Singapore Open Research Conference 2024

Nov 12th, 10:00 AM - 10:20 AM

Is Open Science Possible in Healthcare and Medical Research?

Wilson GOH
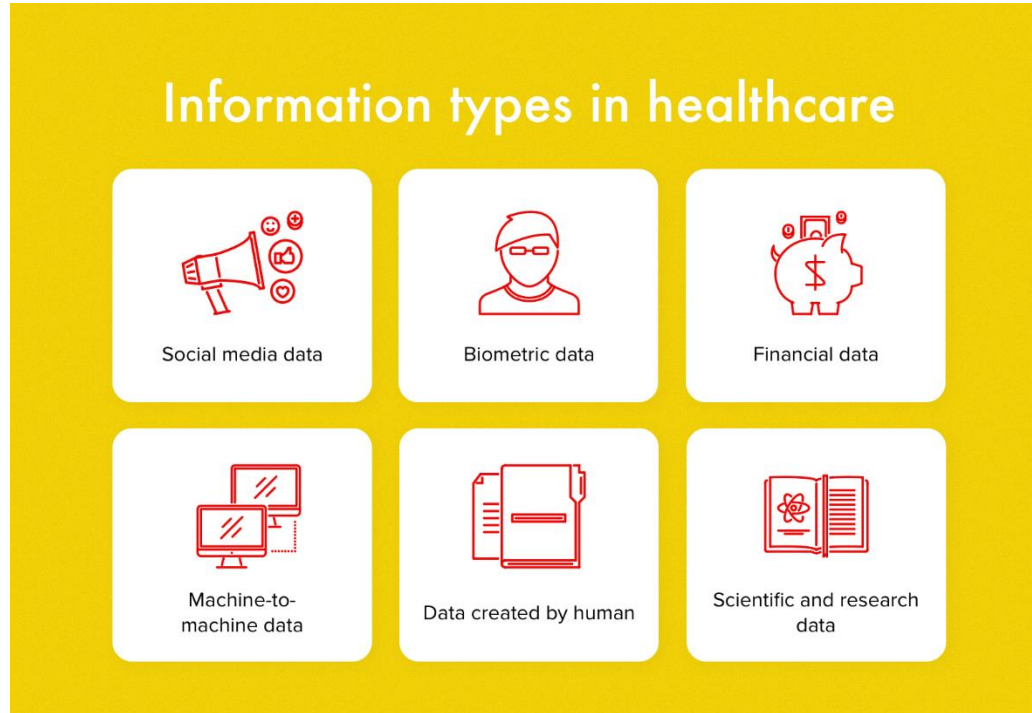*Center of AI in Medicine, Nanyang Technological University (NTU)*

Is open science possible in healthcare and medical research?

Wilson Wen Bin Goh

Chief Data Scientist, NTU Center of AI in Medicine. Head, Data Science and AI Programme, Nanyang Technological University, Singapore

Honorary Senior Lecturer (Associate Professor) Imperial College London, United Kingdom

REDEFINING MEDICINE, TRANSFORMING HEALTHCARE

Healthcare and Medical Data are varied

Information types in healthcare

Social media data · Biometric data · Financial data

Machine-to-machine data · Data created by human · Scientific and research data

https://www.learntek.org/

Benefits of using data in healthcare
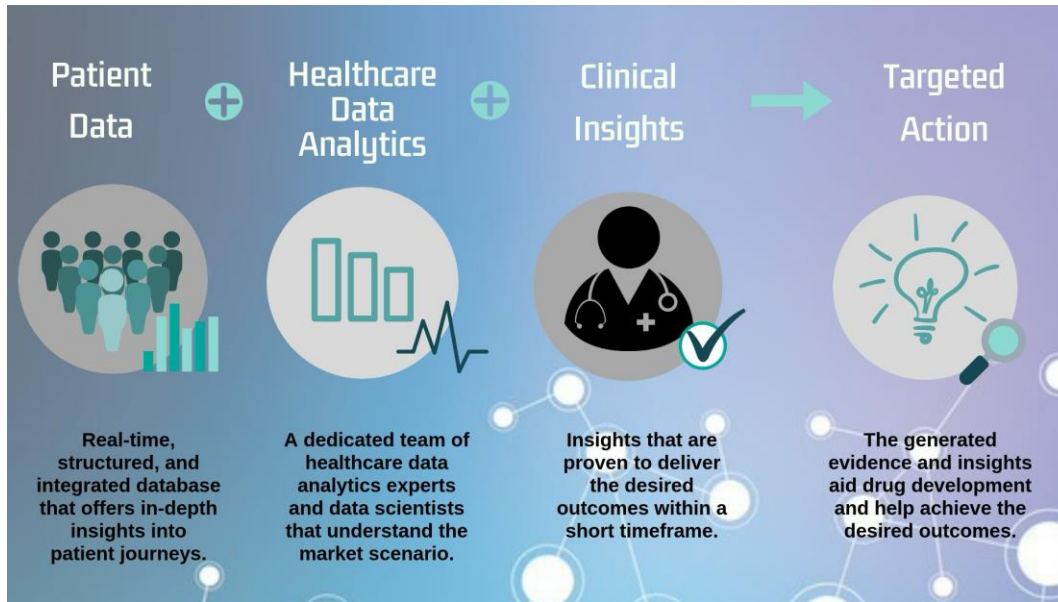
Using patient healthcare data

Patient/Public Concerns

- Safety
- Privacy
- Autonomy

Technology Concerns

- Cybersecurity
- Data Quality and Inter-operability
- Model explainability and interpretability

## Ethical Concerns

- Transparency
- Reproducibility
- Accountability
- Bias, inequities, inequalities

Safeguarding Approaches

- Trusted research environments or TREs
  - (very closed)
  - E.g. MOH Trust
- Open consortias
  - very open
  - E.g. UKBB, MIMIC, UKDRI
- Dynamic provenance and provisioning (hybrid)
  - More relaxed than TREs
  - But riskier and may not be validatable in various contexts

MOH Trust
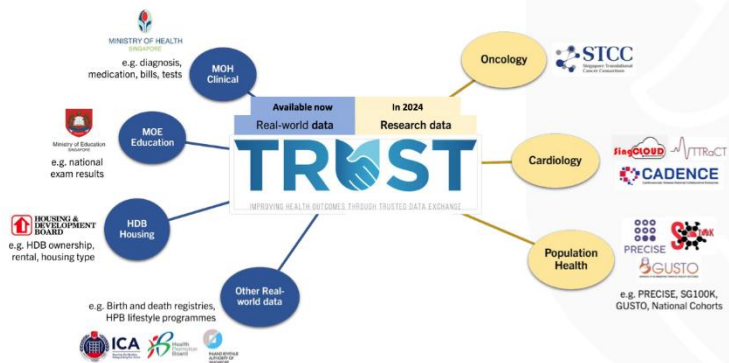
- National health-data exchange platform to facilitate secure data linkage and analysis of anonymised research and real-world data between the public and private sectors.

- Supports research under the Research, Innovation and Enterprise (RIE) 2025 Human Health & Potential (HHP) domain's focus and priorities.
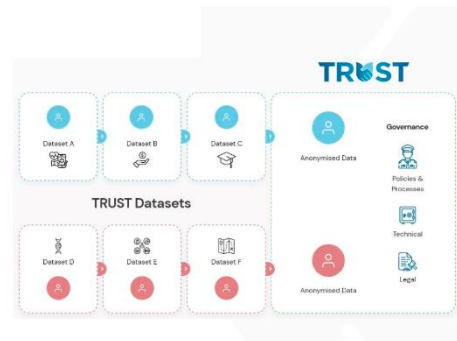
How TRUST ensures data is safe

SAFE PURPOSE
All data request will be reviewed by TRUST Data Access Committee to ensure that purpose of use fulfils public interest and social value.

SAFE PEOPLE
TRUST users must have appropriate credentials for access to TRUST and the approved data for research.

SAFE SETTINGS
TRUST is hosted in a secure environment with government-standard security measures.

SAFE DATA
All data accessed on TRUST are anonymised to government standard to reduce re-identification risks.

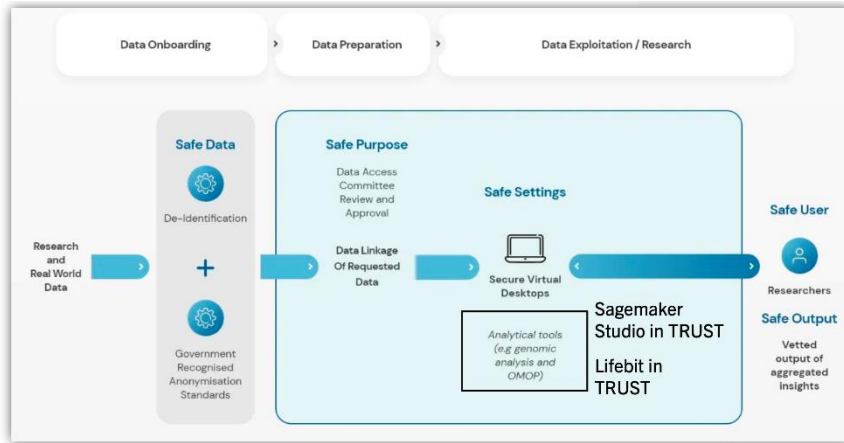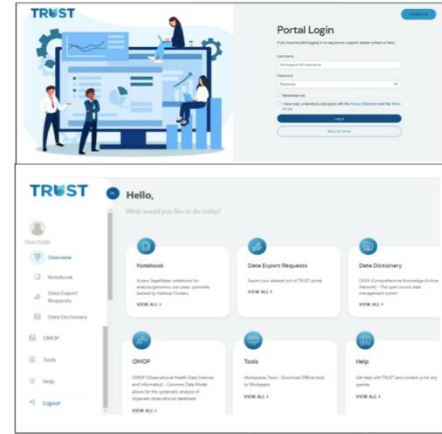**SAFE OUTPUT**
Only verified aggregate data and insights with low re-identification risk can be output.

How to ensure data safety

Data onboarding ensures adherence to privacy and safety standards

Data access is safeguarded by policy/regulatory, digital and physical safeguards
Anything to be taken out needs permission
Publications to be vetted before submission

MOH Trust is well received

Enabling health-data analytics and research

- Since launched, TRUST has approved **8 data requests** and supported approximately **70 users** on their research analytics.
- We are actively engaging researchers to understand their research questions and to support their data needs.
- We strongly encourage all researchers to reach out to TRUST Data Concierge for early review and advice for data request to ensure timely review by DAC.
- TRUST 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

A *TRUST Data Concierge* team supports users throughout their journey
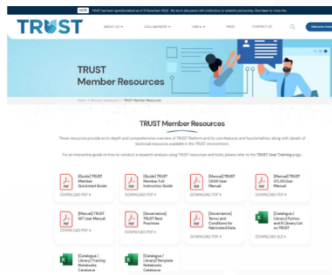
Improve Data Quality And Interoperability ⌃

- ✓ Receive support from TRUST's data cleaning team to clean their data in accordance with a **harmonised format** that meets international standards.
- ✓ Improved data quality while retaining their rights to the ownership of **their cleaned dataset.** Data Contributors will always retain the free and unfettered right to use their cleaned dataset at their own discretion for any purpose outside of TRUST.

Gain Access To Real-World Data ⌄

Retain Exclusive Use Of Their Data During Exclusivity Period* ⌄

Expand Collaboration Network After Exclusivity Period* ⌄
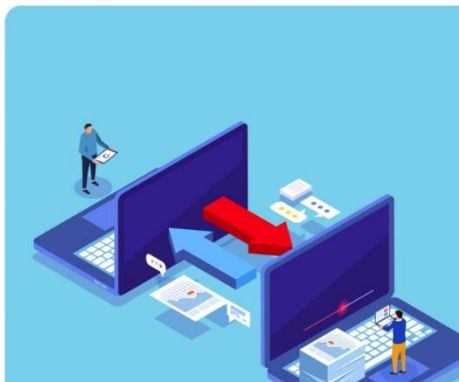
NEW – Demographics, co-morbidities, medications, healthcare service utilisation, health outcomes and healthcare cost of patients with respiratory diseases seen in the primary care

[Learn More]

NEW – Examining factors influencing COVID-19 vaccine adoption through multivariate, time-series and multi-country analysis

[Learn More]

NEW – Studying the short and long-term impact of climate and climate change on infectious and chronic diseases

[Learn More]

NEW – Population-based Observational Study on the Utilisation of Acute Hospital Beds Among Singapore Residents & Study the Utilisation of Hospital-at-home (HaH)

[Learn More]

Predicting the risk of systemic diseases using eye images: Clinical and Cost-Effectiveness

[Learn More]

Long-term outcomes of heart failure cohort (ATTRaCT) and health services analysis

[Learn More]

Benefits to both accessor and contributor

Alternatives to closed systems

- Traditional legal agreements (DSA/MTA, RCAs)
  - Cannot really enforce or control what people do with the data
  - Cannot check or determine in real time if correct permissions and correct use are given
  - Cannot update the data access permissions as circumstances change
  - Cannot update the shared data easily
  - Cannot embed governance principles for data and models
- Deploying a system for dynamic provision and provenance, with AI-based governance tools in place
  - Nascent area
  - More relaxed than a pure TRE
  - Cybersecurity issues are being tackled
  - Real world evidence from use cases, disaster scenarios, useability and deployments needed
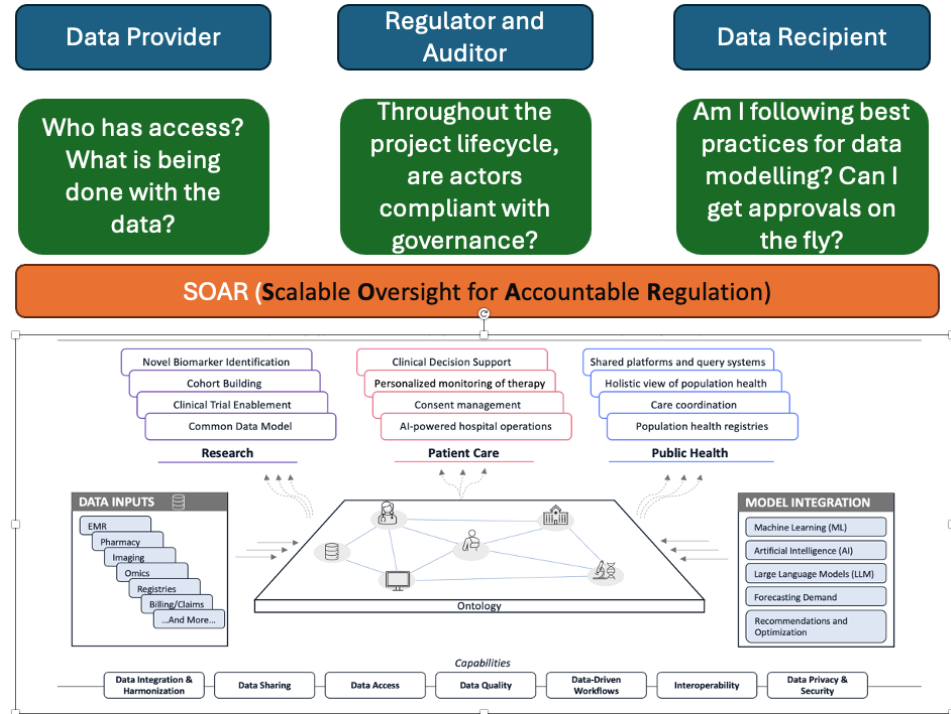
Ongoing…

Public and multi-stakeholder engagement is needed

Public deliberation in the use of health and care data

Delivered by
ONE LONDON Ipsos The King's Fund>

Supported by
LONDON COUNCILS SUPPORTED BY MAYOR OF LONDON NHS

Executive summary

Use of de-personalised data for proactive care

Once the concept was explained, the use of de-personalised data to support proactive care was considered low-risk and its preventative aspect was positively received. This issue proved uncontentious and therefore did not require further deliberation on Day 4 of the Summit.

Recommendation and conditions:

There was overall acceptance for de-personalised data to be used for proactive care, and this came with specific conditions. These conditions aligned with those specified for other uses of de-personalised data, i.e. planning and research.

- Data must not be shared with, or sold to, insurance companies or for marketing purposes

- Data should not be used by policymakers for reasons which are not in the public's interest, and there need to be severe punishments for misuse

- There should be reassurance and information about the process for de-personalising data; who is involved and how is this done

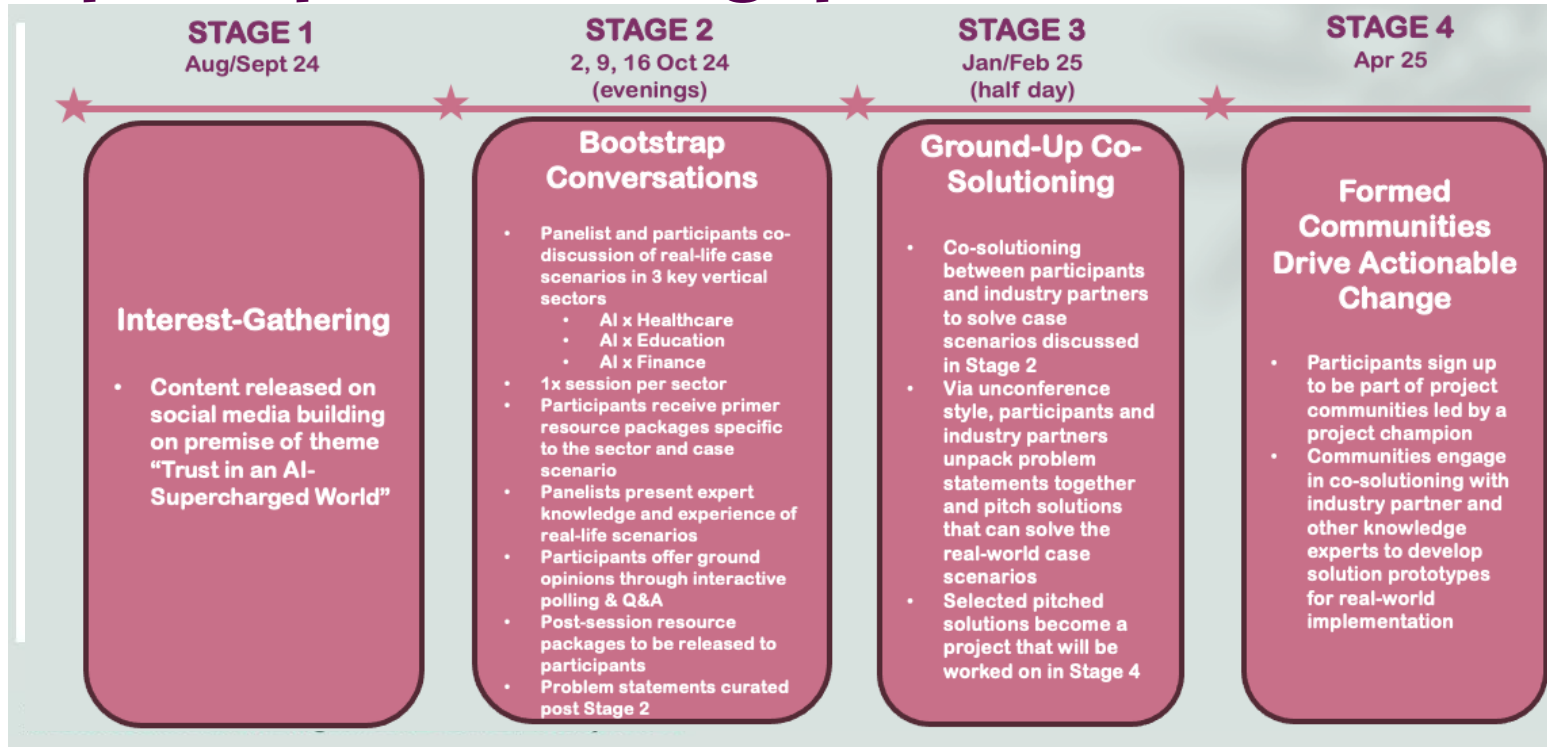### Use of de-personalised data for health and care planning and improvement

Recommendation:

De-personalised health and care data must be shared and used by relevant bodies to plan and improve services and demonstrably benefit health in London.

Conditions:

- Ensure the data being shared is accurate

- All parties must sign up to the 'five safes'[3] in order to access/use data

- Data is shared for an agreed purpose (not for general analysis)

- Only data relevant to the specific analysis are shared

- Shared data must be held securely by all agreed recipients

- Maintain the national 'opt out' option, but also provide an option to opt 'back in'

- There should be legal penalties for misuse of the data (e.g. selling on to 'third parties')

- The organisations that are sharing/using the data should be stated and published somewhere for the public to see

- There should be a defined process for sharing the data that is published somewhere for the public to see

https://imperialcollegehealthpartners.com/portfolio/onelondon/

Catalyst@NL (To get a sense of public perception in Singapore)

STAGE 1
Aug/Sept 24

STAGE 2
2, 9, 16 Oct 24
(evenings)

STAGE 3
Jan/Feb 25
(half day)

STAGE 4
Apr 25

Interest-Gathering

- Content released on social media building on premise of theme "Trust in an AI-Supercharged World"

Bootstrap Conversations

- Panelist and participants co-discussion of real-life case scenarios in 3 key vertical sectors
    - AI x Healthcare
    - AI x Education
    - AI x Finance
- 1x session per sector
- Participants receive primer resource packages specific to the sector and case scenario
- Panelists present expert knowledge and experience of real-life scenarios
- Participants offer ground opinions through interactive polling & Q&A
- Post-session resource packages to be released to participants
- Problem statements curated post Stage 2

Ground-Up Co-Solutioning

- Co-solutioning between participants and industry partners to solve case scenarios discussed in Stage 2
- Via unconference style, participants and industry partners unpack problem statements together and pitch solutions that can solve the real-world case scenarios
- Selected pitched solutions become a project that will be worked on in Stage 4

Formed Communities Drive Actionable Change

- Participants sign up to be part of project communities led by a project champion
- Communities engage in co-solutioning with industry partner and other knowledge experts to develop solution prototypes for real-world implementation

Key Takeaways

- **Health and medical data is highly multi-dimensional**. Its use for analytics and modelling is fraught with ethical, security and technological concerns

- **Trusted research environments (TREs) are the mainstay for now**. These enforce regulatory, physical and cyber security standards. Benefit to both contributors and users by means of data reuse and sharing. But slow..

- **Alternatives to TREs may arrive in the near future**. Using AI and automation to embed governance and safety while providing more relaxed access

- **Multi-stakeholder engagement and with public can help inform on policies**. Are we needlessly strict? What does the patients and public want? Do they want to be more actively engaged?

Acknowledgements

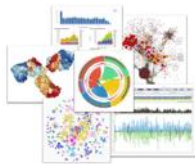Biodata science and education

KEDRI

AI and Proteomics (pi-hub)

Center of AI in medicine

Center for Biomedical Informatics

AI Trust and Health Behavior

Research Fellows
- Allen Chong
- Peng Hui
- Chan Wei Xin
- Neamul Nabir
- Stefano Perna
- James Miller
- Shelly Malik

Research Associates
- Kendrick Chia
- Kalya Kee
- Syahirah
- Phua Serxian

Collaborators
- Joseph Sung
- May Lwin
- Peter Schulz
- Nikola Kasabov
- He Fu Chu
- Tiannan Guo

PhD Students
- Samuel Tan
- Kong Weijia
- Tian Siqi
- Jordon Kho
- Harvard Hann
- Sugam Budhraja
- Zhou Ruwen
- Zhou Ziyuan
- Teddy Tng
- Sehwan Yoo
- Joan Jong
- Song Shangzheng
- Liu Yihang
- Liu Chen
- Cameron Choo
- Kon Wenxuan
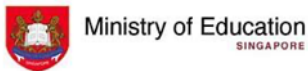
LEE KONG CHIAN SCHOOL OF MEDICINE

Thank you

wilsongoh@ntu.edu.sg