The Australian National Medical Research Data Storage Facility

med.data.edu.au

Jeff Christiansen, Mohammad Islam, Derek Van Dyk, Leonie Hellmers and Ian Gibson
med.data.edu.au

- What is it?
- What are the drivers?
- Where did it come from?
- Who has signed up?
- Data Types
- Why does it matter?
- What will it do and how will it work?
- What’s its status?
What is it?
What is it?

• a National Facility to provide:
  • Highly secure petabyte-scale data storage
  • Related high-speed networked and secure computational services

for Australian medical and health research organisations
What are the drivers?
What are the drivers?

1. Strategic reviews have pointed to an increased requirement to consolidate, aggregate and collaborate to maximise research outcomes.
What are the drivers?
What are the drivers?
What are the drivers?

<table>
<thead>
<tr>
<th>Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Build HMR Capability</strong></td>
</tr>
<tr>
<td>Enhance commercialisation environment (17)</td>
</tr>
<tr>
<td>Foster a culture of commercialisation</td>
</tr>
<tr>
<td>Leverage scale and expertise</td>
</tr>
<tr>
<td><strong>Accelerate Translation</strong></td>
</tr>
<tr>
<td>Enhance commercialisation environment (17)</td>
</tr>
<tr>
<td>Attract clinical trials investment from overseas</td>
</tr>
<tr>
<td><strong>Optimise Investment</strong></td>
</tr>
<tr>
<td>Support research commercialisation (16)</td>
</tr>
<tr>
<td>Matching development grants</td>
</tr>
<tr>
<td>Translational Biotech Fund</td>
</tr>
<tr>
<td><strong>Deliver Outcomes</strong></td>
</tr>
<tr>
<td>Increase longevity and quality of life</td>
</tr>
<tr>
<td>Boost national wealth</td>
</tr>
<tr>
<td>Drive shift to knowledge-based jobs</td>
</tr>
<tr>
<td>Enhance international standing and engagement with Asia</td>
</tr>
</tbody>
</table>

**Vision**

- A Healthy and Wealthy Australia
- The World's Best Health System
- A World-Class HMR Sector

**Build HMR Capability**

- Build health professional research capacity (4)
- Enhance public health research (12)
- Enhance health services research (13)
- Support a range of strategic topics (7)
- Maintain research excellence in discovery and applied research
  - HMR workforce (8)
  - Grant processes (9)

**Accelerate Translation**

- Establish Integrated Health Research Centres (3)
- Accelerate clinical trial reforms (5)
- Drive health system innovation (14)
- Inform policy with evidence (15)

**Optimise Investment**

- Drive research activity in the health system (1)
- Align priority-setting processes (6)
- Attract philanthropy (18)
- Identify new funding sources (19)
- Invest for the future (20)
- Action report recommendations (21)

**Deliver Outcomes**

- Build and maintain the world's best health system
- Deliver evidence-based healthcare and policy through research

- Leverage and extend reforms
- Maintain world-class research
- Focus on translation and impact
- Monitor investment and outcomes

---

- Indirect cost support (10)
- Enabling infrastructure (11)
What are the drivers?

2. Fast growth of the volume of medical research data
What are the drivers?

2. Fast growth of the volume of medical research data – e.g. imaging
What are the drivers?

2. Fast growth of the volume of medical research data – e.g. genomic data

Guy Cochrane, EBI
What are the drivers?

2. Fast growth of the volume of medical research data

Illumina HiSeq X\textsubscript{TEN}

~USD 1000 per genome

0.6 terabases of sequence per day

3-4 terabytes of data per day

10 machines @ the Kinghorn Centre for Clinical Genomics (Garvan Institute)
What are the drivers?

3. The rise of personalised medicine
What are the drivers?

3. The rise of personalised medicine
Where did it come from?
Where did it come from?

- Funded through **RDSI, RDS and ANDS** projects

- All part of the **NCRIS** program
  - Over $1 billion over past few years
  - $150M planned for 2015-16
  - Longer term planning underway
Where did it come from?

- **RDSI**: established 8 large data storage facilities nationally
Where did it come from?

• **RDSI**: established 8 large data storage facilities nationally
  • Intended to store research data of *value for future research*
  • Total storage ~50 petabytes

• Large Collections Program
  • *Medical*, Imaging, Astronomy, Ecology, Genomics
Where did it come from?

• **RDS**: making RDSI storage sustainable.
Where did it come from?

- **RDS**: making RDSI storage sustainable.
  - optimising and maturing the RDSI capability for 9 data-intensive research communities:
  - integrate rapidly expanding data holdings with the cloud and supercomputing infrastructure to provide a truly high-performance data using capability
Where did it come from?

RDSI/RDS
Medical
Where did it come from?

• **ANDS**: ensuring research data is:
  
  • Managed
  • Connected
  • Findable
  • Reusable
Where did it come from?

• **ANDS**: ensuring research data is:
  • Managed
  • Connected
  • Findable
  • Reusable

• Data Management for all users of RDSI
  • Incl. independent Medical Research Institutes
Who has signed up?
Who has signed up?

• 31 letters of support including 20 Medical Research Institutes
Who has signed up?

The University of Adelaide
The University of Sydney
Macquarie University
The University of Newcastle
UNSW
The University of Queensland

QUT
Australian Catholic University
Mater
Women's and Children's Hospital

Flinders University

NSW Government Cancer Institute
GARVAN Institute

Breast Cancer Tissue Bank

Peter Mac

Black Dog Institute

Centenary Institute for Cancer

The Institute of Melbourne

Muir & Gray

Mi phi

MIMR-PHI Institute of Medical Research

Rapid Autopsy in Melanoma Consortium
Melbourne Genomics Health Alliance

SAHMRI

North Australian Health & Medical Research Institute

Children's Cancer Institute

Children's Medical Research Institute

Neuroscience Research Australia

Ludwig Institute for Cancer Research

CHEBA
Centre for Eye Research

Victor Chang
Cardiac Research Institute

Bones Institute

Mim-RPhi Institute

The Nold Laboratory

Melbourne Brain Centre Imaging Unit
Melbourne femur collection
Health and Biomedical Informatics Centre

Mim-RPhi Institute

Westmead Millennium Institute

Murdoch Childrens Research Institute

The Institute

53 organisations
Data types

• Imaging
  • MRI, CT, PET, Xray, histology…

• Clinical
  • Surveys, pathology, video assessments, patient data, longitudinal studies, clinical trials, eHealth, EEG, anatomy/morphology, personalised medicine, drug discovery/therapeutics…

• “Omics”
  • Genomics, Transcriptomics, Proteomics…

• Biobanking

• Computational modelling

• Health economics
Why does it matter?
Why does it matter?

- Compliance and data security/privacy
- Efficiency and cost effectiveness
- Supporting translational impact
- Enabling new research methods and outcomes
What will it do?
What will it do?

- Store data – securely, including identifiable data
- Describe data – at collection and item level
- Find data – at collection level
- Share data – under appropriate conditions
- Use data – via a number of associated tools (TBD) – including secure options

A national facility to provide petabyte-scale research data storage, and related high-speed networked computational services, to Australian medical and health research organisations.
What will it do?

• Store data
  – securely, including identifiable (sensitive) data

• Describe data
  – at collection and item level

• Share data
  – only under appropriate conditions

• Find data
  – at collection level

• Use data
  – via a number of associated tools (TBD) – to include highly secure options
How will it work?
How will it work?

- Each node is an independent entity
- Storage + (HPC, Cloud)
- Connected via AARNet
- Data custodian enters into data storage agreement with node of their choice – typically the local one
- Data custodians retain control over access
How will it work?

• Governance: Advisory Board (current)
What’s its status?
What’s its status?

• Project team established
• Advisory Board established
What’s its status?

Melbourne Brain Centre Imaging Unit
Melbourne femur collection
Health and Biomedical Informatics Centre
What’s its status?

- 80 foundation data collections identified
- 4 PB
- 80 technical analyses underway
  - 53 organisations across 4 states
  - Security
  - Access control
  - Use cases
What happens next?
What happens next?

- Signing up more organisations
- Establish roadmap of services
  - data, metadata standards, data management and curation, APIs, analytics
What happens next?

• Signing up more organisations
• Establish roadmap of services
  – data, metadata standards, data management and curation, APIs, analytics
• Opportunity to align access methods across consenting collections
• Conducting research across aggregated collections
Thanks