about **12 times the risk of T2D** and **three times the risk of broader dysglycaemia** compared with mothers without GDM (Chen et al., 2021).
- **Weight status further influenced progression:** pre-pregnancy overweight/obesity and retaining ≥ 5 kg after delivery increased dysglycaemia risk, whereas gestational weight gain was not consistently associated with later dysglycaemia (Chen et al., 2021).
- **For predicting T2D 4–8 years later,** mid-pregnancy BMI plus GDM status achieved an AUC of **0.86**, and 2-hour OGTT glucose predicted better than fasting glucose; however, this model was trained on only 561 pregnancies and requires external validation (Kumar et al., 2022).

Answered using: Publications  View Records  Export to PDF

For this demo: We focus on the relationship between GDM and T2D

<https://gustodatavault.sg/askai/>

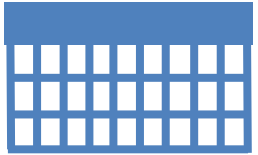


Challenges faced

During OMOP CDM conversion



Challenges: OMOP Mapping + Integration



GUSTO has a lot of survey & questionnaire data



Tedious process to choose the best concept match



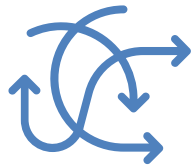
Domain expertise needed for some data



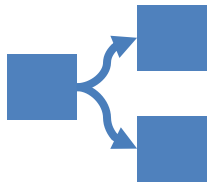
OHDSI tools are largely standalone – not integrated



Challenges: ETL



GUSTO data is standardised in our own format,
custom scripts required



Some data crosses multiple OMOP tables



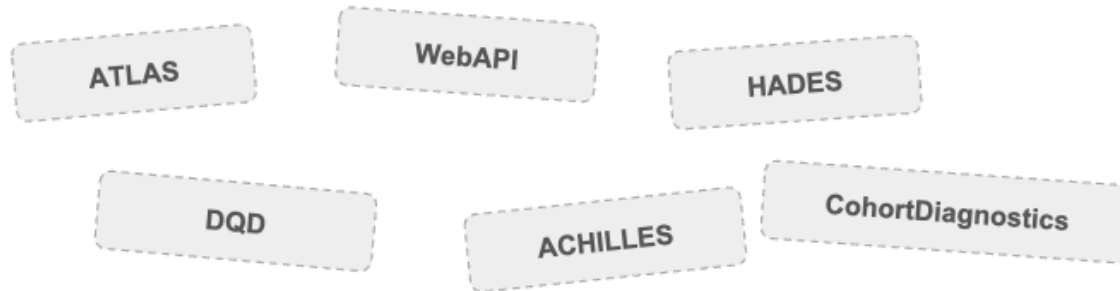
Data2Evidence

<https://github.com/OHDSI/Data2Evidence/>



Using Data2Evidence to set up the system

Today: assembled by hand



- Install & configure each component
- Versions drift across environments
- Manual handoffs between tools

With Data2Evidence: one workspace



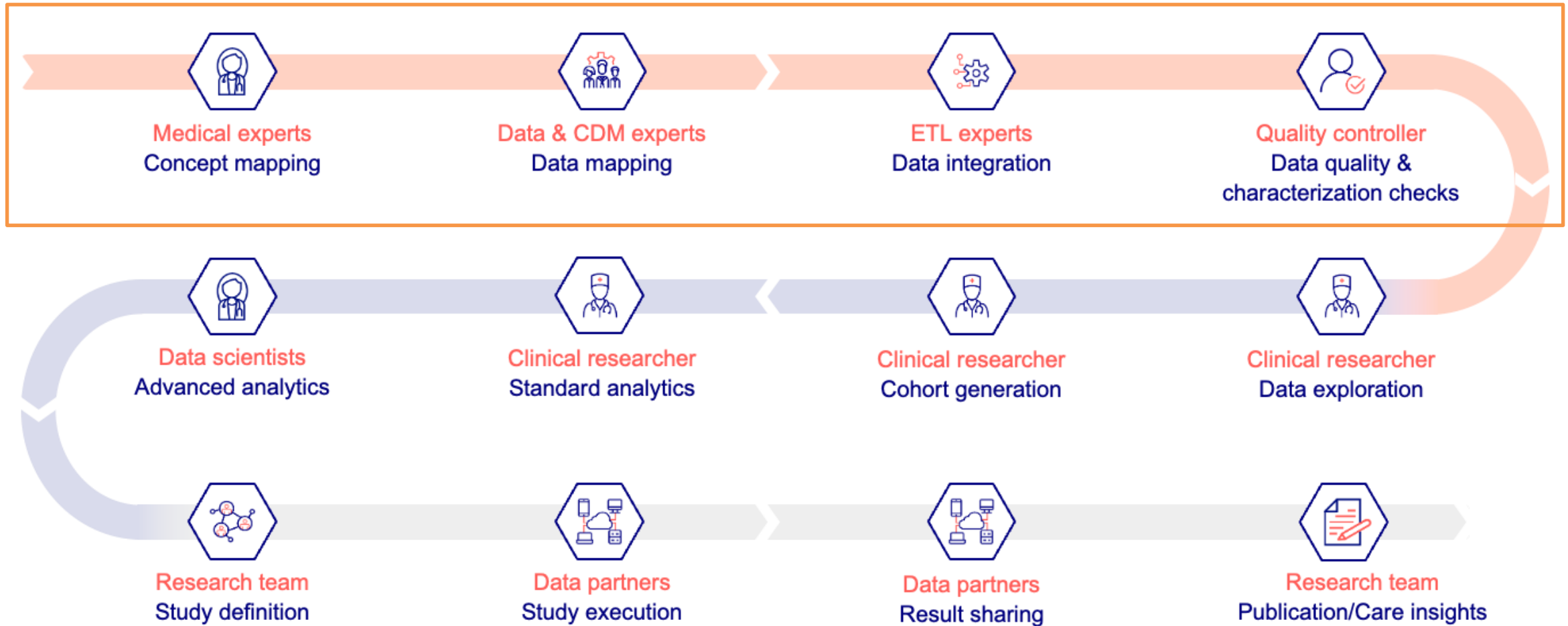
- ✓ One login, one workspace
- ✓ Version-pinned & reproducible
- ✓ Same cohort logic in ATLAS and notebooks

From data to evidence - openly.

Open-source. OMOP-native. Designed to run client-side.



Using Data2Evidence to answer our questions





Data2Evidence: Connect to GUSTO OMOP CDM

Start from GUSTO OMOP CDM import

- Add database
- Add dataset
- Run DQD
- Run data characterization

Setup

Databases

Database connection and credentials

Configure

Users

Datasets

Studies

Jobs

ETL

Setup

Logs

Account

Update dataset metadata

Add dataset

Cache datasets

aa48f590-46...



GUSTO_V1_CACHE

gt_v1_cdm



v5.4

v5.4

omop5-4 [omop_cdm_plugin]

omop

Select action

Select action

Update dataset

Permissions

Resources

Delete dataset

Run data quality

Run data characterization

Update cache

Setup semantic search

Manage dashboard



Data2Evidence DEMO



Challenges addressed



Concept search mapping



Ability to perform ETL in graphical user interface (GUI), easier to understand ETL code



Integration of OHDSI tools

- e.g. "Rabbits", DQD, Achilles, Pheobe, ATLAS, ...



Thank you!