Enabling data driven outcomes: Safe havens

Anitha Kannan
Jerico Revote
Komathy Padmanabhan
Prof. Paul Bonnington
Dr. Steve Quenette
Setting the context..

Keynote
- Potential benefits of sharing data
- Challenges and success stories

BoF
- Defining sensitive data
- Handling sensitive data

Talks
- National ethics policy for managing and sharing data
- Enabling access to sensitive data at Australian Data Archive

#eResAU2018

References:
1. Enabling Data sharing for research, intelligence and public benefit, Prof. David V Ford, Swansea University medical school
2. Sensitive Data – How do you do yours? Frankie Stevens¹, Jeff Christiansen², Kate Le May³, Steve McEachern⁴, Angela Gackle⁵, Ingrid Mason⁶
3. Changes in national ethics policy for managing and sharing human research data Kate LeMay¹, Australian Research Data Commons, Canberra,
4. Enabling access to sensitive data at the Australian Data Archive Dr Steven McEachern¹, Ms Janet McDougall¹, Ms Marina McGale¹

¹Australian Data Archive, ACT, Australia
This talk..

- Quick intro to data safe havens
- Monash eResearch Centre’s approach
- Underlying principles
- Our Journey
  - Innovation pathway
  - Collaboration pathway
- Future landscape
What are data safe havens?

Proposed criteria for a Data Safe Haven

**Data maintenance and release should be socially acceptable and appropriate**
- Criterion 1: Consistent with formal ethical and legal requirements
- Criterion 2: Responsive to emerging issues
- Criterion 3: Discoverable and accessible
- Criterion 4: Transparent and accountable

**Data should be verifiable**
- Criterion 5: Data and metadata fidelity
- Criterion 6: Quality assurance and control
- Criterion 7: Curation and archiving
- Criterion 8: Reliable availability including backup
- Criterion 9: Effective audit

**Data should be safe and secure**
- Criterion 10: Preserve confidentiality, integrity and availability of the repository
- Criterion 11: Appropriate secure access to individually identifying data
- Criterion 12: Appropriate protection of individually identifying data

Our 2-pronged approach

Leverage on existing strengths

Co-design

Build on existing technologies to create seamless workflow

Scale & Customise

Establish gold class standards

Partner with outliers

Frugal Innovation

Collaboration
Underlying principles

1. Safe people
2. Safe projects
3. Safe settings
4. Safe data
5. Safe output

How we started

Big data mining market segmentation of ANZ Bank EFTPOS data

In Australia, the big 4 banks receive large amounts of Electronic Funds Transfer at Point of Sale (EFTPOS) transaction data on a daily basis, but despite this, this information-rich data are not stored nor analysed. The fact that EFTPOS data is both very large and very messy makes it difficult for banks themselves to gain visibility of the characteristics of the stakeholders of the data.

That changed in 2014, when a researcher in Monash’s Faculty of IT, Dr. Grace Rumanis, approached us for assistance in accessing/building a secure analysis environment for a data mining project on a collection of commercially sensitive EFTPOS data obtained through an award winning collaboration with the Australia and New Zealand Banking Group (ANZ). To our knowledge this is the first time market segmentation analyses have been applied to such a large amount of EFTPOS data anywhere in the world.

TWO-TIERED MARKET SEGMENTATION FRAMEWORK

Acquisition and secure storage of de-identified EFTPOS data → EFTPOS Data reduction using RFM analysis → Retailer segmentation via clustering → Finding insights into retailer segments via tree-based classification

2014
Secure VLs for mining the EFTPOS data of ANZ Bank for market segmentation analysis

2015
Secure patient reports for patients and treating team from the survey data
Innovation: Flagship ASPREE Safe Havens
Partnership initiative between Monash eResearch & ASPREE

Background

• ASPREE (ASPirin in Reducing Events in Elderly) is the largest primary prevention aspirin study ever undertaken in healthy older people

• 19,000 Participants – ~ 16500 Australians and 2500 Americans

• 17 sub-studies including Genomics, Hearing, Neuro, Biobank etc.

Challenges

• ASPREE required a secure, curated platform for analysis of a re-identifiable baseline data set. Applicants included international non-Monash collaborators.

• Applications required approval by several committees before data could be provided.

• Limited resources available to manage import and export of data
Innovation: Flagship ASPREE Safe Havens
Partnership initiative between Monash eResearch & ASPREE

Access Management System

- Developed by ASPREE based on their governance policies
- Data custodians publish data catalogue
- Researchers submit expression of interest for the data
- Application reviewed and approval by local & international steering committees, Ethics committee etc. as per governance policies
- Applicant notification on approval and instructions to login to safe havens
Innovation: Flagship ASPREE Safe Havens
Partnership initiative between Monash eResearch & ASPREE
2016-2017

Safe Havens access gateway

- Virtual lab of Safe Havens
- Accessed through 2-factor authentication to Monash VPN
- Potential to scale it

Remote Safe Havens

- Access to approved data
- Range of analysis tools
- Options to request export of analysed data
Underlying architecture

2016-2017
Details of 12 countries nominating to participate in TrueNTH global registry and distribution of 113 participating sites across the 25 local data centres in 12 countries

- Reversibly anonymised patient data from all the 12 countries are collected in central data co-ordinating centre at Monash.
- Data received through a secure online portal.
- Data sharing with central project coordination centre in US and with researchers through safe haven.
- Key criteria for being central data coordination centre is having strict security control and capability to securely share back sensitive data with participating sites for analysis.

Monash SeRP – A Swansea collaboration

UK Secure eResearch Platform

The UK Secure Research Platform (UKSeRP) developed by the Health Informatics Group at Swansea

UKSeRP is an ISO27001 approved independent and customisable technology and analysis platform to allow multiple, complex datasets to be managed, analysed and shared safely, subject to legislative and regulatory requirements


Helix, a health data ecosystem operated by expert staff and underpinned by world-class research infrastructure, spearheads a new era in healthcare research that will drive personalised medicine and significantly accelerate progress towards a healthier society. This capability is an initiative of Monash University working collaboratively with key health partners, including Alfred, Eastern, Monash and Peninsula Health.

HEALTH DATA CAPABILITIES
- Secure data management - REDCap
- Goverred environments for data sharing - Monash SeRP
- Secure virtual environments for statistical data analysis
- High performance computing, Research Cloud and Research data storage - eResearch Centre and MASSIVE

EXPERTISE
The combination of the sheer scale of Monash University’s health datasets - which collate hundreds of thousands of data points within clinical trials, medical and population studies - underpinned by high-performance and specialised IT infrastructure will enable use of the latest developments in technologies like artificial intelligence (AI) offerings.

https://www.monash.edu/__data/assets/pdf_file/0008/1498337/Monash_Uni_Helix_Flyers.PDF
UKSeRP NRDA
(Data custodian portal)

Reference: SAIL Databank + UKSeRP presentation by Simon Thompson, Swansea University
UKSeRP - high level representation

Reference: SAIL Databank + UKSeRP presentation by Simon Thompson, Swansea University
Monash SeRP - Pilot 2018

- Setup on Research Cloud & uses Research Data Storage
- Leverage existing softwares and licenses
- Multi Factor authentication
- Security policies and controls in place
- Includes Data anonymization, data linking, data quality analysis along with simple user interfaces for data custodians and users
- Range of analytical tools and capabilities including NLP and Machine learning
Governance

- Researcher led implementation
- Driven by reference group with researchers across different schools
- Determines the technology and services to be made available on Monash SeRP
- Extendible to other disciplines
- Policies defines the access to data (projects), capabilities and administration
- Quality assurance framework (ISO9001) for management and operations

Data protection by design
Future landscape

Innovative collaborations

Information Security framework (ISO 27001)

Policies and Procedures

Extending to different disciplines

- Humanities
- Legal
- Business & Economics
- Health
- Policy & Governance

Data linking at scale

Potential for future federation for National Safe Haven infrastructure
Thank You

If you have any questions, please reach out to us on

safehaven-support@monash.edu