



Navigating data-driven medical researchers through the IT privacy/security landscape

Jeff Christiansen, Jared Winton and Kathy Dallest

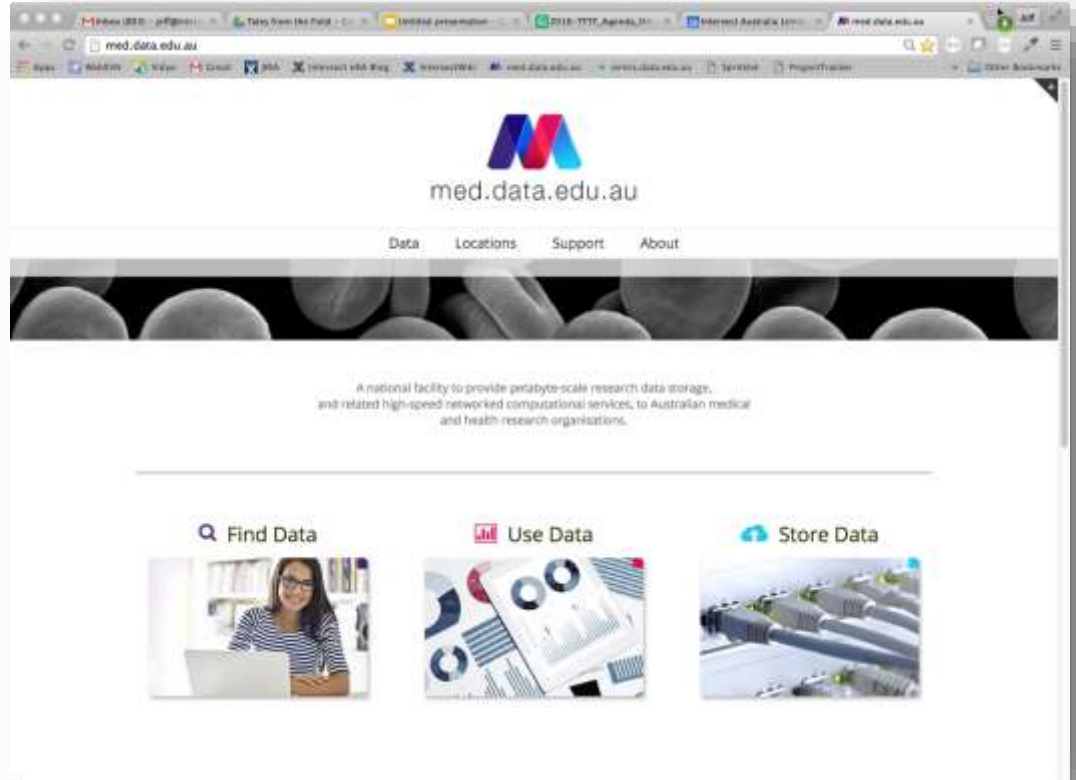


med.data.edu.au

NCRIS-funded project



A national facility for petabyte scale research data storage and high-speed networked computational services to Australian Health and Medical Research organisations



med.data.edu.au

4 RDS nodes:

Intersect, VicNode, eRSA, QCIF

Health/medical research data:

4 states

7 Universities

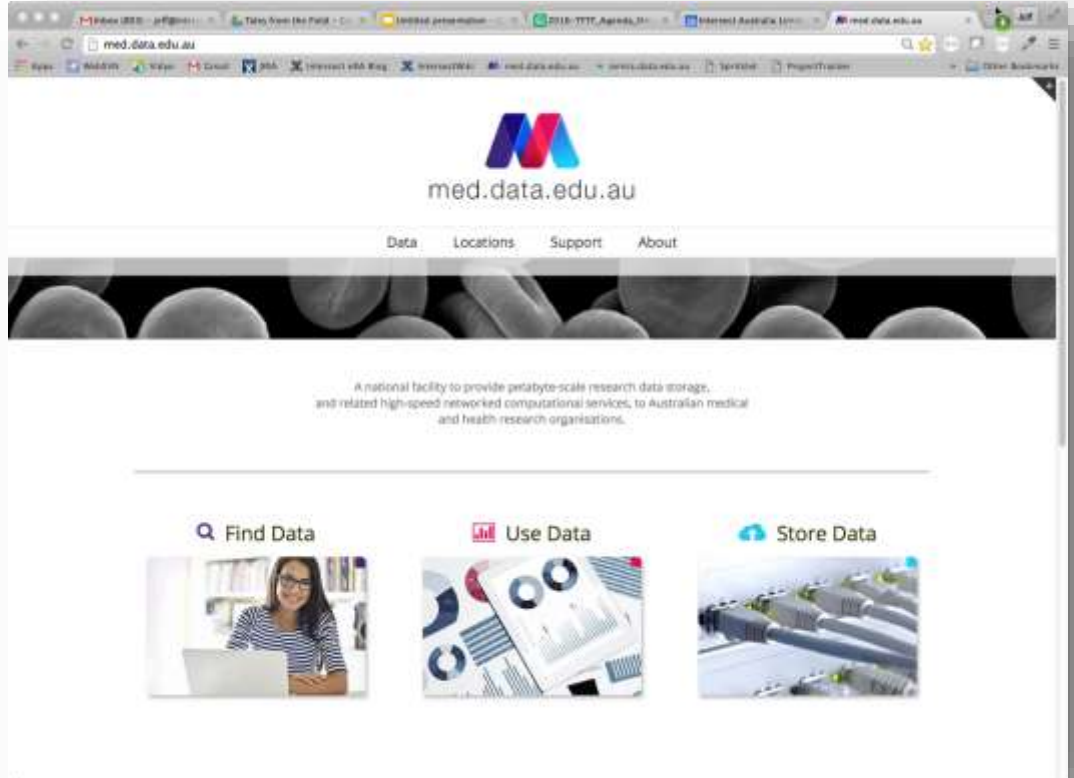
12 Medical Research Institutes

2.2PB (91 data collections)

Human genomic data ~85%

Human 3D imaging ~10%

Other health datasets <5%



The screenshot shows the homepage of the med.data.edu.au website. At the top, there is a navigation bar with the logo (a stylized 'M' in blue and red) and the text 'med.data.edu.au'. Below the logo, there are four menu items: 'Data', 'Locations', 'Support', and 'About'. The main content area features a large image of several grey, pill-like objects. Below this image, there is a text block that reads: 'A national facility to provide petabyte-scale research data storage, and related high-speed networked computational services, to Australian medical and health research organisations.' At the bottom, there are three columns of content, each with an icon and a title: 'Find Data' (magnifying glass icon), 'Use Data' (document icon), and 'Store Data' (cloud icon). Each column also includes a small image: a woman working on a laptop for 'Find Data', a document with charts for 'Use Data', and a server rack for 'Store Data'.

med.data.edu.au

Advisory Board:

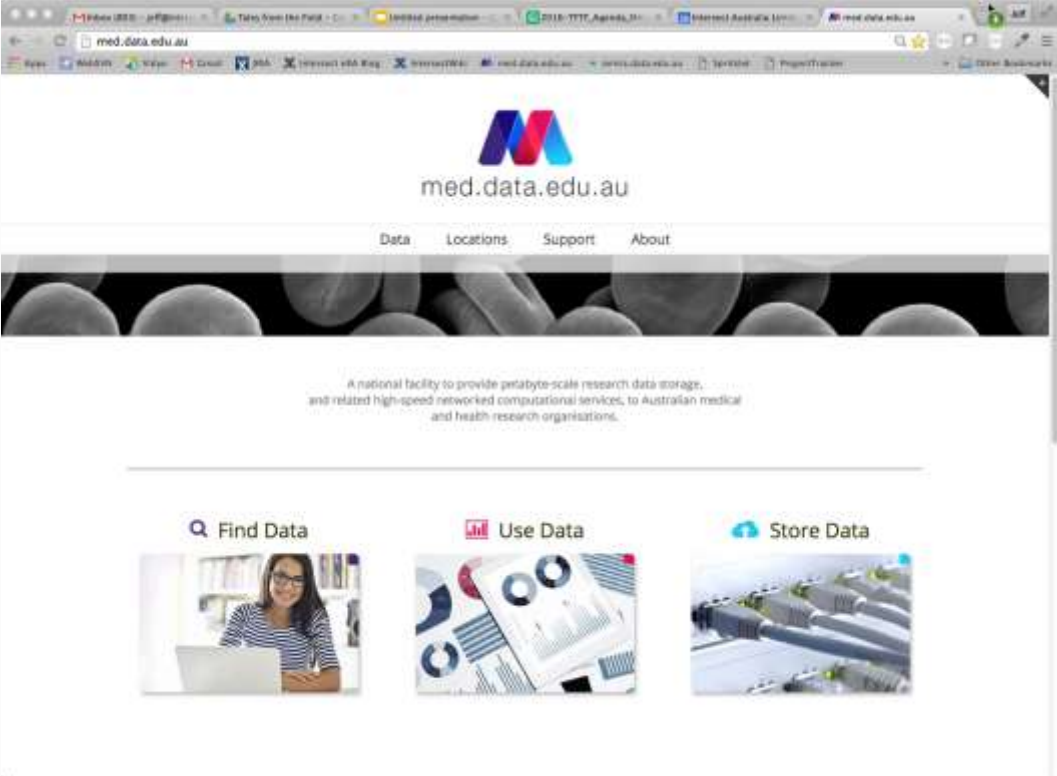


med.data.edu.au

FIND DATA

USE DATA

STORE DATA



The screenshot shows the homepage of the med.data.edu.au website. At the top, there is a navigation bar with the logo (a stylized 'M' in blue and red) and the text 'med.data.edu.au'. Below the logo is a horizontal menu with the items 'Data', 'Locations', 'Support', and 'About'. A large banner image of pills is positioned below the menu. Underneath the banner, a paragraph of text reads: 'A national facility to provide petabyte-scale research data storage, and related high-speed networked computational services, to Australian medical and health research organisations.' Below this text is a horizontal line, and then three main service tiles are displayed. The first tile is titled 'Find Data' with a magnifying glass icon and a photo of a woman at a laptop. The second tile is titled 'Use Data' with a bar chart icon and a photo of data charts. The third tile is titled 'Store Data' with a server rack icon and a photo of server cables.

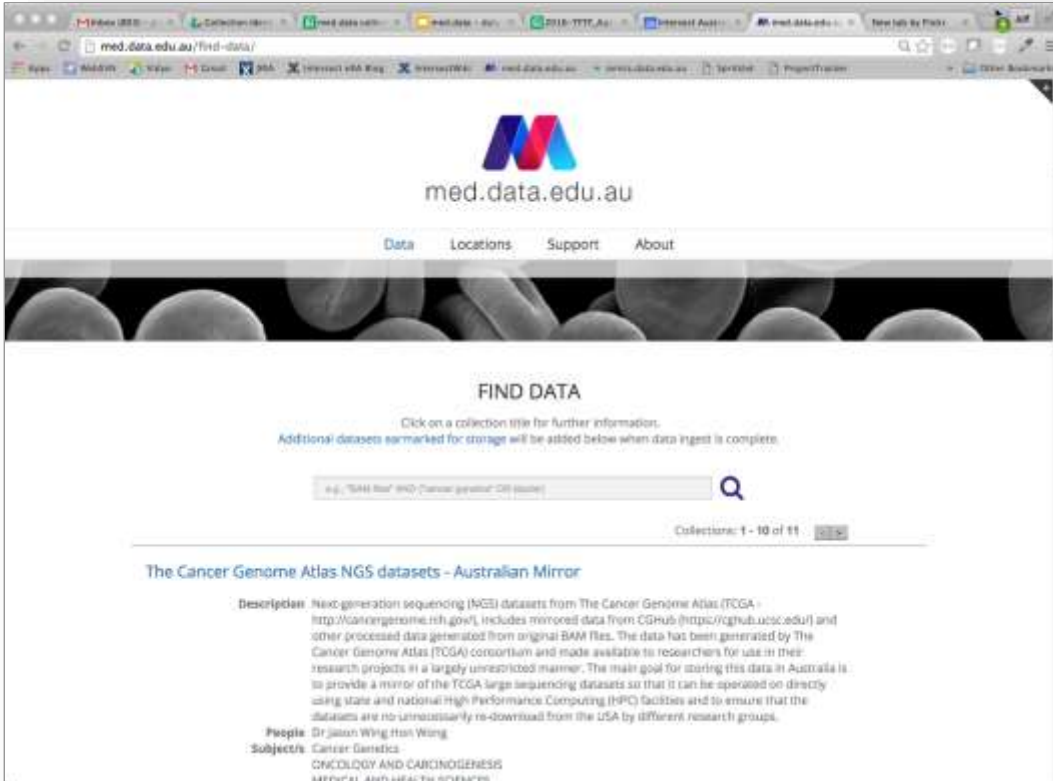
med.data.edu.au

FIND DATA

Dataset descriptions listed

Tools and processes to publish dataset descriptions

Leverages institutional and national infrastructure (ANDS)



The screenshot shows a web browser window displaying the med.data.edu.au website. The page features a navigation menu with links for 'Data', 'Locations', 'Support', and 'About'. Below the navigation is a search bar with the text 'FIND DATA' and a search icon. A search result is displayed for 'The Cancer Genome Atlas NGS datasets - Australian Mirror'. The description of the dataset is as follows:

Description: Next-generation sequencing (NGS) datasets from The Cancer Genome Atlas (TCGA - <http://cancergenome.nih.gov/>), includes mirrored data from CGHub (<http://cghub.usc.edu/>) and other processed data generated from original BAM files. The data has been generated by The Cancer Genome Atlas (TCGA) consortium and made available to researchers for use in their research projects in a largely unrestricted manner. The main goal for storing this data in Australia is to provide a mirror of the TCGA large sequencing datasets so that it can be operated on directly using state and national High Performance Computing (HPC) facilities and to ensure that the datasets are no unnecessarily re-downloaded from the USA by different research groups.

People: Dr Jason Wing Hon Wong

Subjects: Cancer Genetics
ONCOLOGY AND CARCINOGENESIS
MEDICAL AND HEALTH SCIENCES

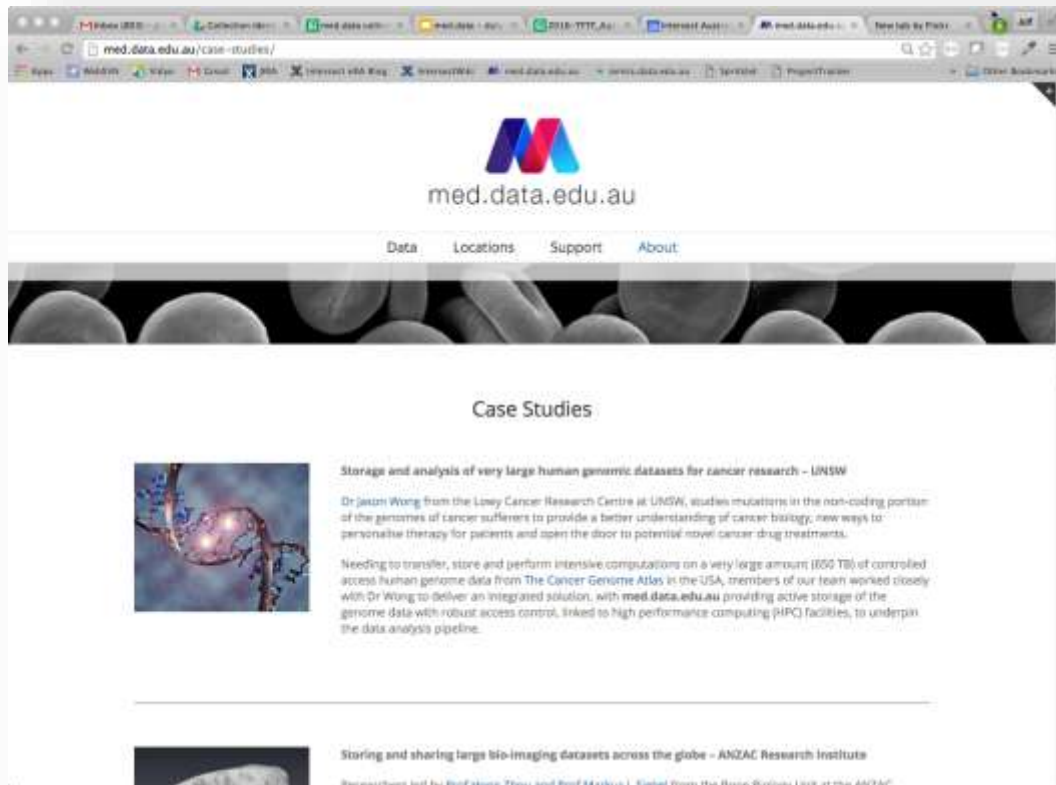
med.data.edu.au

USE DATA

Generally, users requiring large data storage and associated HPC or cloud (Nectar) compute, or

Collaborative research across Institutions

BYO software



The screenshot shows the website med.data.edu.au. The browser address bar displays the URL. The website features a logo with a stylized 'M' in blue and red, followed by the text 'med.data.edu.au'. Below the logo is a navigation menu with links for 'Data', 'Locations', 'Support', and 'About'. A decorative banner image of pills is positioned below the navigation. The main content area is titled 'Case Studies' and lists two examples:

- Storage and analysis of very large human genomic datasets for cancer research - UNSW**
Dr Jaemin Wong from the Linsay Cancer Research Centre at UNSW, studies mutations in the non-coding portion of the genomes of cancer sufferers to provide a better understanding of cancer biology, new ways to personalise therapy for patients and open the door to potential novel cancer drug treatments.
Needing to transfer, store and perform intensive computations on a very large amount (100 TB) of controlled access human genome data from The Cancer Genome Atlas in the USA, members of our team worked closely with Dr Wong to deliver an integrated solution, with med.data.edu.au providing active storage of the genome data with robust access control, linked to high performance computing (HPC) facilities, to underpin the data analysis pipeline.
- Storing and sharing large bio-imaging datasets across the globe - ANZAC Research Institutes**
Researchers led by Prof Steve Zhou and Prof Markus J. Sieber from the Bruce Biotech Unit at the ANZAC

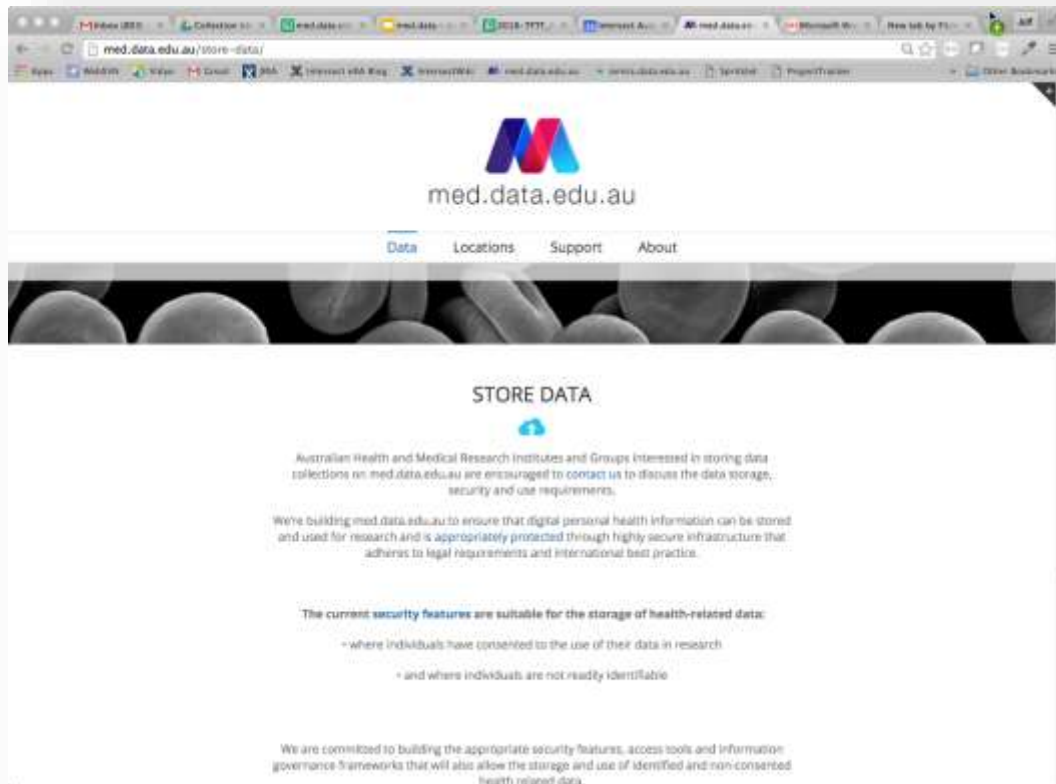
med.data.edu.au

STORE DATA

Data is managed using various platforms

(Mediaflux, Aspera, MyTardis)

Human-derived data is subject to a higher level of protection than other data



The screenshot shows a web browser displaying the 'med.data.edu.au' website. The page features a logo with a stylized 'M' in blue and red, followed by the text 'med.data.edu.au'. Below the logo is a navigation menu with links for 'Data', 'Locations', 'Support', and 'About'. The main content area has a dark banner with a microscopic view of cells. Below this, the heading 'STORE DATA' is displayed with a blue icon of a person. The text explains that Australian Health and Medical Research Institutes and Groups are encouraged to contact the site to discuss data storage, security, and use requirements. It states that the site is built to ensure digital personal health information can be stored and used for research, protected through highly secure infrastructure that adheres to legal requirements and international best practice. A section titled 'The current security features are suitable for the storage of health-related data:' lists two conditions: 'where individuals have consented to the use of their data in research' and 'and where individuals are not readily identifiable'. At the bottom, a commitment is made to building appropriate security features, access tools, and information-governance frameworks for the storage and use of identified and non-consented health-related data.

Health and Medical information



Collected for healthcare provision

Held by health service providers

Derived from or relating to human individuals

It is personal* info: i.e. about an identified or reasonably identifiable individual

It is sensitive* info: i.e. about an individual's Health; or Genetic or Biometric info.

* As defined by The Privacy Act 1988 (Cth)

Health and Medical information



Collected for healthcare provision

Held by health service providers

Derived from or relating to human individuals

It is personal* info: i.e. about an identified or reasonably identifiable individual

It is sensitive* info: i.e. about an individual's Health; or Genetic or Biometric info.

Must be protected (Privacy Act (Cth) and various state health privacy legislation)

Health and Medical information in research



Collected for healthcare provision OR specifically collected for use in research studies

Held by researchers

Derived from or relating to human individuals (i.e. is personal and sensitive)

Health and Medical information in research



Collected for healthcare provision OR specifically collected for use in research studies

Held by researchers

Derived from or relating to human individuals (i.e. is personal and sensitive)

**Must be protected from misuse, interference and loss,
and from unauthorised access, modification or disclosure.**

Health and Medical information in research



POINTS FOR CONSIDERATION

Privacy Protection Legislation

Privacy Protection Legislation

Protecting an individual's privacy (and health privacy) is enshrined in Commonwealth and State Legislation:

Commonwealth

- Privacy Act (1988)

NSW

- Privacy and Personal information Protection Act 1998
- Health Records and Information Privacy Act 2002 (HRIPA)

VIC

- Information Privacy Act 2000 (Vic)
- Privacy and Data Protection Act 2014
- Health Records Act 2001 (Vic)
- Charter of Human Rights and Responsibilities Act 2006 (Vic)

QLD

- Information Privacy Act 2009 (Qld)
- Health Services Act 1991 (Qld)
- Information Standards 42 (general) & 42A (health)
- Public Health Act 2005 Chapter 6, Part 4, Division 2, s281

SA

- Department of the Premier and Cabinet IPPs

WA

- Information Privacy Bill 2007

TAS

- No health specific privacy legislation.
- Personal Information and Protection Act 2004

ACT

- Information Privacy Act 2014 (ACT) (ACT Public Sector Agencies)
- Health Records (Privacy and Access) Act 1997

NT

- No health specific privacy legislation.

Ethics

Ethics

When health data is collected for research purposes, Human Research Ethics Committee (HREC) approval is required first.

HREC approval is also required for its use (usually only approved for use in a specific research study).

Must not be shared in an identifiable form with those outside the HREC-approved research project(s).

Consent

Consent

The guiding principle for researchers is that a person's decision to participate in research is to be voluntary, and based on sufficient information and adequate understanding of both the proposed research and the implications of participation in it.

“Informed consent”

As defined in the National Statement on Ethical Conduct in Human Research (NHMRC, ARC, Universities Australia)

Consent

When consent has not been given to use personal health data in research, the use of health data use may still be permitted for research purposes:

Cth Privacy Act Section 95 – allows Commonwealth managed health data to be used for research in an identifiable form if the proposed research has been approved by a HREC.

Cth Privacy Act Section 95A – allows private sector data to be used for research where gaining consent is not practical, and the research has been approved by a HREC.

Note – neither are applicable for data from State Health Departments - however some states e.g. NSW have comparable legislation (e.g. NSW – HRIPA)

Identifiability of data

Identifiability of data

Individually identifiable data – where the identity of a specific individual can *reasonably be ascertained* (e.g. a name, image, date of birth or address).

Re-identifiable data - where identifiers have been removed and replaced by a code, but it remains possible to re-identify a specific individual by, for example, using the code or linking different data sets.

Non-identifiable data - has never been labelled with individual identifiers or from which identifiers have been permanently removed, and by means of which no specific individual can be identified.

Identifiability of data

Making data non-identifiable, decreases risks associated with inadvertent release

Removal of overt identifying information or identifiers - e.g. name, image, date of birth, address, medicare/patient numbers.

Statistical methods - applied by an expert and the methods must be documented

Legislation affecting trans-border dataflow

Legislation affecting trans-border dataflow

Commonwealth – Privacy Act (1988) Australian Privacy Principle (APP)-8 cross-jurisdictional transfer of personal information out of Australia.

NSW – NSW Health Records and Information Privacy Act 2002 Health Privacy Principle (HPP)-14 Trans-border data flows and data flow to Commonwealth agencies.

VIC – Health Records Act 2001 (Vic) Health Privacy Principle (HPP)-9 Transborder Data Flows

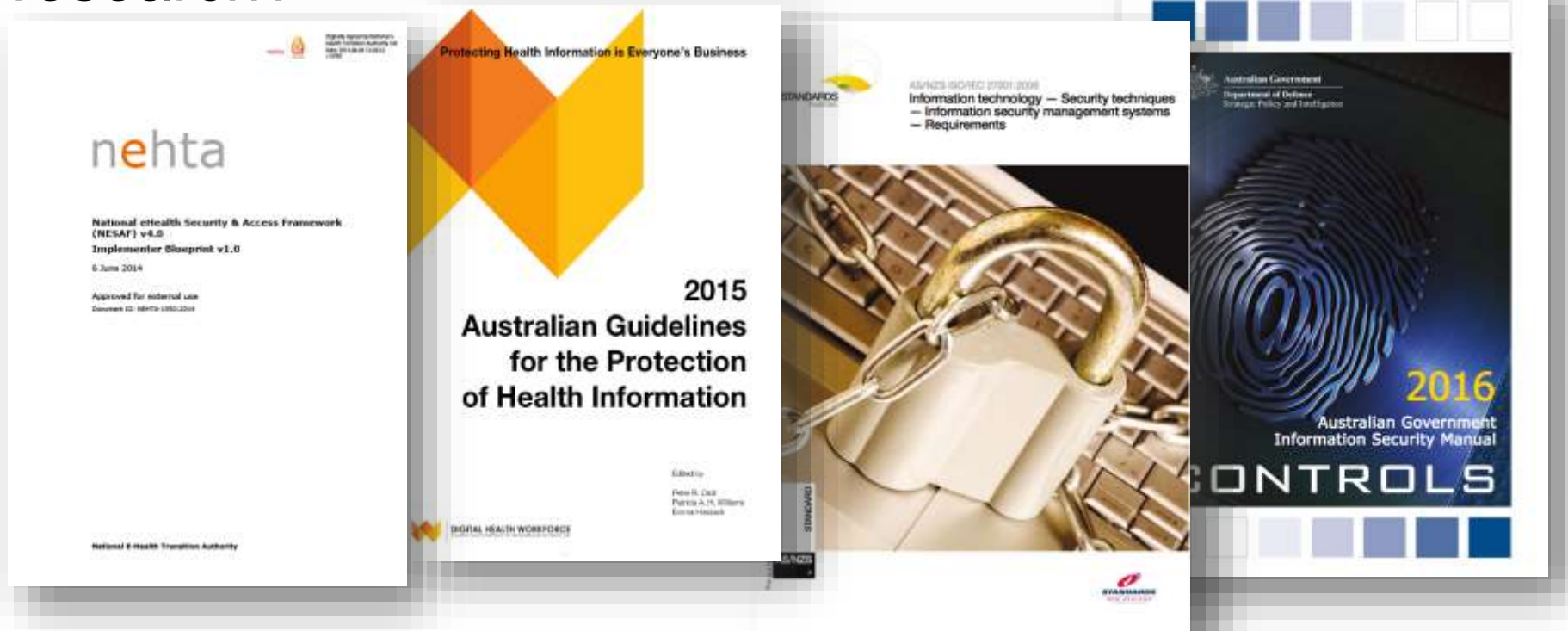
WA – Information Privacy Bill 2007 (WA) Information Privacy Principle 8: Transborder data flows

TAS – Personal Information Protection Act 2004

NT – Information Act: Information Privacy Principle (IPP)-9 Transborder data flows

What controls are needed to protect HM data in research?

What controls are needed to protect HM data in research?



What controls are needed to protect HM data in research?



There IS no definitive checklist - It's about risk management

Discussion paper

Legislative framework
(Commonwealth, State
and International)

Best Practice (Ethics,
Research etc)

IT Security requirements for
Health (Human-derived) data

Roles and Responsibilities
(Data Custodians, Users, Nodes)



Discussion paper

41 pages and heavy going

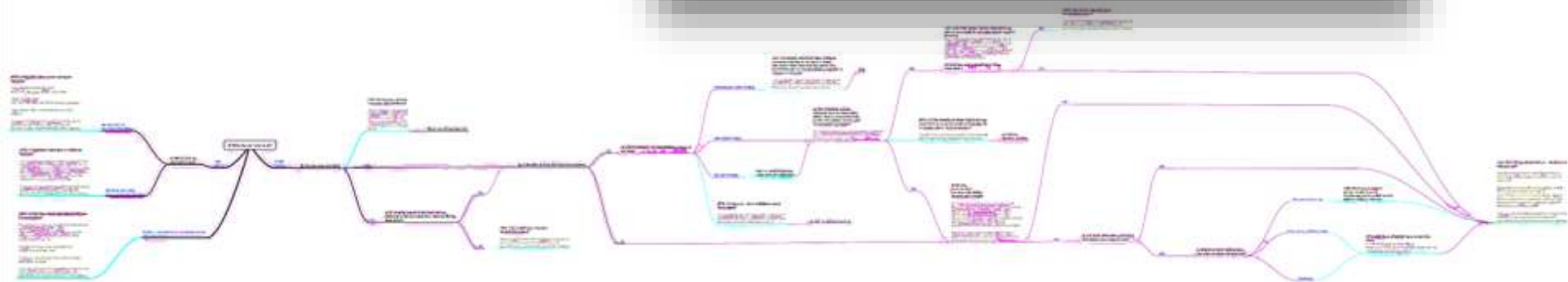
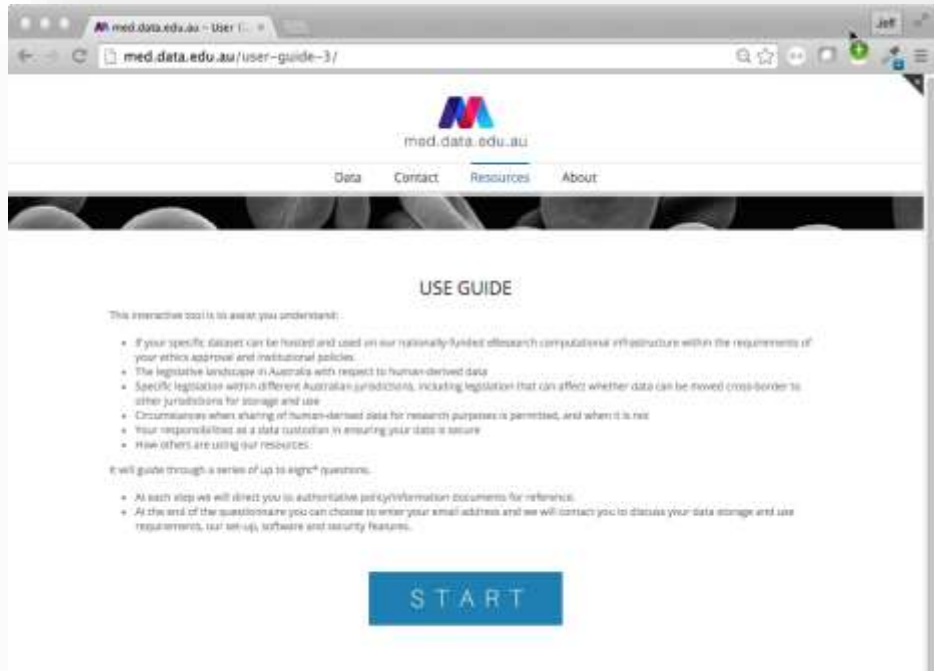
Need something simpler



Use Guide

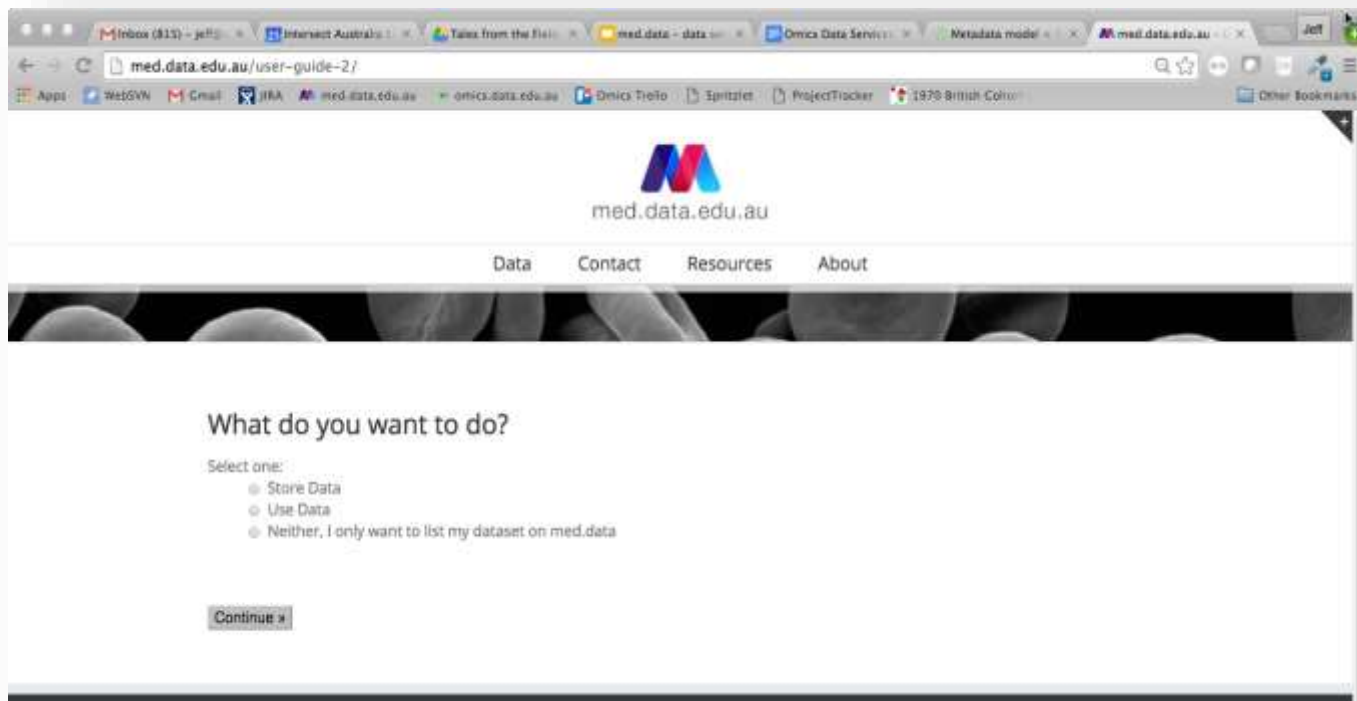
On-line wizard

Navigates the user through relevant info



Use Guide

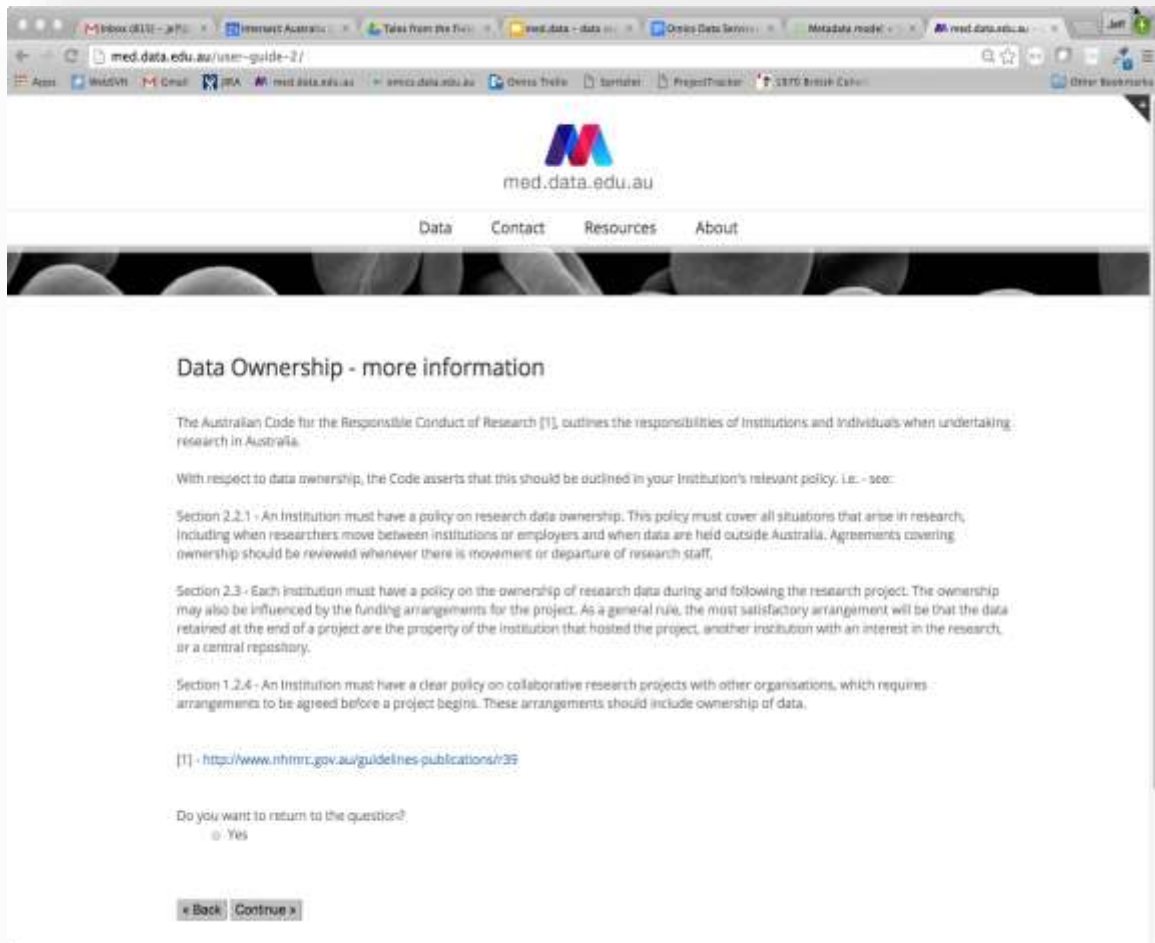
Simple questions



Use Guide

Info when requested

Authoritative references



The screenshot shows a web browser window with the URL med.data.edu.au/user-guide-2/. The page features the med.data.edu.au logo and a navigation menu with links for Data, Contact, Resources, and About. The main content area is titled "Data Ownership - more information" and contains the following text:

The Australian Code for the Responsible Conduct of Research [1], outlines the responsibilities of Institutions and Individuals when undertaking research in Australia.

With respect to data ownership, the Code asserts that this should be outlined in your Institution's relevant policy, i.e. - see:

Section 2.2.1 - An Institution must have a policy on research data ownership. This policy must cover all situations that arise in research, including when researchers move between institutions or employers and when data are held outside Australia. Agreements covering ownership should be reviewed whenever there is movement or departure of research staff.

Section 2.3 - Each institution must have a policy on the ownership of research data during and following the research project. The ownership may also be influenced by the funding arrangements for the project. As a general rule, the most satisfactory arrangement will be that the data retained at the end of a project are the property of the institution that hosted the project, another institution with an interest in the research, or a central repository.

Section 1.2.4 - An Institution must have a clear policy on collaborative research projects with other organisations, which requires arrangements to be agreed before a project begins. These arrangements should include ownership of data.

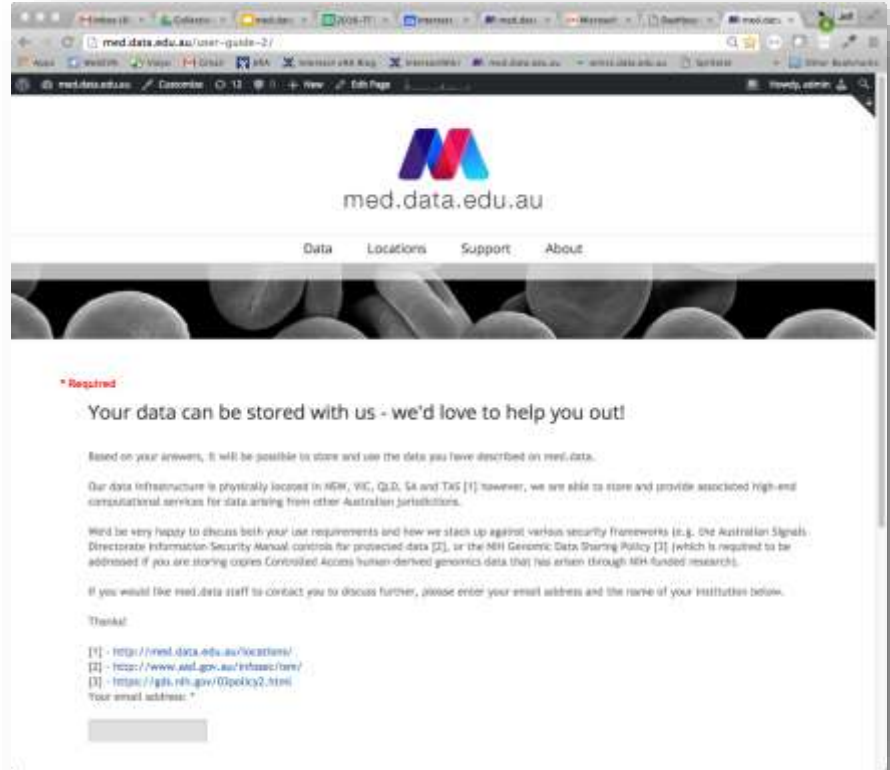
[1] - <http://www.nhmrc.gov.au/guidelines-publications/35>

Do you want to return to the question?

Yes

Use Guide

Option to be contacted for
Node Specific Information



The screenshot shows the homepage of med.data.edu.au. At the top, there is a navigation menu with links for 'Data', 'Locations', 'Support', and 'About'. Below the menu is a banner image of several pills. The main content area features a red heading '* Required' followed by the text 'Your data can be stored with us - we'd love to help you out!'. Below this, there are several paragraphs of text explaining the data storage capabilities and security measures. At the bottom, there is a form for entering an email address and the name of the institution.

*** Required**

Your data can be stored with us - we'd love to help you out!

Based on your answers, it will be possible to store and use the data you have described on med.data.

Our data infrastructure is physically located in NSW, VIC, Q.L.D. SA and TAS [1] however, we are able to store and provide associated high-end computational services for data arising from other Australian jurisdictions.

We'd be very happy to discuss both your use requirements and how we stack up against various security frameworks (e.g. the Australian Signals Directorate Information Security Manual controls for protected data [2], or the NH Genomic Data Sharing Policy [3] (which is required to be addressed if you are storing copies Controlled Access human-derived genomics data that has arisen through NH funded research).

If you would like med.data staff to contact you to discuss further, please enter your email address and the name of your institution below.

Thanks!

[1] - <http://med.data.edu.au/locations/>
[2] - <https://www.asd.gov.au/infocore/infocore/>
[3] - <https://gls.nh.gov/01policy2.html>

Your email address: *

End



Use Guide

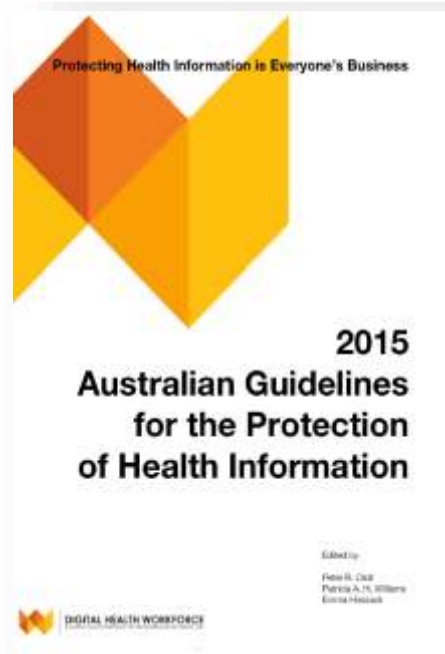
Node Specific Security Info

How do each Node's Storage, HPC and Cloud compute stack up against...

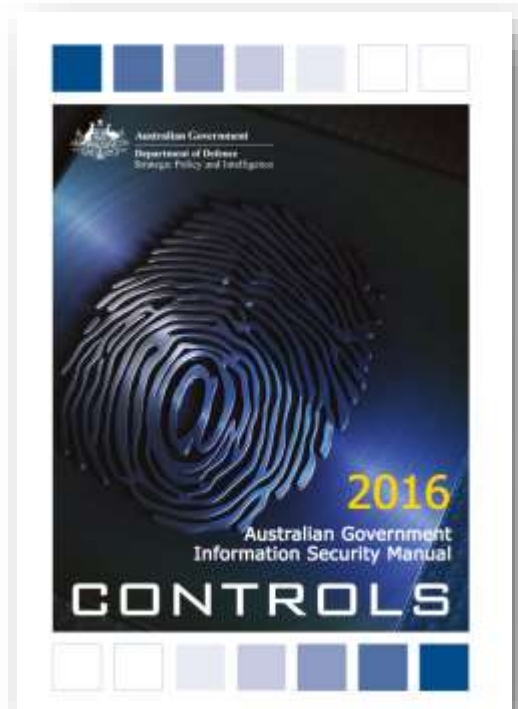
- National Standards for Protected or Sensitive Info (ASD ISM)?

What controls are needed to protect HM data in research?

- User identity
- User authentication
- Access control
- Encryption
- Secure audit
- General security
- Anonymisation



- Controls for
“Protected” and
“Sensitive”
Data
- IT Security set-up
- Security Policy



Use Guide

Node Specific Security Info

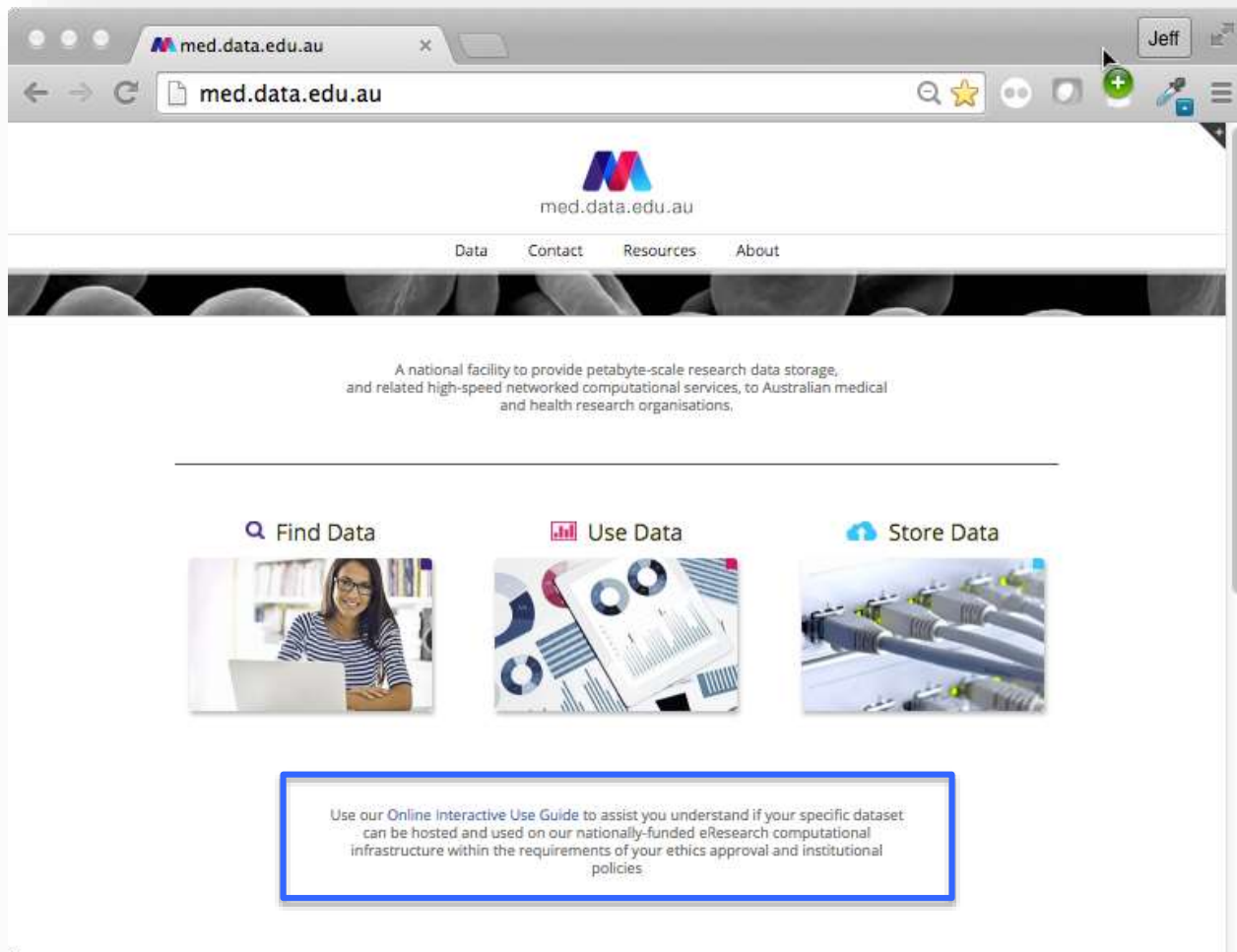
How do each Node's Storage, HPC and Cloud compute stack up against...

- National Standards for Protected or Sensitive Info (ASD ISM)?
- Common other standards (e.g. NIH Best Practices for storage of Human Genomic Sequence)?

Required for an informed conversation about risk with each Data Custodian (Node set-up, Roles and Responsibilities of Data Custodians, Data Users and Node Operators).

Use Guide

Available Now



The screenshot shows a web browser window with the URL med.data.edu.au. The page features a logo with a stylized 'M' in blue and red, followed by the text 'med.data.edu.au'. A navigation menu includes 'Data', 'Contact', 'Resources', and 'About'. Below the menu is a banner image of several grey spheres. The main content area contains a paragraph: 'A national facility to provide petabyte-scale research data storage, and related high-speed networked computational services, to Australian medical and health research organisations.' Below this are three columns: 'Find Data' with a magnifying glass icon and a photo of a woman at a laptop; 'Use Data' with a bar chart icon and a photo of data charts; and 'Store Data' with a cloud icon and a photo of network cables. At the bottom, a blue-bordered box contains the text: 'Use our Online Interactive Use Guide to assist you understand if your specific dataset can be hosted and used on our nationally-funded eResearch computational infrastructure within the requirements of your ethics approval and institutional policies.'

THANK YOU
jeff@intersect.org.au

Acknowledgements



NCRIS
National Research
Infrastructure for Australia

An Australian Government Initiative