THE UNIVERSITY OF MELBOURNE

Centre for
Digital Transformation of Health

Unlocking the Value of Health Data with Large Scale Analytics

Assoc Prof Graeme Hart
Austin Health
Clinical Informatics Director (Austin)
Centre for Digital Transformation of Health, University of Melbourne

Austin HEALTH

Acknowledgements: Assoc Prof Doug Boyle, Dr Christine Hallinan, Dr Roger Ward
Mr Andrew Howard.

Austin Health and the University of Melbourne acknowledge the Traditional Owners of the land on which we meet today, the Wurundjeri people and all members of the Kulin nation.

We would like to pay our respects to Elders past and present and extend that respect to other Aboriginal and Torres Strait Islander people who are here today.

Data is Generated Everywhere – in the petabytes

- Health is no exception
  - How best to harness this information and transform into actionable knowledge?
  - Domains of Practice and Influence/Control
  - Administrative – local, state and federal
  - Clinical – Audit, Financial,
  - Clinical Quality Registries – Professional Societies, ACSQHC, Jurisdictions
  - Research – clinician led (+/- Grant funded), Academic, industry led
    - Single site, multi site, multi jurisdictional, linked, unlinked

What is “VALUE”

- Value derives from the formative use of information to influence and innovate behaviours and therapies and evaluate the consequences of our actions and change.

- The efficient creation, dissemination and use of information provides best value to the community

SHORT COURSE Learning Healthcare Systems Academy

Centre for Digital Transformation of Health

MELBOURNE

Applied Learning Healthcare Systems

Training the future champions of digital transformation of the health system.

Do you want to learn to use health data to inform clinical practice and to design and validate digital health-enabled models of care?

The Centre for Digital Transformation of Health (DT4Health) is offering selected enrolment in an online pilot of a short course on creating a digital health-enabled Learning Healthcare System.

This course is to support you, your organisation and ultimately the health sector in developing and implementing digital health innovations to improve clinical decision making and healthcare quality.

What will I learn?

You will gain the fundamental knowledge and skills in clinical informatics and digital health that you need to apply the Learning Healthcare System approach to practice improvement:

- gain hands-on experience working through the phases of a Learning Healthcare System
- step through a scenario of improving diabetes management

Data to Knowledge
- use real healthcare data to gain new knowledge to drive practice change

Knowledge to Practice
- design, build and validate components of a remote monitoring system

Practice to Data
- create an implementation and evaluation plan
- measure whether digital health innovations actually work

Information session

5:30-6:30pm
Mon 23 August 2021
Hear from program leaders, followed by a Q & A.

REGISTER

What are the prerequisites?

- strong interest in the topic and commitment to complete the course and engage in real-time sessions
- no programming experience or healthcare expertise required
- agreement to take part in evaluation of this pilot, so that we can use the learnings to tailor future programs to meet health sector needs

What is “VALUE”: Perspective Statements for data custodians, researchers, & funders

Data custodians, researchers, and funders have various expectations from the data market place, which is driven by their varying ideas for how to apply data for analytical research, nature of research undertaken, and transformative changes it brings to health practice.

Data Custodians

Data derived from Point of Care, extracted and managed in secure, modern “warehouse” for internal management and clinical use.
Standards based, complete, ready for reporting, audit, research

Researchers

High Quality, linked data – easy to find, appropriate governance and access
Security and Privacy
Least cost access

Consumers

Information used to provide better care for me and for the community.
Information enabling “Patient Choice”
Security and Privacy

Funders

High quality patient focused data for system management, audit and research : faster, more efficient resulting in new knowledge, practice change, improved outcomes and increased efficiency

Complication: Problem statements for data custodians, researchers, & funders

Data custodians, researchers, and funders have various expectations from the data market place, which is driven by their varying ideas for how to apply data for analytical research, nature of research undertaken, and transformative changes it brings to health practice.

Data Custodians

How do we ensure that researchers respect the confidentiality and privacy of the patients and that the research is conducted ethically and derives translational outcomes?

Researchers

Is ease of access and ability to use the consolidated data on their analytic platform the top priority for us? Are we keen to see the outcomes of our research being implemented to change health practices?

Consumers

Is my biggest concern the level of trust in third party access to, and appropriate management of, my personal health data, or the knowledge that the investment will improve my health outcomes?

Funders

Is knowing that higher quality research is being done faster resulting in practice change and improved outcomes our top priority?

Where Does the Data Live?

Research Data Australia

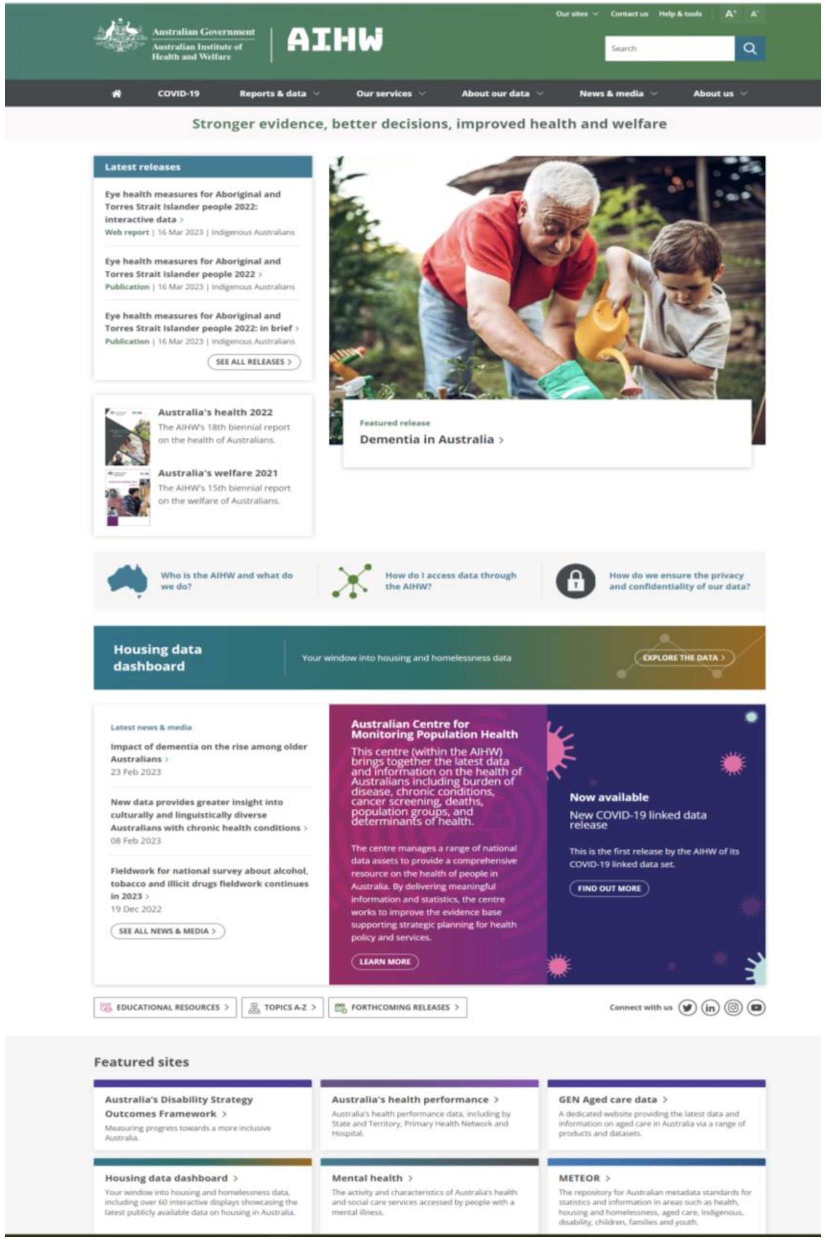
Research Data Australia

More than ever we're recognising the value of data and the huge amount of untapped potential in our nation's extensive research data collections.

That's why services such as Research Data Australia (RDA) are essential. The RDA is an online portal for finding research data and associated projects, researchers, and data services. You can find, access, and reuse data for research from over one hundred Australian research organisations, government agencies, and cultural institutions. RDA covers a broad spectrum of research fields from science, technology,

Australian Research Data Commons

AIHW



Stronger evidence, better decisions, improved health and welfare

Latest releases

Eye health measures for Aboriginal and Torres Strait Islander people 2022: interactive data >
Web report | 16 Mar 2023 | Indigenous Australians

Eye health measures for Aboriginal and Torres Strait Islander people 2022 >
Publication | 16 Mar 2023 | Indigenous Australians

Eye health measures for Aboriginal and Torres Strait Islander people 2022: in brief >
Publication | 16 Mar 2023 | Indigenous Australians

Australia's health 2022
The AIHW's 18th biennial report on the health of Australians.

Australia's welfare 2021
The AIHW's 15th biennial report on the welfare of Australians.

Featured release
Dementia in Australia >

Who is the AIHW and what do we do?

How do I access data through the AIHW?

How do we ensure the privacy and confidentiality of our data?

Housing data dashboard
Your window into housing and homelessness data

Latest news & media

Impact of dementia on the rise among older Australians >
23 Feb 2023

New data provides greater insight into culturally and linguistically diverse Australians with chronic health conditions >
08 Feb 2023

Fieldwork for national survey about alcohol, tobacco and illicit drugs fieldwork continues in 2023 >
19 Dec 2022

Australian Centre for Monitoring Population Health

This centre (within the AIHW) brings together the latest data and information on the health of Australians including burden of disease, chronic conditions, cancer screening, deaths, population groups, and determinants of health.

The centre manages a range of national data assets to provide a comprehensive resource on the health of people in Australia. By delivering meaningful information and statistics, the centre works to improve the evidence base supporting strategic planning for health policy and services.

Now available
New COVID-19 linked data release

This is the first release by the AIHW of its COVID-19 linked data set.

Featured sites

Australia's Disability Strategy Outcomes Framework >
Measuring progress towards a more inclusive Australia.

Australia's health performance >
Australia's health performance data, including by State and Territory, Primary Health Network and Hospital.

GEN Aged care data >
A dedicated website providing the latest data and information on aged care in Australia via a range of products and datasets.

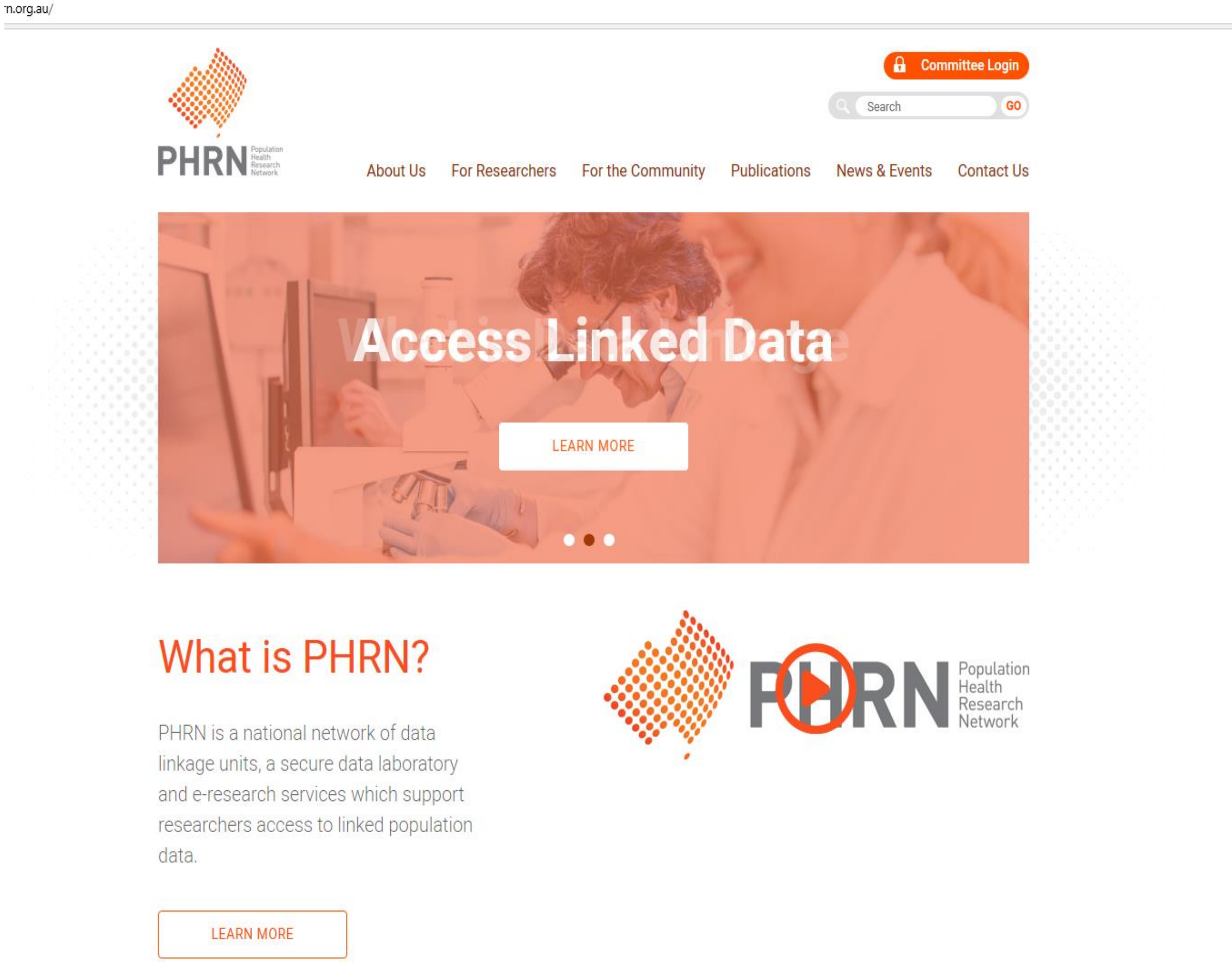
Housing data dashboard >
Your window into housing and homelessness data, including over 60 interactive displays showcasing the latest publicly available data on housing in Australia.

Mental health >
The activity and characteristics of Australia's health and social care services accessed by people with a mental illness.

METEOR >
The repository for Australian metadata standards for statistics and information in areas such as health, housing and homelessness, aged care, Indigenous, disability, children, families and youth.

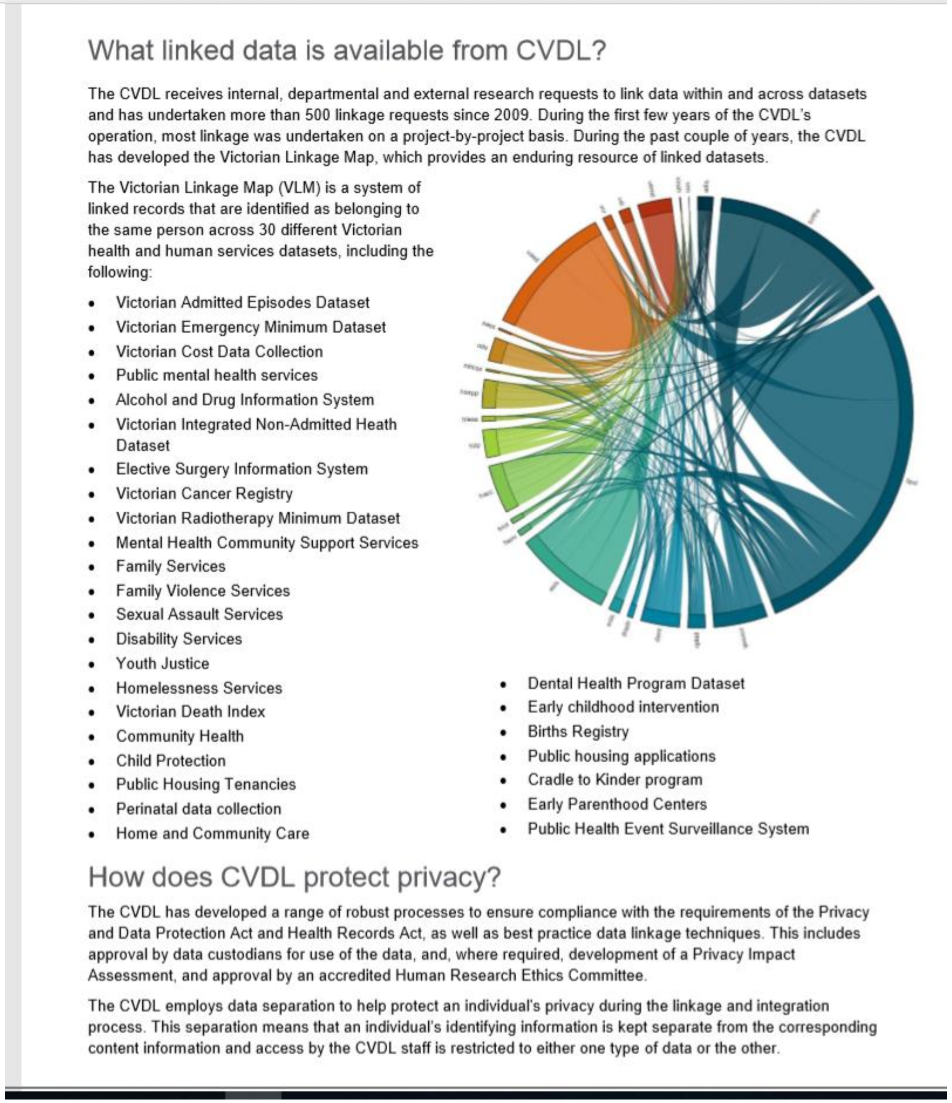
Public Health Research Network

Centre for Victorian Data Linkage



What is PHRN?

PHRN is a national network of data linkage units, a secure data laboratory and e-research services which support researchers access to linked population data.



What linked data is available from CVDL?

The CVDL receives internal, departmental and external research requests to link data within and across datasets and has undertaken more than 500 linkage requests since 2009. During the first few years of the CVDL's operation, most linkage was undertaken on a project-by-project basis. During the past couple of years, the CVDL has developed the Victorian Linkage Map, which provides an enduring resource of linked datasets.

The Victorian Linkage Map (VLM) is a system of linked records that are identified as belonging to the same person across 30 different Victorian health and human services datasets, including the following:

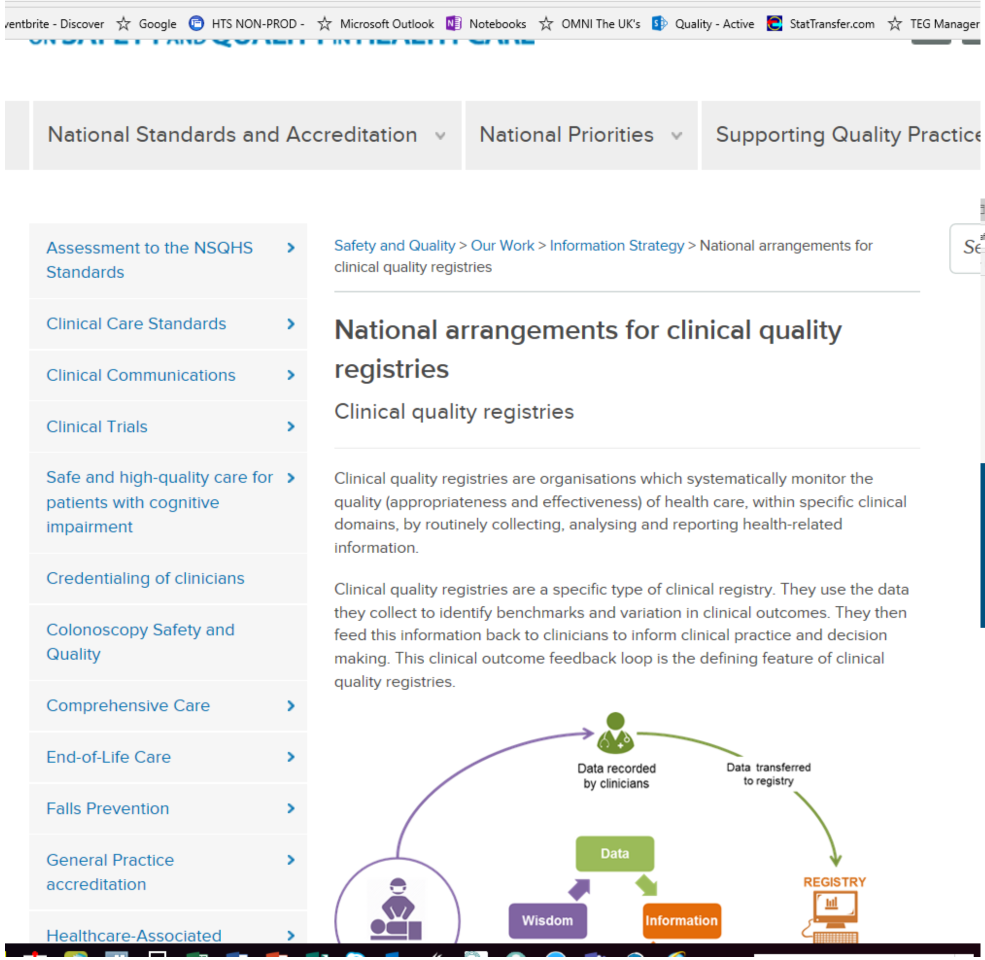
- Victorian Admitted Episodes Dataset
- Victorian Emergency Minimum Dataset
- Victorian Cost Data Collection
- Public mental health services
- Alcohol and Drug Information System
- Victorian Integrated Non-Admitted Heath Dataset
- Elective Surgery Information System
- Victorian Cancer Registry
- Victorian Radiotherapy Minimum Dataset
- Mental Health Community Support Services
- Family Services
- Family Violence Services
- Sexual Assault Services
- Disability Services
- Youth Justice
- Homelessness Services
- Victorian Death Index
- Community Health
- Child Protection
- Public Housing Tenancies
- Perinatal data collection
- Home and Community Care
- Dental Health Program Dataset
- Early childhood intervention
- Births Registry
- Public housing applications
- Cradle to Kinder program
- Early Parenthood Centers
- Public Health Event Surveillance System

How does CVDL protect privacy?

The CVDL has developed a range of robust processes to ensure compliance with the requirements of the Privacy and Data Protection Act and Health Records Act, as well as best practice data linkage techniques. This includes approval by data custodians for use of the data, and, where required, development of a Privacy Impact Assessment, and approval by an accredited Human Research Ethics Committee.

The CVDL employs data separation to help protect an individual's privacy during the linkage and integration process. This separation means that an individual's identifying information is kept separate from the corresponding content information and access by the CVDL staff is restricted to either one type of data or the other.

CQR resources



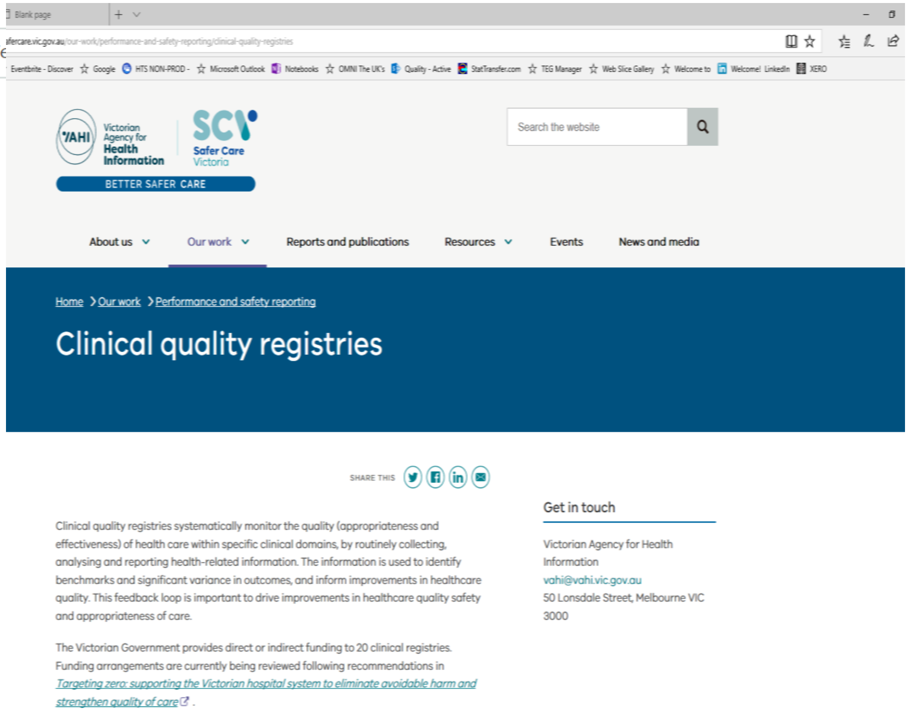
Safety and Quality > Our Work > Information Strategy > National arrangements for clinical quality registries

National arrangements for clinical quality registries

Clinical quality registries

Clinical quality registries are organisations which systematically monitor the quality (appropriateness and effectiveness) of health care, within specific clinical domains, by routinely collecting, analysing and reporting health-related information.

Clinical quality registries are a specific type of clinical registry. They use the data they collect to identify benchmarks and variation in clinical outcomes. They then feed this information back to clinicians to inform clinical practice and decision making. This clinical outcome feedback loop is the defining feature of clinical quality registries.



Clinical quality registries

Clinical quality registries systematically monitor the quality (appropriateness and effectiveness) of health care within specific clinical domains, by routinely collecting, analysing and reporting health-related information. The information is used to identify benchmarks and significant variance in outcomes, and inform improvements in healthcare quality. This feedback loop is important to drive improvements in healthcare quality safety and appropriateness of care.

The Victorian Government provides direct or indirect funding to 20 clinical registries. Funding arrangements are currently being reviewed following recommendations in *Targeting zero: supporting the Victorian hospital system to eliminate avoidable harm and strengthen quality of care* .

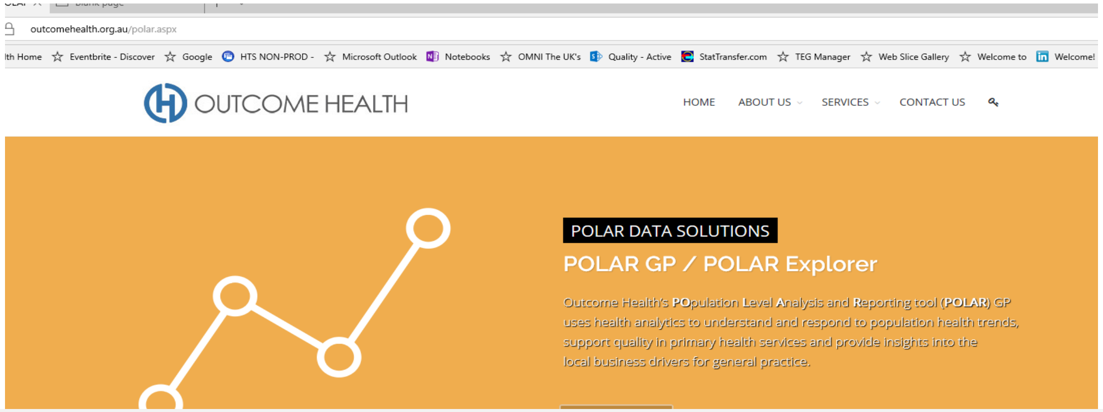
Get in touch

Victorian Agency for Health Information
vahi@vahi.vic.gov.au
50 Lonsdale Street, Melbourne VIC 3000

Monash Dept Epidemiology and Preventive Medicine

Cancer & Blood Diseases, Cardiovascular Disease, ,Critical Care, Trauma & Perioperative Medicine, Development, Stem Cells & Regenerative Medicine, Infection, Inflammation & Immunity, Metabolism, Obesity & Men's Health, Neurosciences & Mental Health, Public Health & Health Systems Improvement, Women's, Children's & Reproductive Health

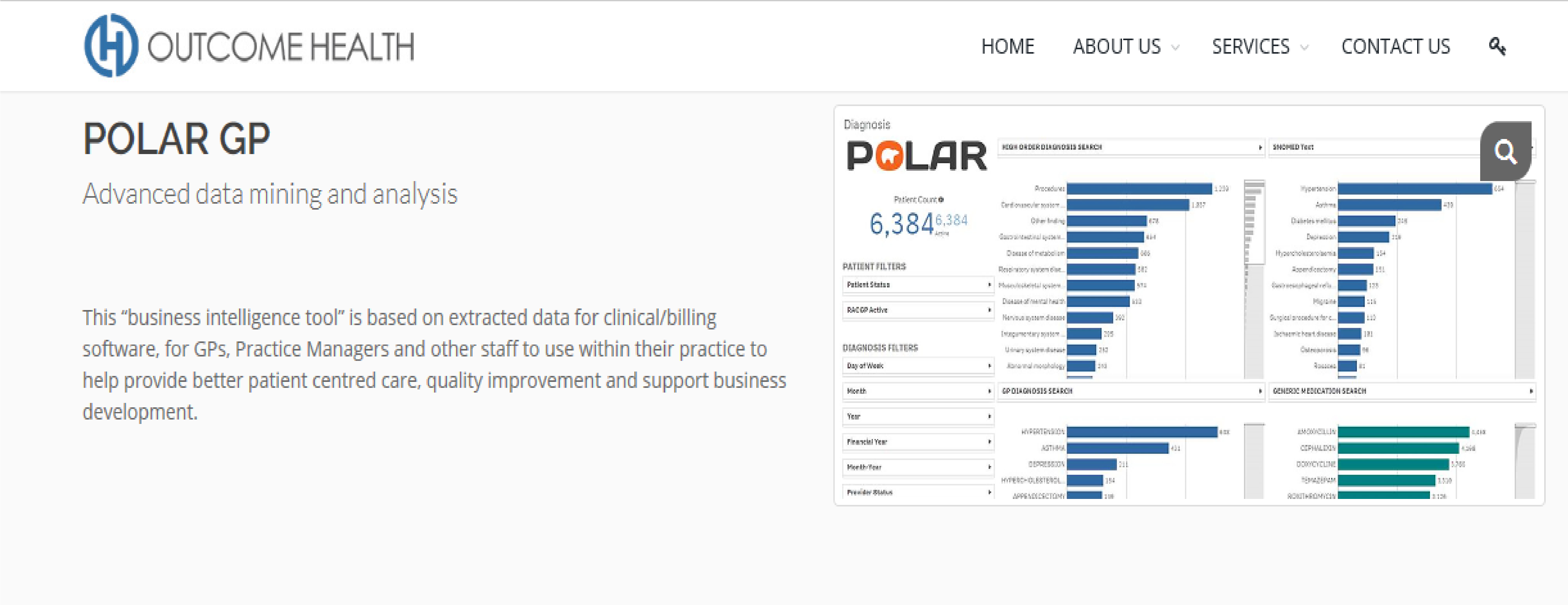
POLAR



POLAR DATA SOLUTIONS

POLAR GP / POLAR Explorer

Outcome Health's **PO**pulation **L**evel **A**nalysis and **R**eporting tool (**POLAR**) GP uses health analytics to understand and respond to population health trends, support quality in primary health services and provide insights into the local business drivers for general practice.



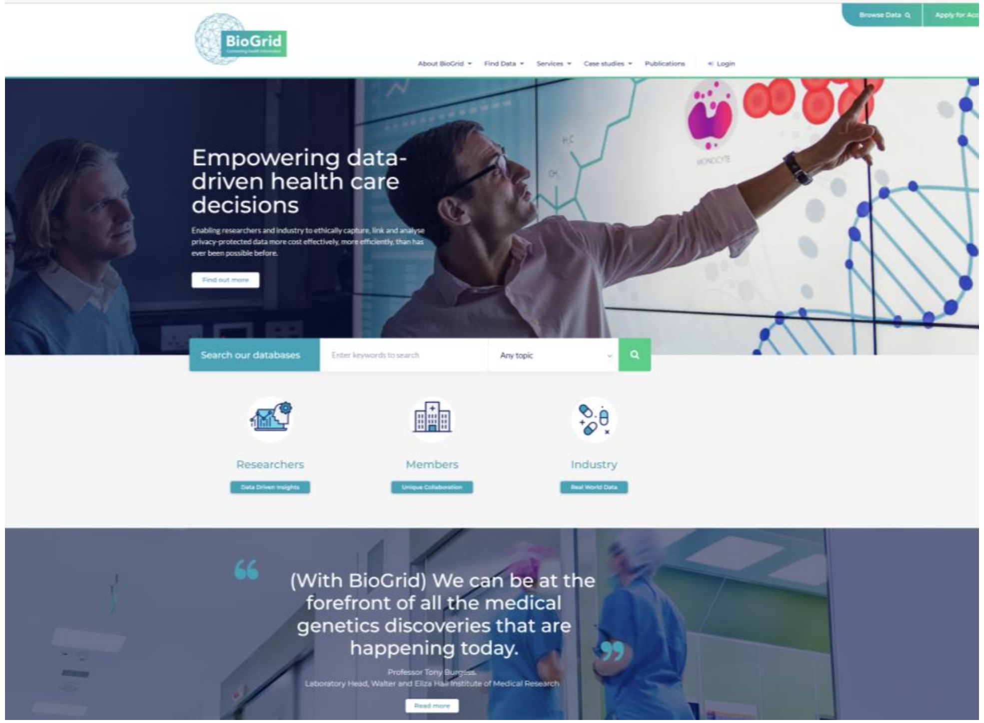
POLAR GP

Advanced data mining and analysis

This "business intelligence tool" is based on extracted data for clinical/billing software, for GPs, Practice Managers and other staff to use within their practice to help provide better patient centred care, quality improvement and support business development.

Other local Data Linkage / GP systems

- Bio Grid :



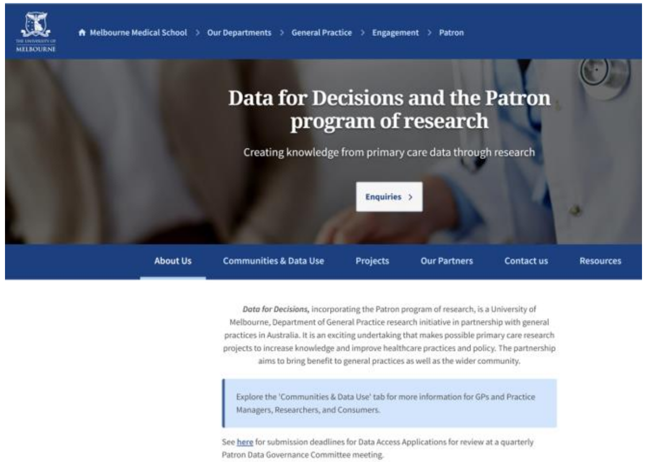
Empowering data-driven health care decisions

Enabling researchers and industry to ethically capture, link and analyse privacy-protected data more cost effectively, more efficiently, than has ever been possible before.

“(With BioGrid) We can be at the forefront of all the medical genetics discoveries that are happening today.”

Professor Tony Burgess.
Laboratory Head, Walter and Eliza Hall Institute of Medical Research

- PATRON



Data for Decisions and the Patron program of research

Creating knowledge from primary care data through research

Data for Decisions, incorporating the Patron program of research, is a University of Melbourne, Department of General Practice research initiative in partnership with general practices in Australia. It is an exciting undertaking that makes possible primary care research projects to increase knowledge and improve healthcare practices and policy. The partnership aims to bring benefit to general practices as well as the wider community.

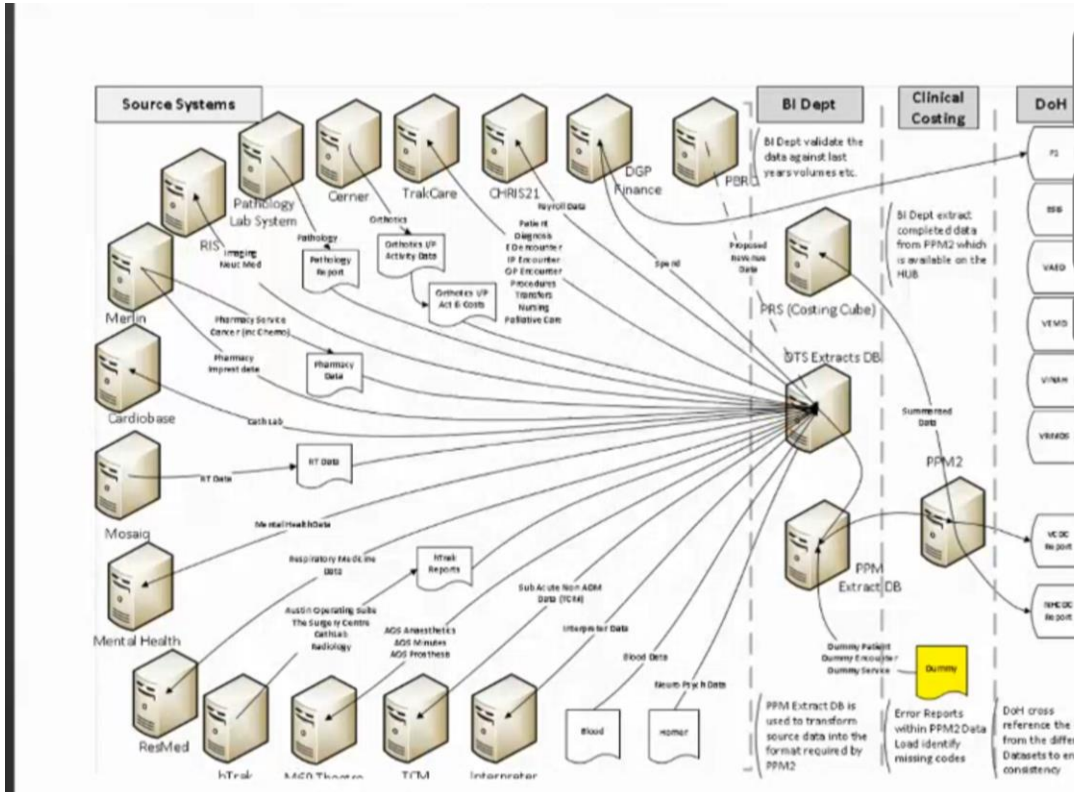
Explore the 'Communities & Data Use' tab for more information for GPs and Practice Managers, Researchers, and Consumers.

See here for submission deadlines for Data Access Applications for review at a quarterly Patron Data Governance Committee meeting.

http://www.datafordecisions.com.au/

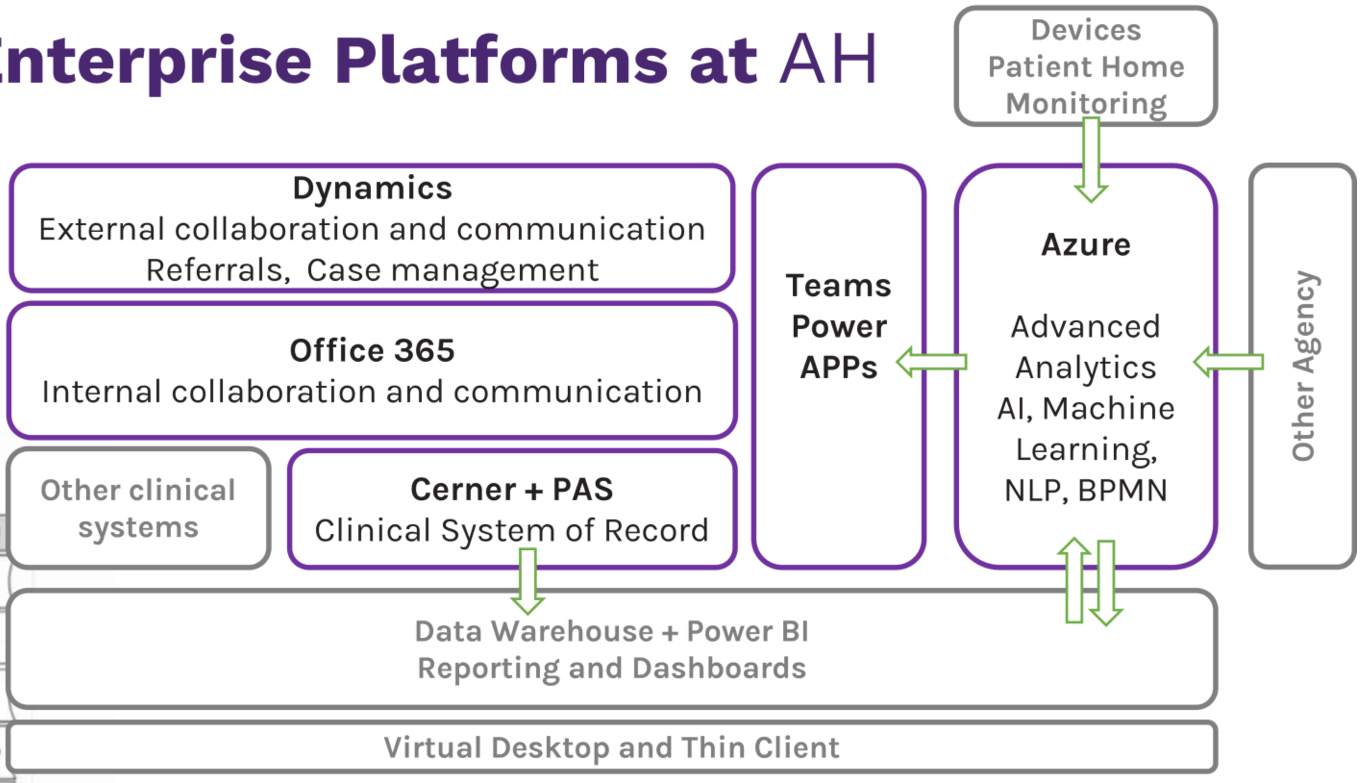
- Largest PBRN in Australia. - 140 practices, 2M patients
- Epidemiological research
- Direct research agreement between participating practices and the University hospital networks..
- Translational research including trialling new care processes, interventions and clinical decision support tools like Future Health Today https://futurehealthtoday.com.au/.

A very complicated Austin Health data “terroir”



Courtesy Ronald Ma, Clinical Costing Specialist, Austin Health

Enterprise Platforms at AH



Devices
Patient Home
Monitoring

Dynamics
External collaboration and communication
Referrals, Case management

Office 365
Internal collaboration and communication

Other clinical
systems

Cerner + PAS
Clinical System of Record

Teams
Power
APPs

Azure

Advanced
Analytics
AI, Machine
Learning,
NLP, BPMN

Other Agency

Data Warehouse + Power BI
Reporting and Dashboards

Virtual Desktop and Thin Client

PRxMs collected via CRM

Courtesy Alan Pritchard
Director EMR and ICT Services, Austin Health

OPEN ACCESS Freely available online

PLOS | ONE

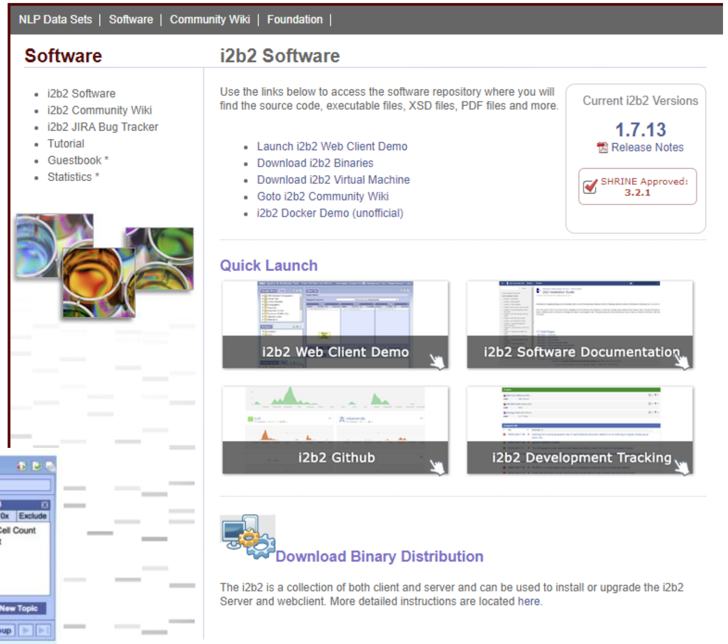
SHRINE: Enabling Nationally Scalable Multi-Site Disease Studies

Andrew J. McMurry[1,2,3,4]*, Shawn N. Murphy[3,5,6], Douglas MacFadden[1], Griffin Weber[3,7], William W. Simons[1], John Orechia[8], Jonathan Bickel[2,9], Nich Wattanasin[5], Clint Gilbert[1], Philip Trevvett[1], Susanne Churchill[3,5], Isaac S. Kohane[1,2,3]

1 Center for Biomedical Informatics, Harvard Medical School, Boston, Massachusetts, United States of America, 2 Children's Hospital Informatics Program, Children's Hospital Boston, Boston, Massachusetts, United States of America, 3 i2b2 National Center for Biomedical Computing, Brigham and Women's Hospital, Boston, Massachusetts, United States of America, 4 Bioinformatics Program, Boston University, Boston, Massachusetts, United States of America, 5 Partners Healthcare System, Research Computing, Boston, Massachusetts, United States of America, 6 Massachusetts General Hospital, Boston, Massachusetts, United States of America, 7 Beth Israel Deaconess Medical Center and Harvard Medical School Information Technology, Boston, Massachusetts, United States of America, 8 Clinical Research Information Technology, Dana-Farber Cancer Institute, Boston, Massachusetts, United States of America, 9 Information Systems Department, Children's Hospital Boston, Boston, Massachusetts, United States of America

Abstract

Results of medical research studies are often contradictory or cannot be reproduced. One reason is that there may not be enough patient subjects available for observation for a long enough time period. Another reason is that patient populations may vary considerably with respect to geographic and demographic boundaries thus limiting how broadly the results apply. Even when similar patient populations are pooled together from multiple locations, differences in medical treatment and record systems can limit which outcome measures can be commonly analyzed. In total, these differences in medical research settings can lead to differing conclusions or can even prevent some studies from starting. We thus sought to create



NLP Data Sets | Software | Community Wiki | Foundation |

Software

- i2b2 Software
- i2b2 Community Wiki
- i2b2 JIRA Bug Tracker
- Tutorial
- Guestbook *
- Statistics *

i2b2 Software

Use the links below to access the software repository where you will find the source code, executable files, XSD files, PDF files and more.

- Launch i2b2 Web Client Demo
- Download i2b2 Binaries
- Download i2b2 Virtual Machine
- Goto i2b2 Community Wiki
- i2b2 Docker Demo (unofficial)

Current i2b2 Versions

1.7.13

Release Notes

SHRINE Approved: 3.2.1

Quick Launch

Download Binary Distribution

The i2b2 is a collection of both client and server and can be used to install or upgrade the i2b2 Server and webclient. More detailed instructions are located here.

pcornet

The National Patient-Centered Clinical Research Network

Data

PCORnet represents data from everyday encounters with more than 30 million people annually across the U.S.

Data is the backbone of PCORnet, and the scale, quality, and security of PCORnet-accessible data is a differentiator for the Network. PCORnet Network Partners perform rigorous work upfront that enables users to ask the same question to millions of people across the United States simultaneously, with fast answers delivered in a single, standardized format.

>30 Million in the Network

PCORnet represents data from everyday encounters with more than 30 million people across the U.S. each year.

Click to enlarge or download

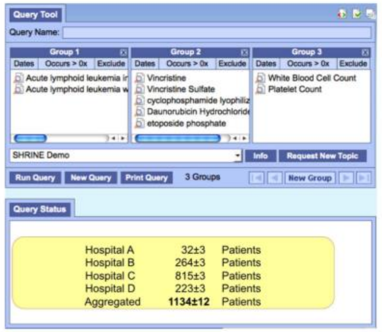
Data Sources

Data accessible via the PCORnet distributed network draw from millions of electronic health records (EHRs) with growing links to patient-reported and payor data to create a powerful, standard data set that facilitates large-scale, multi-site research.

Data Resources

PCORnet Network Partners have developed policies and other critical documentation to ensure the quality, facilitate the accessibility, and govern the use of the PCORnet Common Data Model and resources.

- 200 US AcuteHospitals
- EMR extracts to Common metadata repository
- Researcher Interface allows search criteria
- Platform queries metadata repositories to identify location and number of patients
- Researcher contacts data custodians for permission to contact patients or access to data



Hospital A	32±3	Patients
Hospital B	264±3	Patients
Hospital C	815±3	Patients
Hospital D	223±3	Patients
Aggregated	**1134±12**	Patients

International Data Models for Research

Possible Path through the Tower of Babel??

Hospitals

Cerner

Epic

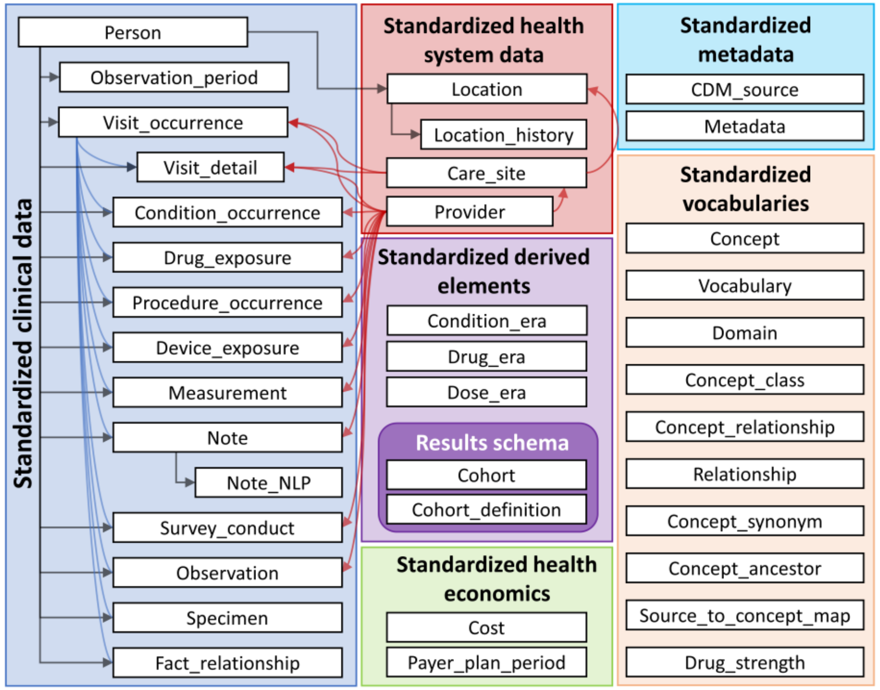
Allscripts

TRAKCare

PAS s

PRxMs

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS


Standardized clinical data
- Person
- Observation_period
- Visit_occurrence
- Visit_detail
- Condition_occurrence
- Drug_exposure
- Procedure_occurrence
- Device_exposure
- Measurement
- Note
- Note_NLP
- Survey_conduct
- Observation
- Specimen
- Fact_relationship

Standardized health system data
- Location
- Location_history
- Care_site
- Provider

Standardized derived elements
- Condition_era
- Drug_era
- Dose_era

Results schema
- Cohort
- Cohort_definition

Standardized health economics
- Cost
- Payer_plan_period

Standardized metadata
- CDM_source
- Metadata

Standardized vocabularies
- Concept
- Vocabulary
- Domain
- Concept_class
- Concept_relationship
- Relationship
- Concept_synonym
- Concept_ancestor
- Source_to_concept_map
- Drug_strength

Clinical
Quality
Registries

Consolidated Reports to Hospitals, Jurisdictions

? Consumers ?

Single Site Outcome Research

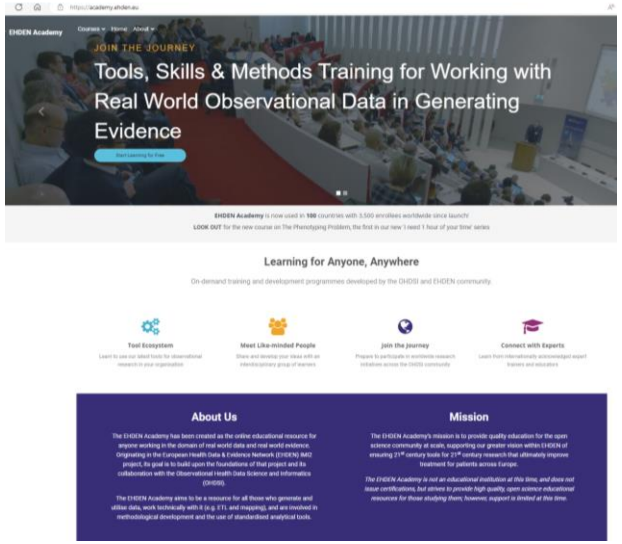
Multicentre Outcome Research

OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

The European Health Data & Evidence Network Portal

The European Health Data & Evidence Network (EHDEN) project aspires to be the trusted observational research ecosystem to enable better health decisions, outcomes and care. Its mission is to provide a new paradigm for the discovery and analysis of health data in Europe, by building a large-scale, federated network of data sources standardized to the OMOP common data model.



JOIN THE JOURNEY

Tools, Skills & Methods Training for Working with Real World Observational Data in Generating Evidence

EHDEN Academy is now used in **100** countries with 3,500 enrollees worldwide since launch!

LOOK OUT for the new course on The Phenotyping Problem, the first in our new 'I need 1 hour of your time' series

Learning for Anyone, Anywhere

On-demand training and development programmes developed by the OHDSI and EHDEN community

Tool Ecosystem
Learn to use our latest tools for observational research in your organisation

Meet Like-minded People
Share and develop your ideas with an interdisciplinary group of learners

Join the Journey
Prepare to participate in worldwide research initiatives across the OHDSI community

Connect with Experts
Learn from internationally acknowledged expert trainers and educators

About Us

The EHDEN Academy has been created as the online educational resource for anyone working in the domain of real world data and real world evidence. Originating in the European Health Data & Evidence Network (EHDEN) IMI2 project, its goal is to build upon the foundations of that project and its collaboration with the Observational Health Data Science and Informatics (OHDSI).

The EHDEN Academy aims to be a resource for all those who generate and utilise data, work technically with it (e.g. ETL and mapping), and are involved in methodological development and the use of standardised analytical tools.

Mission

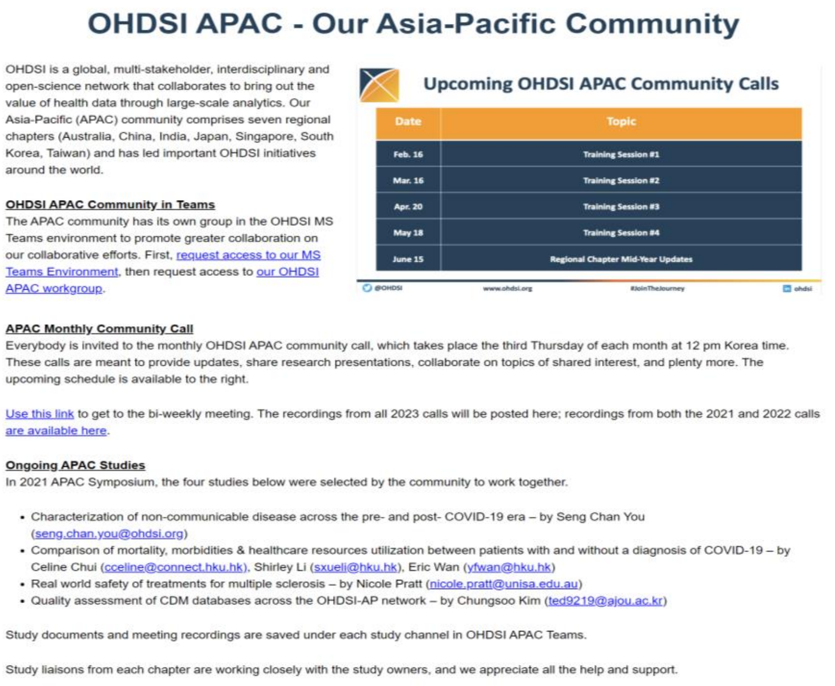
The EHDEN Academy's mission is to provide quality education for the open science community at scale, supporting our greater vision within EHDEN of ensuring 21st century tools for 21st century research that ultimately improve treatment for patients across Europe.

The EHDEN Academy is not an educational institution at this time, and does not issue certifications, but strives to provide high quality, open science educational resources for those studying them; however, support is limited at this time.

www.ohdsi.org

OHDSI APAC conference
July 13-14, 2023

John Niland Scientia
building (The Galleries)
at
UNSW
Randwick Campus.



OHDSI APAC - Our Asia-Pacific Community

OHDSI is a global, multi-stakeholder, interdisciplinary and open-science network that collaborates to bring out the value of health data through large-scale analytics. Our Asia-Pacific (APAC) community comprises seven regional chapters (Australia, China, India, Japan, Singapore, South Korea, Taiwan) and has led important OHDSI initiatives around the world.

OHDSI APAC Community in Teams
The APAC community has its own group in the OHDSI MS Teams environment to promote greater collaboration on our collaborative efforts. First, request access to our MS Teams Environment, then request access to our OHDSI APAC workgroup.

Upcoming OHDSI APAC Community Calls

Date	Topic
Feb. 16	Training Session #1
Mar. 16	Training Session #2
Apr. 20	Training Session #3
May 18	Training Session #4
June 15	Regional Chapter Mid-Year Updates

APAC Monthly Community Call
Everybody is invited to the monthly OHDSI APAC community call, which takes place the third Thursday of each month at 12 pm Korea time. These calls are meant to provide updates, share research presentations, collaborate on topics of shared interest, and plenty more. The upcoming schedule is available to the right.

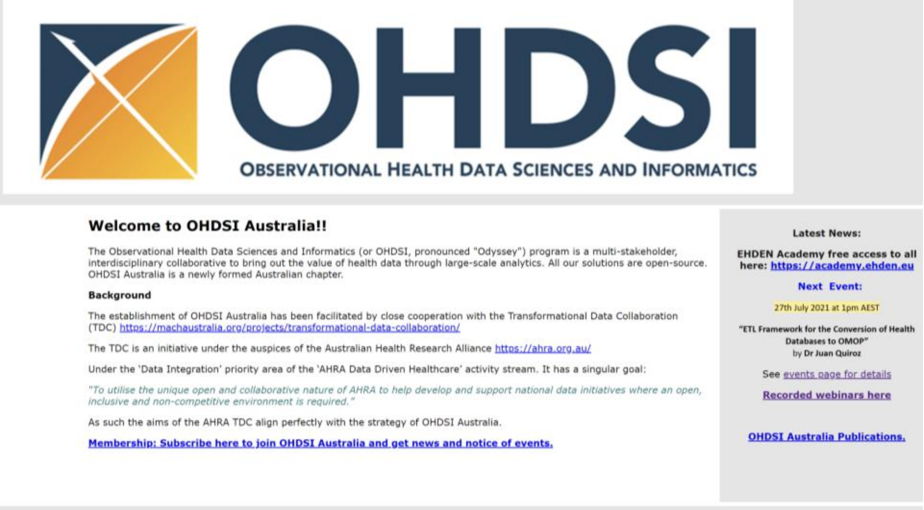
Use this link to get to the bi-weekly meeting. The recordings from all 2023 calls will be posted here; recordings from both the 2021 and 2022 calls are available here.

Ongoing APAC Studies
In 2021 APAC Symposium, the four studies below were selected by the community to work together.

- Characterization of non-communicable disease across the pre- and post- COVID-19 era – by Seng Chan You (seng.chan.you@ohdsi.org)
- Comparison of mortality, morbidities & healthcare resources utilization between patients with and without a diagnosis of COVID-19 – by Celine Chui (cceline@connect.hku.hk), Shirley Li (sxueli@hku.hk), Eric Wan (yfwan@hku.hk)
- Real world safety of treatments for multiple sclerosis – by Nicole Pratt (nicole.pratt@unisa.edu.au)
- Quality assessment of CDM databases across the OHDSI-AP network – by Chungsoo Kim (ted9219@ajou.ac.kr)

Study documents and meeting recordings are saved under each study channel in OHDSI APAC Teams.

Study liaisons from each chapter are working closely with the study owners, and we appreciate all the help and support.



Welcome to OHDSI Australia!!

The Observational Health Data Sciences and Informatics (or OHDSI, pronounced "Odyssey") program is a multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics. All our solutions are open-source. OHDSI Australia is a newly formed Australian chapter.

Background

The establishment of OHDSI Australia has been facilitated by close cooperation with the Transformational Data Collaboration (TDC) https://machaustralia.org/projects/transformational-data-collaboration/

The TDC is an initiative under the auspices of the Australian Health Research Alliance https://ahra.org.au/

Under the 'Data Integration' priority area of the 'AHRA Data Driven Healthcare' activity stream. It has a singular goal:

"To utilise the unique open and collaborative nature of AHRA to help develop and support national data initiatives where an open, inclusive and non-competitive environment is required."

As such the aims of the AHRA TDC align perfectly with the strategy of OHDSI Australia.

Membership: Subscribe here to join OHDSI Australia and get news and notice of events.

Latest News:

EHDEN Academy free access to all here: https://academy.ehden.eu

Next Event:

27th July 2021 at 1pm AEST

"ETL Framework for the Conversion of Health Databases to OMOP"
by **Dr Juan Quiroz**

See events page for details

Recorded webinars here

OHDSI Australia Publications.

What is OMOP?

The Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM)
‘is a standard data schema that uses standardised terms’
to enable the systematic analysis of multiple distinct observational databases

The concept behind the **OMOP approach**:

- Transform data contained within a unique repository (database) into a common format (data model)
- Enable a common representation of terminologies, vocabularies, and coding schemes through harmonisation
- Perform systematic analyses using a library of standard analytic routines using a common format

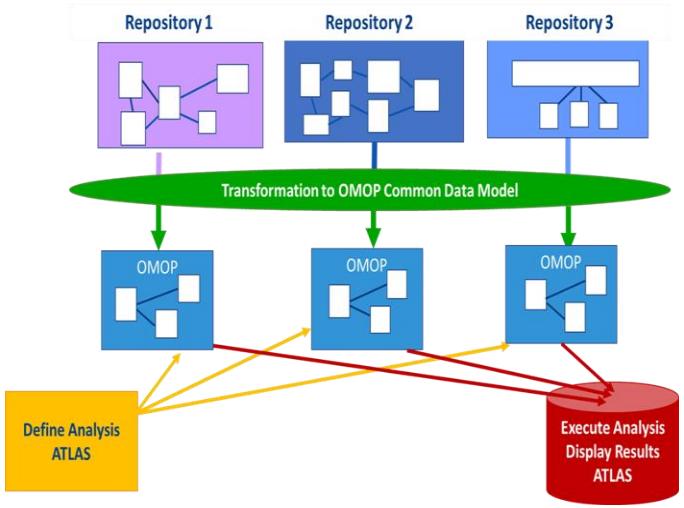
OMOP can be used for:

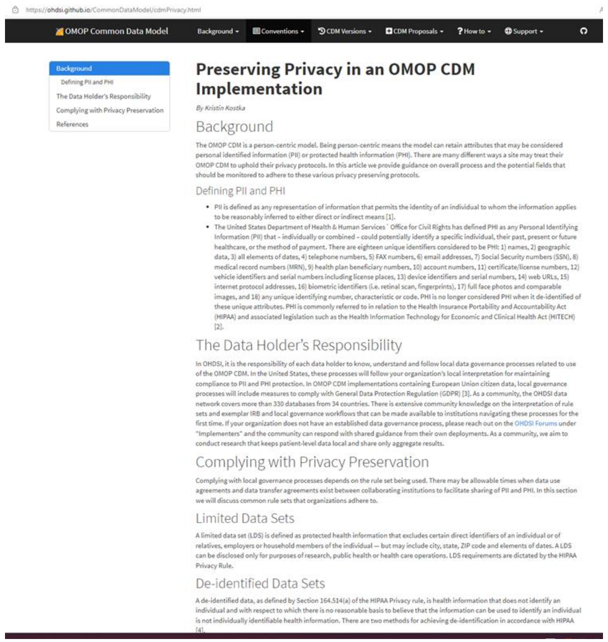
- Independent institutional research that incorporates advanced analytic and prediction techniques
- Research within a research network where research requests are delivered to partner institutions
- Clinical decision support systems for patient-specific medical treatment
- Pharmacovigilance for the active surveillance of drug safety
- Multi-institutional research where ‘validation’ analyses are rapidly conducted under the same conditions using tools such as ATLAS

What is ATLAS?

ATLAS is a web-based search and navigation tool that enables the filtering of data within the OMOP platform.
Once data in OMOP is captured by ATLAS, the tool schematically represents the cohort with tables, graphs, sunburst charts.

What about OHDSI OMOP makes it different ?





Preserving Privacy in an OMOP CDM Implementation

By Kristin Kostka

Background

The OMOP CDM is a person-centric model. Being person-centric means the model can retain attributes that may be considered personal identified information (PII) or protected health information (PHI). There are many different ways a site may treat their OMOP CDM to uphold their privacy protocols. In this article we provide guidance on overall process and the potential fields that should be monitored to adhere to these various privacy preserving protocols.

Defining PII and PHI

- PII is defined as any representation of information that permits the identity of an individual to whom the information applies to be reasonably inferred to either direct or indirect means [1].
- The United States Department of Health & Human Services ' Office for Civil Rights has defined PHI as any Personal Identifying Information (PII) that – individually or combined – could potentially identify a specific individual, their past, present or future healthcare, or the method of payment. There are eighteen unique identifiers considered to be PHI: 1) names, 2) geographic data, 3) all elements of dates, 4) telephone numbers, 5) FAX numbers, 6) email addresses, 7) Social Security numbers (SSN), 8) medical record numbers (MRN), 9) health plan beneficiary numbers, 10) account numbers, 11) certificate/license numbers, 12) vehicle identifiers and serial numbers including license places, 13) device identifiers and serial numbers, 14) web URLs, 15) internet protocol addresses, 16) biometric identifiers (i.e. retinal scan, fingerprints), 17) full face photos and comparable images, and 18) any unique identifying number, characteristic or code. PHI is no longer considered PHI when it de-identified of these unique attributes. PHI is commonly referred to in relation to the Health Insurance Portability and Accountability Act (HIPAA) and associated legislation such as the Health Information Technology for Economic and Clinical Health Act (HITECH) [2].

The Data Holder's Responsibility

In OHDSI, it is the responsibility of each data holder to know, understand and follow local data governance processes related to use of the OMOP CDM. In the United States, these processes will follow your organization's local interpretation for maintaining compliance to PII and PHI protection. In OMOP CDM implementations containing European Union citizen data, local governance processes will include measures to comply with General Data Protection Regulation (GDPR) [3]. As a community, the OHDSI data network covers more than 330 databases from 34 countries. There is extensive community knowledge on the interpretation of rule sets and exemplar IRB and local governance workflows that can be made available to institutions navigating these processes for the first time. If your organization does not have an established data governance process, please reach out on the OHDSI Forums under "Implementers" and the community can respond with shared guidance from their own deployments. As a community, we aim to conduct research that keeps patient-level data local and share only aggregate results.

Complying with Privacy Preservation

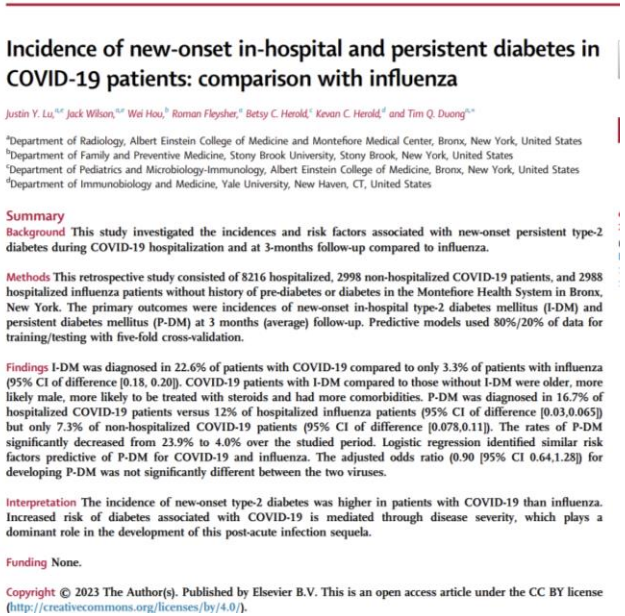
Complying with local governance processes depends on the rule set being used. There may be allowable times when data use agreements and data transfer agreements exist between collaborating institutions to facilitate sharing of PII and PHI. In this section we will discuss common rule sets that organizations adhere to.

Limited Data Sets

A limited data set (LDS) is defined as protected health information that excludes certain direct identifiers of an individual or of relatives, employers or household members of the individual — but may include city, state, ZIP code and elements of dates. A LDS can be disclosed only for purposes of research, public health or health care operations. LDS requirements are dictated by the HIPAA Privacy Rule.

De-identified Data Sets

A de-identified data, as defined by Section 164.514(a) of the HIPAA Privacy rule, is health information that does not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual is not individually identifiable health information. There are two methods for achieving de-identification in accordance with HIPAA [4].



Incidence of new-onset in-hospital and persistent diabetes in COVID-19 patients: comparison with influenza

Justin Y. Lu,[a,e] Jack Wilson,[a,e] Wei Hou,[b] Roman Fleysher,[a] Betsy C. Herold,[c] Kevan C. Herold,[d] and Tim Q. Duong[a,]*

[a]Department of Radiology, Albert Einstein College of Medicine and Montefiore Medical Center, Bronx, New York, United States
[b]Department of Family and Preventive Medicine, Stony Brook University, Stony Brook, New York, United States
[c]Department of Pediatrics and Microbiology-Immunology, Albert Einstein College of Medicine, Bronx, New York, United States
[d]Department of Immunobiology and Medicine, Yale University, New Haven, CT, United States

Summary

Background This study investigated the incidences and risk factors associated with new-onset persistent type-2 diabetes during COVID-19 hospitalization and at 3-months follow-up compared to influenza.

Methods This retrospective study consisted of 8216 hospitalized, 2998 non-hospitalized COVID-19 patients, and 2988 hospitalized influenza patients without history of pre-diabetes or diabetes in the Montefiore Health System in Bronx, New York. The primary outcomes were incidences of new-onset in-hospital type-2 diabetes mellitus (I-DM) and persistent diabetes mellitus (P-DM) at 3 months (average) follow-up. Predictive models used 80%/20% of data for training/testing with five-fold cross-validation.

Findings I-DM was diagnosed in 22.6% of patients with COVID-19 compared to only 3.3% of patients with influenza (95% CI of difference [0.18, 0.20]). COVID-19 patients with I-DM compared to those without I-DM were older, more likely male, more likely to be treated with steroids and had more comorbidities. P-DM was diagnosed in 16.7% of hospitalized COVID-19 patients versus 12% of hospitalized influenza patients (95% CI of difference [0.03,0.065]) but only 7.3% of non-hospitalized COVID-19 patients (95% CI of difference [0.078,0.11]). The rates of P-DM significantly decreased from 23.9% to 4.0% over the studied period. Logistic regression identified similar risk factors predictive of P-DM for COVID-19 and influenza. The adjusted odds ratio (0.90 [95% CI 0.64,1.28]) for developing P-DM was not significantly different between the two viruses.

Interpretation The incidence of new-onset type-2 diabetes was higher in patients with COVID-19 than influenza. Increased risk of diabetes associated with COVID-19 is mediated through disease severity, which plays a dominant role in the development of this post-acute infection sequela.

Funding None.

Open Source
Strong international User Community
Repository of Tools to Share
Extensible – NLP, ML

Rapid Response to new Health Threats
Standing Infrastructure
Strong Research Community
Sharable Resources

OMOP Transformations underway

- ARDC Funded
- CERNER sites
  - Queensland Health
  - University of NSW hospital affiliates
  - Austin Health
  - Western Health

- Independently Funded
  - Parkville Precinct (EPIC)
  - Patron GP

Matters of Importance - OMOP Patron Data Governance

DATA GOVERNANCE

Patron OMOP data may only be utilised for projects *approved* by the Patron Data Governance Committee.

Re-use of the same OMOP dataset for a new project *requires* new ethics and Patron Data Governance Committee approvals.

Project scope is shaped by the *limits* placed on the project by the Patron Data Governance Committee.

Patron Data Governance Committee considerations include the:

- Nature of a prospective researchers IT system
- Quality of systems and protections used to safeguard data
- Nature of the research institution
- Research objectives, proposed uses, data disclosure, and commercial entity involvement
- Prior contraventions of privacy laws or data breaches

CONSENT

Patron-OMOP practices need to provide consent for their practice data to be accessed for research purposes. Patron practices can change their consent options or withdraw their consent to participate at any time, without prejudice.

Opt out

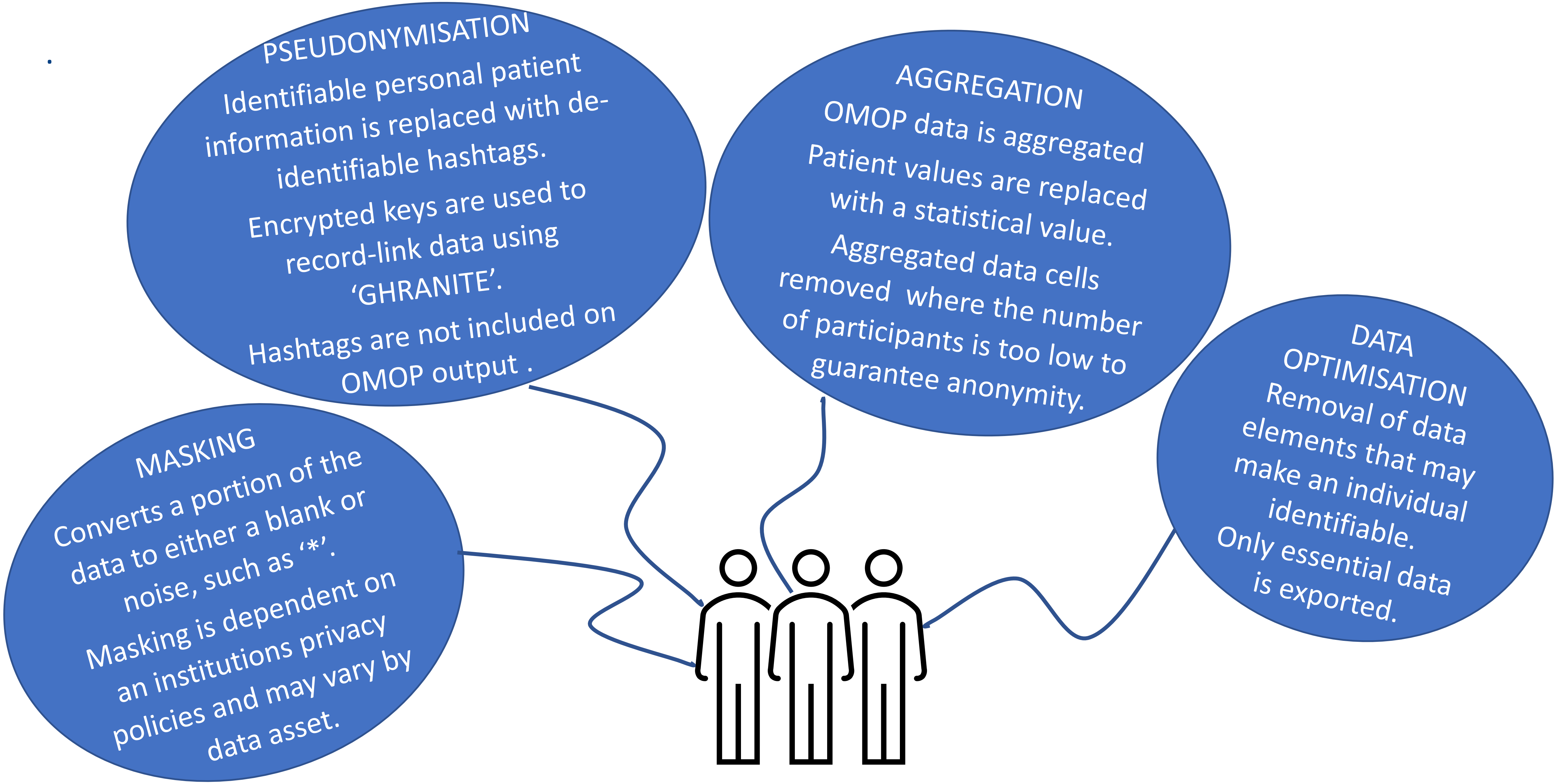
If a research project is deemed by an ethics committee to be of low risk to the individual - consent is conventionally managed in an opt-out manner.

Waiver of consent

It is not always possible to obtain patient consent. If certain criteria are met, waivers of consent will be granted by an ethics committee. This is a common model for Patron-OMOP projects.

Data Governance – How Patron OMOP Anonymises Personally Identifiable Information

.



PSEUDONYMISATION

Identifiable personal patient information is replaced with de-identifiable hashtags.

Encrypted keys are used to record-link data using ‘GHRANITE’.

Hashtags are not included on OMOP output .

AGGREGATION

OMOP data is aggregated

Patient values are replaced with a statistical value.

Aggregated data cells removed where the number of participants is too low to guarantee anonymity.

DATA OPTIMISATION

Removal of data elements that may make an individual identifiable.

Only essential data is exported.

MASKING

Converts a portion of the data to either a blank or noise, such as ‘*’.

Masking is dependent on an institutions privacy policies and may vary by data asset.

Compatibility with International Governance Standards

The FAIR Guiding Principles

F	Findability	Metadata and data should be easily found by both humans and computers through the assignment of globally unique and permanent identifier to enable the automatic discovery of datasets and services via machine learning [31, 32].
A	Accessibility	Metadata and data should easily retrieved by authorised and authenticated users via a standard communication protocol[31, 32] .
I	Interoperabilty	Data from one data source can be integrated with data from other sources so that it can be aggregated into a single, unified view, and refers to the integration and exchange of applications, analysis, storage, and workflow processing across different data sources[31, 32].
R	Reusability	Metadata and data characteristics is specified in detail to enable replication and or linkage in different settings. Reusability includes the release of data usage licenses, provenance details, and disclosure around community standards relevant to the domain [31, 32].

The Five Safes Framework

People	Safe People Is the researcher appropriately trained and authorised to access and use the data? [34]
Projects	Safe Projects Is data used for an appropriate purpose that is valid and of public benefit ? [34]
Settings	Safe Settings Does IT access and physical environment prevent unauthorised use? [34]
Data	Safe Data Has appropriate and sufficient protection been applied to the data to avoid risk of disclosure? [34]
Outputs	Safe Outputs Are the statistical results non-disclosive? [34]

The CARE Principles

C	Collective Benefit	Collective benefit include that where the wellbeing of Indigenous Peoples' rights and are of primary concern[33]
A	Authority	Indigenous Peoples' rights and interests in data about their peoples, communities, cultures, and territories is part of reclaiming control of data is clearly articulated[33]
R	Researcher	Researchers have a responsibility to develop and nurture respectful relationship with Indigenous Peoples' from whom the data originate [33]
E	Ethics	Minimise harm and maximise benefit for Indigenous Peoples, for justice and future use [33]

Summary

Health data is ubiquitous but heavily siloed

Research and Innovation capacity and efficiency is constrained

Silos, data conformance, costs of "research readiness", Privacy fears

Complexity and Commercial Designs of EMR vendor databases

Cost, time, reproducibility

Adding Value

OHDSI OMOP CDM offers a real opportunity to simplify and streamline

Advantages for large scale, rapid implementation multicentre research