med.data.edu.au

DISCUSSION PAPER

Legal, Best Practice and Security Frameworks for consideration in operation of the Australian National Medical Research Data Storage Facility.

Jan 2016

1. INTRODUCTION

Seed funding from the Commonwealth Department of Education through the National Collaborative Research Infrastructure Strategy (NCRIS)[1] and via the Research Data Storage Initiative (RDSI)[2] and Research Data Services (RDS)[3] projects has been received to establish a National Medical Research Data Storage Facility (med.data.edu.au or "med.data" for short).

med.data is intended to function as a national petabyte-scale data storage facility that can store Health and Medical Research data on behalf of a wide range of the Australian Research Community (i.e. from Universities, Hospitals and Medical Research Institutes).

Infrastructure and services that underpin the med.data "brand" will be operated and maintained by several independent organisations (or 'nodes') around the country[4], each of which have different computational infrastructure, governance and processes. Therefore when mature, it is envisaged that the med.data brand will point to a variety of different infrastructures and services at these various nodes. These will enable different Data Custodian/Stewards around the country to store their data securely, and provide services for that data to be (a) well managed, structured and described, (b) accessible to state of the art computational facilities (e.g. HPC and cloud systems) through a secure high-speed network, and (c) allow the Data Custodian/Stewards to share specific datasets or items in a seamless manner with others when appropriate.

As the primary focus of the facility is Health and Medical Research data, it is envisaged that a proportion of the data stored will include personal or sensitive information that is derived from human participants. As such, med.data as a whole needs a well-defined data governance framework that can provide clear guidelines to each node operator as to their responsibilities, as well as assurance to Data Custodian/Stewards that data will be held and managed in a manner that is aligned with security standards that are required by current laws. med.data will also need to provide tools and services that allow Data Custodian/Stewards to use the dispersed facility in alignment with best practice in the use of human participant data in research.

The overall aim of this discussion paper is to describe relevant Legislation, Codes, Policies and Standards that will impact on the operation of med.data as a whole, and to describe a shared model of responsibility, where Data Custodian/Stewards and the operators of med.data have clearly defined roles and responsibilities to ensure that data (including personal information) held on the facility is stored and used with a high degree of confidence with regards to security.

This Discussion Paper is currently focussed on NSW and Victoria and the policies, legislation and standards listed in this paper are not necessarily comprehensive. As such, this document will be reviewed and extended over time.

[1] https://education.gov.au/national-collaborative-research-infrastructure-strategy-ncris
[2] https://www.rdsi.edu.au/
[3] https://www.rds.edu.au/
[4] The node operators Intersect, VicNode, QCIF, eResearch SA and TPAC are based in are based in NSW, VIC, QLD, SA and TAS respectively.

2. FUNCTIONALITY

med.data is intended to allow Data Custodian/Stewards to Store, Describe, Share and Use their data. Any data sharing will be determined and controlled by the Data Custodian/Steward.

The full functionality of med.data will be informed by ongoing analysis with stakeholders to entirely understand the needs and requirements of the diverse health and medical research sector, however initially, med.data proposed to have four broad primary functions: to "Store", "Describe", "Share" and "Use" data.

2.1 STORE Data

med.data will provide managed data storage, and as a facility will need to be compliant with the ethical, legal and technical standards required to hold sensitive information, as it is anticipated that clinically derived health information will be held as part of the research data holdings. This information may be duplicated from health records or health information collected from individuals as part of a research project.

2.2 DESCRIBE Data

med.data will have a publicly facing catalogue of high level descriptions of the data sets stored at each node. Descriptions of the data holdings will be created by the Data Custodian/Steward and published through a single agreed mechanism provided by the Node Operators. Post-publishing, 3rd party researchers will be able to find datasets through various mechanisms including a web search engine.

2.3 SHARE Data

A significant proportion of funding for the med.data data storage hardware at each node has been provided via the NCRIS RDSI project. In order to comply with RDSI guidelines to attract a subsidy for data storage costs, some level of data sharing is expected for each data collection. This may be with specific collaborators or the general public.

Data sharing is a broad term and in terms of the function of med.data will be determined and controlled by the Data Custodian/Steward. It is not proposed that any Node Operator of med.data will determine who may access data, rather that med.data will provide the technical functionality that enables Data Custodian/Stewards to make their data assets both discoverable and shareable to other researchers or policy makers in the appropriate format (e.g. aggregated data, de-identified data sets or identified datasets for data linkage), and under the appropriate circumstances.

Access to the actual data will only be available under conditions that have been determined by the Data Custodian/Steward as being appropriate to the type of data held e.g. ranging from freely downloadable public data, through to only being available to registered users (e.g. those who have appropriate approval from a Human Research Ethics Committee (HREC)) via a highly defined application mechanism that is controlled either by the Data Custodian/Steward or a trusted 3rd party. The Node Operators will provide the mechanisms to allow data to be shared according to the conditions stipulated by the Data Custodian/Steward.

2.4 USE Data

Med.data will provide networked access to associated data analysis platforms. Node Operators will implement these platforms over the lifetime of the facility based on needs of the user community and it is acknowledged that needs are likely to change over time. Functionality that this paper will consider is the addition of and/or connecting to data analysis and data-linkage tools that will point to data stored in med.data.

3. DATA TYPES TO BE HELD

It is envisaged that for med.data to function in a fit-for-purpose manner to underpin cutting edge research in the Health and Medical research sector, a variety of data types will be held, ranging from: sensitive to non-sensitive, identifiable to non-identifiable, and across a wide range of formats (e.g. genetic sequences, clinically and non-clinically derived images and test results, participant surveys, etc.).

3.1 PERSONAL AND SENSITIVE INFORMATION

med.data will hold a proportion of data that pertains to health, genetic or biometric information about individuals. These data are considered to be sensitive.

The Commonwealth Privacy Act (1988)[5] ("The Privacy Act") defines what is considered personal and sensitive information in Australia. **Personal Information** means information about an identified individual, or an individual who is reasonably identifiable, and of relevance to med.data, **Sensitive Information** includes: Information about an individual's Racial or ethnic origin or Sexual orientation or practices; Health information; Genetic information and Biometric information. Med.data is therefore likely to held a proportion of data that can be considered personal and sensitive.

3.2 IDENTIFIABLE INFORMATION

med.data may hold a proportion of data pertaining to individuals or groups of individuals that will include identification information (e.g. names and addresses) or identifiers (Medicare numbers, Medical Record Numbers), as well as data that is potentially re-identifiable, and data that is non-identifiable. It is recommended that Node Operators agree on appropriate technical storage solutions for each category of data identified (individually identifiable, re-identifiable, and non-identifiable).

The Privacy Act defines what is considered identifying information about an individual.

> **Identification information about an individual** means: the individual's full name; alias; date of birth; sex; current, last known or previous address or employer; or driver's licence number. An **Identifier of an individual** means a number, letter or symbol, or a combination thereof, that is used to identify or verify the identity of the individual, but does not include the individual's name. For instance, an identifier of an individual may include a Medicare Number or Hospital/Medical Record Number. The Act also defines what is meant by de-identified i.e. "personal

[5] http://www.oaic.gov.au/privacy/privacy-act/the-privacy-act

> information is de-identified if the information is no longer about an identifiable individual or an individual who is reasonably identifiable".

The National Health and Medical Research Council (NHMRC), Australian Research Council (ARC) and Australian Vice Chancellors Committee in their National Statement on Ethical Conduct in Human Research[6], define the following terms:

> **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**, which have 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. A subset of non-identifiable data are those that can be linked with other data so it can be known that they are about the same data subject, although the person's identity remains unknown. Note that the National Statement avoids the term **de-identified data** as its meaning is unclear. As the Statement points out, "while it is sometimes used to refer to a record that cannot be linked to an individual ('non-identifiable'), it is also used to refer to a record in which identifying information has been removed but the means still exist to re-identify the individual. When the term 'de-identified data' is used, researchers and those reviewing research need to establish precisely which of these possible meanings is intended".

It is recommended that Node Operators agree on appropriate technical storage solutions for each category of data identified (individually identifiable, re-identifiable and non-identifiable).

3.3 DATA TYPES BY FORMAT

med.data will hold a wide range of health and medical research data spanning 'bench', clinical and community research.

Globally, various types of Health and Medical data, including demographics, clinical and genomic information, are increasingly collected and stored in both Electronic Medical Record (EMR) systems and biomedical research repositories. Such data have traditionally been used in automating healthcare workflows, but have recently been recognised as an invaluable source for performing large-scale and low-cost biological, medical, and healthcare analysis and decision-making. These tasks are essential for the discovery of new drugs and therapies, and are a key step towards realising the vision of personalised medicine.

A broad range of data types, have been identified as initial candidates for storage on med.data, including: demographic; clinical; health economics; and "-omic" types. It is envisaged that the types of data that researchers will wish to hold on med.data will continue be highly varied and be derived from the entire spectrum of the Health and Medical research endeavour:

> **Bench** - Animal research models; Cell/Tissue culture (human or animal); Tissue from biobanks or specific collection for study, animal/cell/tissue culture-derived –omics data.

[6] https://www.nhmrc.gov.au/guidelines-publications/e72

Clinical - Pathology; Clinical notes and history; Questionnaires/Surveys; Data on research specific procedures or parameters; Imaging/photography; Health services research (mix of service and patient data e.g. Length of Stay, infection rates, short and long term follow up); Clinical Trials, human-derived –omics data

Community - Qualitative data; Aggregated statistics from agencies; Evaluation data (impact of public health programs); Health economic data; Health services research (mix of service and patient data).

4. ROLES AND RESPONSIBILITIES

4.1 ROLES

There are three key roles that have responsibilities for data governance within med.data. These are: (a) the Data Sponsor/Owner, (b) the Data Custodian/Steward and (c) the Node Operator.

4.1.1 Data Sponsors (Owners) and Data Custodians (Stewards)

Responsibility for defining how data may be used and/or shared resides firmly with two roles.

The NSW Health Privacy Manual[7] (Section 15.14.3) provides titles and definitions of these roles and their responsibilities: **Data Sponsor** - Each data collection has a nominated data sponsor who undertakes the *duties of ownership* on behalf of the relevant organisation, including: Defining the purpose of the data collection; Establishing the scope and coverage of the collection; and Defining access and custody arrangements. **Data Custodian** - The data sponsor appoints a custodian for each data collection who is responsible for the *day-to-day management of the data collection* i.e.: Data storage and disposal; Compliance of data with relevant legislation and policies administration; Quality assurance; and Data access and release (i.e. authorising disclosure of information to other researchers or for other research purposes).

Other organisations may use alternative terms, however the meanings are similar. For example, the Australian Standard AS ISO 27799-2011 Information Security Management in Health using ISO/IEC 27002[8] refers to an **Asset Owner** rather than a Data Sponsor, but the intention is the same: i.e. "The asset owner should be responsible for: a) ensuring that information and assets associated with information processing facilities are appropriately classified; b) defining and periodically reviewing access restrictions and classifications, taking into account applicable access control policies". The Global Alliance for Genomics and Health[9] define the term **Data Steward** as opposed to Data Custodian – i.e. "an entity responsible for assuring the quality, integrity, and access arrangements of data and metadata in a manner that is consistent with applicable law, institutional policy, and individual permissions".

[7] NSW Health Privacy Manual, Version 2 http://www0.health.nsw.gov.au/policies/pd/2005/pdf/PD2005593.pdf
[8] http://infostore.saiglobal.com/store/details.aspx?ProductID=1461737
[9] https://genomicsandhealth.org/files/public/GA4GH_REWG_Data Sharing Lexicon.pdf

Depending on the size of the research group and the organisational structure, the roles of Data Owner/Sponsor and Data Custodian/Steward may be performed by one person or by different people. They may be dedicated roles or be in addition to other duties. The **Data Owner/Sponsor** may be a researcher or may be a Senior Officer within the organisation. The **Data Custodian/Steward** is responsible for the day-to-day management of the data collection and may authorise disclosure of information to other researchers or for other research purposes. Organisational involvement and sign off is important where data collections are independent of any one researcher or they endure beyond the tenure of their original sponsor/custodian.

It is recommended that staff at each organisation storing data on med.data, understand these roles and associated responsibilities and identify relevant person(s) to undertake these roles. It is also recommended that prior to data ingestion into med.data, institutional authorisation should be obtained as well as clarity over any concerns or dispute of data ownership.

4.1.2 Node Operator

The Node Operator is the organisation that provides med.data services at a particular location. It is proposed that each Node Operator will meet stated standard terms of conditions of service provision. It is anticipated that these terms and conditions will describe the service in terms of availability, data management, back up, business continuity, minimum security, access and identification management standards. Where a data collection has requirements over and above the Terms and Conditions, the Data Sponsor and The Node Operator can establish a service level agreement (SLA).

med.data nodes may wish to consider developing a consistent internal governance checklist to ensure that each data collection has an identified Data Owner/Sponsor and Data Custodian/Steward. There may also be a Technical Contact for each data collection, but this is not a requirement.

4.2 RESPONSIBILITIES

4.2.1 Data Custodians/Stewards

Data Custodians/Stewards have a responsibility to ensure that it is appropriate the data they choose to store on med.data is done so in compliance with a number of relevant laws and codes discussed in this paper. Commonly questions asked by a Data Custodian/Steward include: Is my data stored protected on secure servers? Where is my data physically located? How is access to my data controlled? Are the controls available appropriate for my data? Concerns of this type are not new and should be examined by a Data Custodian/Steward whether data is stored in a shared facility such as med.data, or within institutional infrastructure.

In terms of data sharing within med.data, this will solely be determined and controlled by the Data Custodian/Steward and that this role alone will make their data assets both discoverable and shareable to other researchers or policy makers in the appropriate format (e.g. aggregated data, de-identified data sets or identified datasets), only under the appropriate circumstances.

4.2.2 Node Operators

Node operators have a responsibility to ensure that Data Custodian/Stewards can securely store and control access to their data (to a level that is appropriate for their particular type of data) with a high degree of confidence. Data storage sites and security controls need to be clearly defined. It is not proposed that any node operator of med.data will ever determine who, apart from the Data Custodian/Steward may access data, rather that med.data will provide the technical functionality that enables each Data Custodian/Steward to provide access to others in a secure manner when appropriate.

4.2.3 Shared Responsibility Model

The responsibility of data security in med.data is shared between both the Data Custodian/Stewards and the Node Operators. Node Operators need to ensure the med.data infrastructure operates securely. Data Custodian/Stewards need to ensure they secure the data they store on the infrastructure with a view of meeting relevant regulatory and compliance requirements.

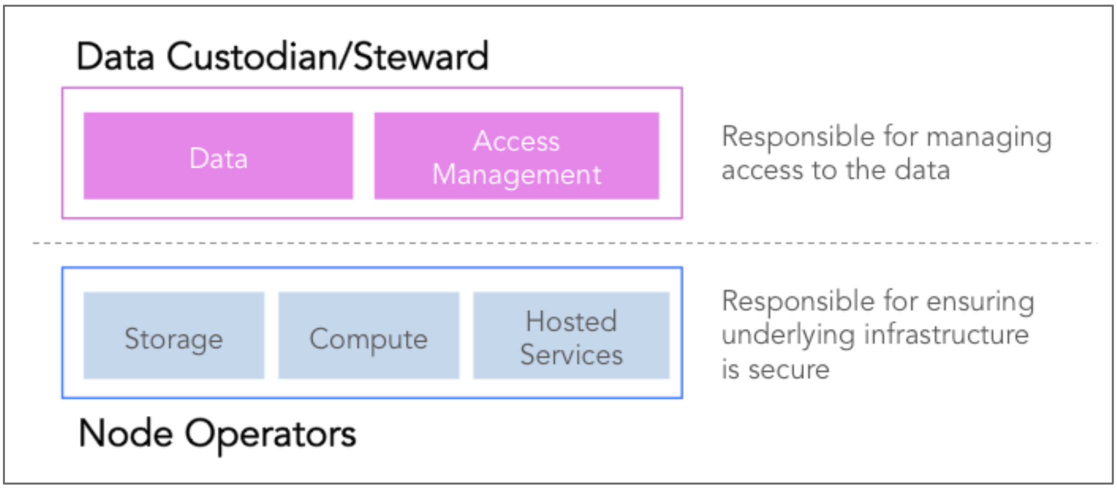
Data Custodian/Steward
Data
Access Management
Responsible for managing access to the data
Storage
Compute
Hosted Services
Responsible for ensuring underlying infrastructure is secure
Node Operators

Figure 1: Proposed Shared Responsibility Model

5. RELEVANT LEGISLATION

5.1 PRIVACY LAWS

The Australian Privacy Framework consists of Federal and State information privacy legislation and some sector-specific legislation at a State level. Note that not all jurisdictions have state or health specific privacy legislation, for example South Australia and Western Australia[10].

5.1.1 Federal Privacy Legislation

The Privacy Act (1988) applies to the collection, use, storage and disclosure of personal information in Australia. When an individual consents to their health information being collected for Health and Medical Research purposes, the Privacy Act does not apply. In cases where an individual does not consent, data re-use may be possible if approved by a Human Research Ethics Committee.

The Privacy Act explicitly describes circumstances where the use and disclosure of genetic information when informed consent has not been sought is allowed, however this does not include the use of genetic information in research. Guidelines for the use of genetic information in research are covered in the (non-legally binding) NHMRC's National Statement o Custodian n Ethical Conduct in Human Research.

If data stored on med.data includes personal or sensitive data of any form, it is the responsibility of each Data Custodian/Steward to control access to their data.

The Privacy Act (1988) applies to the collection, use, storage, disclosure of and access to any personal information[11] as held by Federal Government Departments and Agencies as well as the Private Sector. It includes 13 Australian Privacy Principles (APPs)[12] that cover the collection, use, disclosure and storage of personal information.

The Privacy Act and Health and Medical Research

The Privacy Act permits the handling of health information for health and medical research purposes in certain circumstances, where researchers are unable to seek individuals' consent. In the health and medical research context, privacy legislation generally only applies where consent is not sought or is considered impracticable. In cases where an individual consents to their health information being collected, used, stored or disclosed for health and medical research purposes, the Privacy Act does not apply.

The NHMRC is also authorised under section 95 (s95) of the Privacy Act to issue further guidance on the use of health and personal information for research purposes. The NHMRC s95 guidelines[13] provide a framework for the conduct of research using information held by Commonwealth agencies where identified information needs to be used without consent. In these situations, a Commonwealth agency may disclose, in identifiable form, records for research purposes without infringing the Privacy Act if the proposed research has been approved by a properly constituted Human Research Ethics Committee (HREC). The sister

[10] http://www.oaic.gov.au/privacy/other-privacy-jurisdictions/state-and-territory-privacy-law
[11] http://www.oaic.gov.au/privacy/privacy-act/the-privacy-act
[12] http://www.oaic.gov.au/privacy/privacy-act/australian-privacy-principles
[13] https://www.nhmrc.gov.au/health-ethics/human-research-ethics-committees-hrecs/health-research-privacy-framework

s95A guidelines[14] apply to the collection, use or disclosure of health information by the private sector for the purposes of research where it is impracticable to seek consent from the individual(s) involved. In both cases the s95 and s95A guidelines provide a framework for HRECs and those involved in conducting research to weigh the public interest in the use of the health information for specific purposes against the individuals' interest in privacy. The approving HREC holds responsibility for determining if information should be disclosed for research purposes[15].

Note that the s95 and s95A guidelines do not apply to State managed public health organisations, including public hospitals, arguably the richest source of health information. Some jurisdictions have elected to apply the s95 and s95A guidelines as Best Practice in the absence of jurisdictional specific health or privacy legislation.

With respect to the use of human genetic information in research, the Privacy Act does not prevent a health service provider using or disclosing a patient's genetic information if the patient has given informed consent. In cases where a health service provider has not been able to obtain consent from the patient, the Privacy Act allows the use and disclosure of genetic information where there is a serious threat to life, health or safety of a genetic relative. Note that the guidelines outlined in Section 95AA of the Privacy Act do not cover the use of genetic information in human research[16]. This is discussed in Chapter 3.5 of the National Statement on Ethical Conduct in Human Research[17] – See below).

5.1.2 State Privacy Legislation

5.1.2.1 New South Wales

In NSW, all health information is covered by HRIPA, which discerns between primary and secondary uses of health information. In cases where informed consent has not been given by an individual for use of their information in a research study, access to the information by the researcher must be approved by a HREC.

For researchers based in NSW, if data stored on med.data includes personal or sensitive data of any form, it is the responsibility of each Data Custodian/Steward to control access to their data.

In NSW, all health information is covered by the Health Records and Information Privacy Act 2002[18] (HRIPA), whether information is collected in the public or private sector. HRIPA regulates the collection, holding and use of Health Information through 15 Health Privacy Principles[19]. The Privacy and Personal Information Protection Act 1998[20] (PPIPA) is also applied to non-health information collected, stored and used in NSW, however as HRIPA imposes a higher standard it is generally applied and is not inconsistent with the NSW

[14] https://www.nhmrc.gov.au/guidelines/publications/pr2

[15] https://www.nhmrc.gov.au/health-ethics/human-research-ethics-committees-hrecs/health-research-privacy-framework

[16] https://www.nhmrc.gov.au/_files_nc/publications/attachments/pr3_use_of_genetic_information_s95aa_140311.pdf

[17] https://www.nhmrc.gov.au/book/chapter-3-5-human-genetics

[18] http://www.legislation.nsw.gov.au/fragview/inforce/act+71+2002+whole+0+N?nohits=y&tocnav=y&xref=Type%3Dact%20AND%20Year%3D2002%20AND%20no%3D71

[19] http://www.legislation.nsw.gov.au/fragview/inforce/act+71+2002+sch.1+0+N?nohits=y&tocnav=y&xref=Type%3Dact%20AND%20Year%3D2002%20AND%20no%3D71

[20] http://www.legislation.nsw.gov.au/xref/inforce/?xref=Type%3Dact%20AND%20Year%3D1998%20AND%20no%3D133&nohits=y

PPIPA. As with the Commonwealth Privacy Act, HRIPA only applies where consent is not sought and is a determination made by the HREC to disclose personal health information.

HRIPA clearly delineates between primary and secondary use of health information. The primary purpose is the main purpose for which the information was collected. In the health and medical research context, the primary purpose is a research project where consent is presumably sought. A secondary use would be an unrelated research project. HRIPA would apply where consent is NOT obtained from individuals to disclose and use their information for an unrelated, future health research project. An application would need to be made to an HREC and the HREC may or may not provide approval under the provisions in HRIPA (or s95A outside of NSW).

Personal health information held in medical records is collected for the primary purpose of providing treatment to an individual. Therefore, any use of medical records for research purposes without consent is considered a secondary use of information and HRIPA must be approved by a HREC. There are some exemptions to this – primarily information collected under legislation for public health purposes such as the NSW Central Cancer Registry[21].

5.1.2.2 Victoria

In Victoria all health information is covered by Victorian privacy legislation, which includes the *National Statement on Ethical Conduct in Research Involving Humans 1999*, *Health Records Act 2001*, and the *Privacy and Data Protection Act 2014*. The Victorian Department of Health has adopted 15 Health Privacy Principles and 11 Information Privacy Principles which are recommended as best practices[22].

The Privacy and Data Protection Act 2014 was passed to provide for responsible collection and handling of personal information in the Victorian public sector, establish a protective data security regime for the Victorian public sector, establish a regime for monitoring and assuring public sector data security and to establish the office of the Commissioner for Privacy and Data Protection and repeal the Information Privacy Act of 2000. The Commissioner is in the process of establishing the standards, protocols and the Victorian Protective Data Security Framework that all applicable agencies must implement over a two-year transition period for the protection of public sector data[23].

In regards to research use, when research or the compilation of statistics that are in the public interest cannot be undertaken with de-identified information, and where it is impractical to seek the individual's consent, the research or compilation of statistics will be carried out in accordance with the *National Statement on Ethical Conduct in Research Involving Humans* issued by the National Health and Medical Research Council (1999) and in accordance with the Health Services Commissioner guidelines[24].

The following legislation may be relevant for research data stored in Victoria:
Copyright Act 1968 (Commonwealth)
Privacy Act 1988 (Commonwealth)
Public Records Act 1973 (Victoria)
Health Records Act 2001 (Victoria)
Evidence Act 2008 (Victoria)
Freedom of Information Act 1982 (Victoria)
Crimes Act 1958 (Victoria)

[21] http://www.cancerinstitute.org.au/data-and-statistics/cancer-registries/nsw-central-cancer-registry-data-access
[22] http://www.health.vic.gov.au/privacy.htm#principles
[23] http://www.cpdp.vic.gov.au/practitioners/privacy/laws-and-standards
[24] http://www.education.vic.gov.au/pages/privacypolicy.aspx

5.1.2.3 Other Jurisdictions in Australia

Not all Australian states/territories have specific privacy legislation, or health specific privacy legislation. In these cases the Commonwealth Privacy Act is applied.

A summary of applicable Privacy legislation in all Australian State and Territories is outlined in Appendix One.

5.1.3 International Privacy Legislation

Whilst not legally binding within Australia and therefore not directly relevant for med.data.edu.au, summaries of the EU Data Protection Directive[25], US Health Information Portability and Accountability Act (HIPAA)[26] and US Genetic Information Nondiscrimination Act (GINA)[27] are included in Appendix Two.

5.1.4 Laws concerning cross-jurisdictional personal data transfer

Cross-jurisdictional transfer of personal information out of Australia is permitted under the Australian Privacy Principle (APP) 8, as long as the entity disclosing the information takes reasonable steps to ensure the overseas recipient does not breach any of the APPs.

In NSW the HRIPA Health Privacy Principle (HPP) 14 outlines that HREC and researcher togther must ensure that any recipient outside NSW is subject to substantially similar privacy standards or laws.

If data stored on med.data includes personal data of any form and is to be transferred to a 3rd party in another country or state, it is the responsibility of each Data Custodian/Steward to ensure that the recipient is subject to substantially similar privacy standards or laws. It is the responsibility of the Node Operators to clearly indicate the physical storage location of all data on med.data.

The Commonwealth Privacy Act 1998 (Australian Privacy Principle (APP) 8—cross-border disclosure of personal information)[28], outlines the conditions that must be met to enable such a transfer internationally from Australia. The key points of APP8 are that before an entity discloses personal information to an overseas recipient, the entity must take reasonable steps to ensure that the overseas recipient does not breach any of the APPs[29] in relation to the information (APP 8.1) and that the entity that discloses personal information to an overseas recipient is accountable for any acts or practices of the overseas recipient in relation to the information that would breach any of the APPs (s 16C).

[25] http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:31995L0046
[26] https://www.gpo.gov/fdsys/pkg/PLAW-104publ191/html/PLAW-104publ191.htm
[27] https://www.govtrack.us/congress/bills/110/hr493/text
[28] http://www.oaic.gov.au/privacy/applying-privacy-law/app-guidelines/chapter-8-app-8-cross-border-disclosure-of-personal-information
[29] http://www.oaic.gov.au/privacy/privacy-act/australian-privacy-principles

In NSW, HRIPA Health Privacy Principle (HPP)-14: Trans-border data flows and data flow to Commonwealth agencies[30] requires that before transferring health information out of New South Wales the HREC and researcher must ensure that the recipient is subject to substantially similar privacy standards or laws. If equivalent privacy protections do not exist, then it is only possible to transfer the health information out of NSW under certain circumstances as outlined in HRIPA[31]. It is recommended that either individuals or an HREC provide consent for the transfer of data across borders and this is best established at the data collection stage. In the absence of this, clear contractual provisions must be in place to unsure the security and integrity of the data.

5.2 LAWS RELEVANT TO CLINICAL TRIALS

The US Food and Drug Administration (FDA)'s Title 21 Code of Federal Regulations (CFR), Part 11, concerns the requirement for electronic records and signatures to be trustworthy and reliable. Its intention is to prevent fraud while enabling the use of technology to accelerate clinical research processes. Compliance with 21CFR11 is often cited in contracts and protocols for clinical trials originating out of the USA or specifically seeking FDA approval.

If data stored on med.data includes information arising from clinical trials originating in the USA or requiring FDA approval, it is a responsibility of the Node Operators to ensure that the processes and infrastructure is in place to enable securing, monitoring and auditing electronic data systems.

21CFR11 is the US Food and Drug Administration (FDA)'s Title 21 Code of Federal Regulations (CFR), Part 11, concerning electronic records and electronic signatures[32]. It establishes the requirement for electronic records and signatures to be trustworthy and reliable. Its intention is to prevent fraud while enabling the use of technology to accelerate clinical research processes. The CFR outlines requirements for securing, monitoring and auditing electronic data systems. Compliance with 21CFR11 is frequently cited in contracts and protocols for clinical trials originating out of the USA or specifically seeking FDA approval. **Note that 21CFR11 is not reflected in Australian legislation and compliance in Australia is a matter of enhanced industry practice.** Note that several Clinical Trial Management Tools including open source options such as OpenClinica[33] do support 21CFR11 and may be candidates for hosting on med.data.

6. CODES AND POLICIES

In addition to legislation outlined previously, which governs the use of personal and health information in research, a number of Codes of best practice in research ethics and data management and associated Policies are also relevant to users of med.data and its operation.

[30] http://www.legislation.nsw.gov.au/fullhtml/inforce/act+71+2002+FIRST+0+N#sch.1
[31] http://www.ipc.nsw.gov.au/sites/default/files/filemanager/hripahealthhandbook.pdf (Section 2.9 (p47))
[32] http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm?cfrpart=11&showfr=1
[33] https://www.openclinica.com/

The NHMRC, ARC, Universities Australia and various International Organisations have issued a number of Codes and Statements pertaining to the use of data (including health data) in research. Whilst not legally binding documents, they are widely respected as authoritative information on the conduct of ethical research in Australia. They do not provide specific guidance regarding relevant technical specifications, compliance and operation policies needed for med.data.

For any data stored on med.data, Node Operators need to demonstrate to Data Custodian/Stewards that their node aligns to their local institutional (or other) policies and procedures for each dataset stored.

6.1 NATIONAL STATEMENT ON ETHICAL CONDUCT IN HUMAN RESEARCH

The Australian National Statement on Ethical Conduct in Human Research (2007) – Updated March 2014[34] consists of a series of Guidelines made in accordance with the National Health and Medical Research Council Act 1992[35] and outlines the principles and values that guide the conduct of ethical research.

It is primarily applied by HRECs who consider the risks and benefits research poses to intended participants. The National Statement provides advice on the design and conduct of research across a number of research methodologies including qualitative and quantitative research, interventions and clinical trials, genetic research, biobanking and data banking. It also provides detailed guidance on conducting research, including consent in participant groups that may be vulnerable. This includes people highly dependent on medical care who may be unable to give consent, the presence of unequal relationships, children and young people, Aboriginal and Torres Strait Islander peoples and pregnant women and people with cognitive impairment or intellectual disability.

The National Statement is not a legally binding document but is widely respected as the authority on the conduct of ethical research in Australia. Therefore its application should be considered mandatory to all research activities involving humans, even if the risk posed to participants is low.

6.2 AUSTRALIAN CODE FOR THE RESPONSIBLE CONDUCT OF RESEARCH

The Australian Code for the Responsible Conduct of Research (2007)[36] (published jointly by the ARC, NHMRC and Universities Australia) is the overarching guide for the responsible conduct for research in Australia. It assigns researchers and their institutions a shared responsibility to manage research data and primary materials well, by addressing aspects of ownership, storage and retention, and accessibility.

The code outlines that researchers are required to manage their data – using methods appropriate to the discipline and to the nature of the data – to the highest standards including legislation, policies, funding agency requirements, technical protocols, audit and accreditation processes, discipline norms and the expectations of the broader community.

However the code does not provide specific guidance to researchers and institutions, rather delegating the need for specific policies to institutions. Although the Code is a key document and compliance is compulsory for institutions that receive NHMRC funding – it

[34] https://www.nhmrc.gov.au/book/national-statement-ethical-conduct-human-research
[35] http://www.comlaw.gov.au/Series/C2004A04516
[36] http://www.nhmrc.gov.au/guidelines-publications/r39

lacks clear standards applicable to data management. The Institution ultimately decides if its policies and procedures meet the Code.

6.3 NHMRC AND ARC DATA ACCESS POLICIES

Both the NHMRC and the ARC have released policies[37,38], that touch on the open access dissemination of research findings where an NHMRC or ARC Grant funds the research. These policies are intended to maximise the benefits of publically funded research through requiring that any publication arising from ARC or NHMRC supported research be made available through an institutional repository within 12 months of publication. Furthermore, associated funding agreements from the NHMRC and ARC[39,40], also explicitly refer to the dissemination of related data via deposition in an appropriate publically accessible subject and/or institutional data repository.

Moreover, the NHMRC has recently undertaken a Targeted Consultation on the Draft Principles for Accessing and Using Publicly Funded Data for Health Research[41], that aims to provide a framework for researchers and data Custodian/Stewards to consider when requests or applications are made for access to existing health and health-related datasets for research purposes. The principles aim to clarify responsibilities of Custodian/Steward and users and in doing so improve access to data for research purposes.

Additionally, the 2015 NHMRC Statement on Data Sharing[42] acknowledges the position of the NHMRC with respect to encouraging data sharing and providing access to data and other research outputs (e.g. metadata, analysis code, study protocols, study materials and other collected data) arising from NHMRC supported research. The NHMRC acknowledges that the level of detail in which data could be shared may be limited by a wide range of factors (e.g. ethics (particularly consent), legal, IP, data format and standards and variable ontologies used to describe data). Researchers are therefore encouraged by the NHMRC to share data with as much breadth and depth as possible, while taking into account their ethical-legal obligations, and providing sufficient metadata to allow others to reuse their data. The NHMRC also notes in this statement that the infrastructure and mechanisms for data sharing are currently available through individual institutions (e.g. Universities, Medical Research Institutes), government repositories (e.g. data.gov.au[43]), international repositories (e.g. Dryad[44]), established networks (e.g. Population Health Research Network[45], Research Data Storage Infrastructure[46] (which underpins the storage component of med.data) or nationwide registry and data organisations which offer varying levels of support (e.g. Australian National Data Service[47] (a partner in developing robust data and metadata management capabilities of med.data), Biogrid[48] and Intersect[49] (the operator of the NSW node of med.data)).

[37] https://www.nhmrc.gov.au/grants-funding/policy/nhmrc-open-access-policy

[38] http://www.arc.gov.au/applicants/openaccess.htm

[39] https://www.nhmrc.gov.au/filesnhmrc/file/grants/funding/funded/manage/policy/nhmrcfundingagreement1january2014.pdf (See Section 12.9)

[40] http://www.arc.gov.au/pdf/DP16/DPDiscoveryProgram2015-16fundingrules.pdf (see Section A11.5.2)

[41] http://consultations.nhmrc.gov.au/publicconsultations/funded-data

[42] http://www.nhmrc.gov.au/grants-funding/policy/nhmrc-statement-data-sharing

[43] http://data.gov.au/

[44] http://datadryad.org/

[45] http://www.phrn.org.au/

[46] https://www.rdsi.edu.au/

[47] http://www.ands.org.au/

[48] https://www.biogrid.org.au/

[49] http://www.intersect.org.au/

6.4 NATIONAL PRINCIPLES OF INTELLECTUAL PROPERTY MANAGEMENT FOR PUBLICLY FUNDED RESEARCH

Intellectual Property (IP) covers a wide range of intangible property that is the result of the creative and intellectual effort of individuals and organisations, which includes scientific discoveries, computer programs and databases[50].

In Australia, guidance is provided for the ownership, promotion, exploitation, and where appropriate, protection of IP generated through Australian Government funded research through the National Principles of Intellectual Property Management for Publicly Funded Research[51]. These principles are relevant to government-funded research that have been funded through grants awarded by ARC, NHMRC, and other government research funding schemes. In cases where research is not government-funded (e.g. research conducted by government departments / agencies for its own purposes), issues of IP should be addressed by internal procedures and agency contracts.

The National Principles state that ownership and the associated rights of all IP generated as a result of Australian Government competitively funded research will initially be vested in the Research Institution receiving and administering the grants. Research Institutions must therefore have policies in place relating to the ownership and availability for exploitation of IP generated as a result of Australian Government competitive funding. These policies must foster the most valuable use of this IP by industry and commercial ventures, governments, and the research sector by both: (a) Making the IP openly accessible through licensing and accessibility arrangements which allow for its use and re - use, including potentially for commercial exploitation; and (b) Protecting the IP through licensing and accessibility arrangements which provide exclusive opportunities to undertake commercial exploitation.

Research Institutions should, in addition, assist in the management of IP by providing systems to: (a) Identify where data (including datasets and databases), generated by Australian Government funded research, constitutes IP; and (b) Support the management of the data from which the IP was derived - including data which constitutes intellectual property, and data which does not constitute intellectual property - in order to maximise the benefits from the research, including the documentation and safe storage of data for future use.

6.5 BEST PRACTICE GUIDELINES FOR CLINICAL RESEARCH

The most commonly applied standard is the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use – Good Clinical Practice (ICH-GCP)[52]. These are the technical requirements of conducting clinical trials for the registration of pharmaceutical products for human use.

Locally, these guidelines are implemented through policy issued by the Australian Government's Therapeutic Goods Administration (TGA) [53] . The TGA has adopted CPMP/ICH/135/95 in principle but has recognised that some elements are, by necessity, overridden by the National Statement on Ethical Conduct in Human Research[54], (and therefore not adopted) and that others require explanation in terms of 'local regulatory requirements'.

While ICH-GCP does not make significant commentary on the storage or use of data from the perspective of service provision; it does provide clear advice on Data Handling and

[50] http://www.ag.gov.au/RightsAndProtections/IntellectualProperty/Documents/IntellectualPropertyManual.pdf
[51] https://www.nhmrc.gov.au/grants-funding/policy/intellectual-property-management/national-principles-intellectual-property-man
[52] http://www.ich.org/products/guidelines/efficacy/efficacy-single/article/good-clinical-practice.html
[53] https://www.tga.gov.au/publication/note-guidance-good-clinical-practice
[54] https://www.nhmrc.gov.au/book/national-statement-ethical-conduct-human-research

Record Keeping for clinical trials as well as the roles and responsibilities of the Trial Sponsor through an Independent Data-Monitoring Committee (IDMC) in Section 5.5[55] (See Appendix Three).

6.6 GENOMIC DATA SHARING POLICIES

Whilst no specific policy yet exists within Australia with respect to the sharing of genomic data, in 2015, the US National Institutes of Health (NIH) issued a Genomic Data Sharing (GDS) Policy[56]. This Policy applies to all NIH-funded research that generates large-scale human or non-human genomic data (genome-wide association studies (GWAS), single nucleotide polymorphisms (SNP) arrays, genome sequence, transcriptomic, metagenomic, epigenomic, and gene expression data) as well as the use of these data for subsequent research.

NIH expects investigators and their institutions to provide a basic "Genomic Data Sharing Plan" in funding applications. Any resources that may be needed to support a proposed genomic data sharing plan should be included in the project's budget.

Non-human genomic data may be made available through any widely used data repository, whether NIH-funded or not such as the Gene Expression Omnibus (GEO)[57], Sequence Read Archive (SRA)[58], Trace Archive[59], ArrayExpress[60], Mouse Genome Informatics (MGI)[61], WormBase[62], Zebrafish Model Organism Database (ZFIN)[63], GenBank[64], European Nucleotide Archive (ENA)[65] or the DNA Data Bank of Japan (DDBJ)[66].

Conversely, human-derived genomic data:

(1) should be submitted with relevant associated data (e.g. phenotype and exposure data) to an NIH-designated data repository (i.e. a data repository that is maintained or supported by NIH either directly or through collaboration[67]). Investigators should de-identify human genomic data that they submit to the NIH-designated data repository according to the standards set forth in the HHS Regulations for the Protection of Human Subjects[68] to ensure that the identities of research subjects cannot be readily ascertained. Investigators should also strip the data of identifiers according to the HIPAA Privacy Rule (see section 5.1.3. of this Discussion Paper). The de-identified data should be assigned random, unique codes by the investigator, and the key to other study identifiers held by the submitting institution.

(2) With respect to data repositories, NIH-funded researchers are expected to register all studies with human genomic data that fall within the scope of the GDS Policy in dbGaP[69], regardless of which NIH-designated data repository (e.g. dbGaP, GEO, SRA, the Cancer Genomics Hub[70]) will receive the data. The policy also states that NIH-funded data repositories need not be the exclusive source for facilitating the sharing of genomic data, however under these circumstances investigators should ensure that appropriate data security measures are in place within the non NIH-designated data repository, and that confidentiality, privacy, and data use measures are consistent[71] with the GDS Policy (see

[55] https://www.tga.gov.au/sites/default/files/ich13595an.pdf
[56] http://gds.nih.gov/03policy2.html
[57] http://www.ncbi.nlm.nih.gov/geo/
[58] http://www.ncbi.nlm.nih.gov/sra/
[59] http://www.ncbi.nlm.nih.gov/Traces/home/
[60] https://www.ebi.ac.uk/arrayexpress/
[61] http://www.informatics.jax.org
[62] https://www.wormbase.org
[63] http://zfin.org
[64] http://www.ncbi.nlm.nih.gov/genbank/
[65] http://www.ebi.ac.uk/ena
[66] http://www.ddbj.nig.ac.jp
[67] http://gds.nih.gov/02dr2.html
[68] http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html#46.102.
[69] http://www.ncbi.nlm.nih.gov/gap
[70] https://cghub.ucsc.edu
[71] http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/GetPdf.cgi?document_name=dbgap_2b_security_procedures.pdf

Section 7.5 for further information outlining the NIH security requirements).
(3) The informed consent under which the data or samples were collected is the basis for the submitting institution to determine whether the data should be available through unrestricted or controlled access. Controlled-access data in NIH-designated data repositories are to be made available for secondary research only after investigators have obtained approval from NIH to use the requested data for a particular project.
(4) For research that falls within the scope of the GDS Policy, submitting institutions, through their IRB (the US equivalent of a HREC), are to review the informed consent materials to determine whether it is appropriate for data to be shared for secondary research use, and
(5) The responsible Institutional Signing Official of the submitting institution should provide an Institutional Certification prior to award consistent with the genomic data-sharing plan that states whether the data will be submitted to an unrestricted- or controlled-access database.
Additionally, the responsibilities of Investigators Accessing and Using Genomic Data are outlined in the GDS Policy, where requests for Controlled-Access Data are to be reviewed by NIH Data Access Committees (DACs), and investigators approved to download controlled-access data from an NIH-designated data repository and their institutions are expected to abide by the NIH Genomic Data User Code of Conduct[72] through their agreement to the Data Use Certification[73].

7. SECURITY

A number of Frameworks and Standards pertaining to information security exist and whilst not generally specifically focussed on health data, these are potentially relevant to the storage of personal and sensitive data on med.data.

7.1 AUSTRALIAN GOVERNMENT NATIONAL INFORMATION GOVERNANCE FRAMEWORK

The Australian Government Protective Security Policy and the associated Information Security Manual outline personnel, information and physical security measures that must be observed when undertaking Commonwealth Government business. These may be relevant to med.data if Commonwealth data such as Medicare data is held in the facility.

7.1.1 Australian Government Protective Security Policy

The Australian Government's protective security policy is organised in a tiered, hierarchical structure—the Protective Security Policy Framework (PSPF)[74]. It describes the higher-level protective security outcomes and identifies the mandatory requirements in undertaking Government Business. The core policies cover personnel security, information security and physical security for the following types of agencies and bodies[75]:

- Non-corporate Commonwealth entities subject to Public Governance, Performance and Accountability Act 2013 (PGPA Act)

[72] http://gds.nih.gov/pdf/Genomic_Data_User_Code_of_Conduct.pdf
[73] http://gds.nih.gov/pdf/Model_DUC_7-26-13.pdf

[74] http://www.protectivesecurity.gov.au/Pages/default.aspx
[75] http://www.protectivesecurity.gov.au/governance/Pages/Applicability-of-the-Protective-Security-Policy-Framework.aspx

- Corporate Commonwealth entities and companies subject to the PGPA Act that have received Ministerial direction to apply the protective security policies of the Australian Government, and
- Other bodies established for a public purpose under a law of the Commonwealth and other Australian Government agencies, where the body or agency has received a notice from the relevant Minister that the PSPF applies to them.

A complete list of the requirements, are outlined in Appendix Four. In instances where med.data nodes hold data (or duplicate data) derived from Federal Government Departments or Agencies (e.g. Medicare Data), compliance would be desirable.

7.1.2 Australian Government Information Security Manual (ISM)

The Australian Signals Directorate (ASD) produces the Australian Government Information Security Manual (ISM)[76]. The manual is the standard that governs the security of government ICT systems and details controls to mitigate and manage threats to information security. It complements the Protective Security Policy Framework discussed above.

The purpose of the ISM is to assist Australian government agencies in applying a risk-based approach to protecting their information and systems and applies to:

- Australian government agencies that are subject to the Financial Management and Accountability Act 1997, or the Public Governance, Performance and Accountability Act 2013.
- Bodies that are subject to the Commonwealth Authorities and Companies Act 1997, or the Public Governance, Performance and Accountability Act, and that have received notice in accordance with that Act that the ISM applies to them as a general policy of the Government.
- Other bodies established for a public purpose under the law of the Commonwealth and other Australian government agencies, where the body or agency has received a notice from their Portfolio Minister that the ISM applies to them.
- State and territory agencies that implement the Australian Government Protective Security Policy Framework.
- Organisations that have entered a Deed of Agreement with the Government to have access to sensitive or classified information.

The ISM Controls Manual[77] outlines in detail requirements with respect to: Industry Engagement and Outsourcing, Roles And Responsibilities, Information Security Documentation, System Accreditation, Information Security Monitoring, Cyber Security Incidents, Physical Security For Systems, Personnel Security For Systems, Communications Security, Product Security, Media Security, Software Security, Email Security, Access Control, Secure Administration, Network Security, Cryptography, Cross Domain Security, Data Transfers And Content Filtering.

Appendix Five of the ISM Controls Manual applies to outsourcing of data storage and processing of data from a relevant Australian Government agency in a public cloud (such as Amazon Web Services[78], Microsoft Azure[79] or Google Cloud[80]), or offshore facilities. As med.data does not represent either a public cloud, or an offshore solution, these controls

[76] http://www.asd.gov.au/infosec/ism/
[77] http://www.asd.gov.au/publications/InformationSecurityManual2014Controls.pdf
[78] http://aws.amazon.com/
[79] http://azure.microsoft.com/en-us/
[80] https://cloud.google.com/

are not directly relevant to med.data per se, but do point to considerations that any Government Agency may wish to consider if any data may be held on outsourced facilities such as med.data.

7.2 NATIONAL eHEALTH SECURITY & ACCESS FRAMEWORK (NESAF)

The National eHealth Security and Access Framework (NESAF) is focussed at providing guidance for IT professionals within the clinical care sector to ensure patient data is adequately protected. It is however potentially useful for determining the security best practices that should be employed when storing health-derived information for research.

The NESAF recommends that where health information is used for purposes other than primary health care (e.g. for research), any such data is de-identified.

The National eHealth Security and Access Framework (NESAF)[81] provides guidelines for the Australian clinical care sector to implement secure systems that protect patient data and eHealth-related assets, while providing the provenance required for ensuring patient safety and privacy. It offers a risk-based approach where risk identification and analysis is used to establish appropriate security and access controls within an organisation. It is not a legally binding document.

The NESAF has been developed by the National E-Health Transition Authority (NEHTA)[82], in consultation with various subject matter experts (security and privacy experts, clinicians, consumers, etc.).

It contains six component documents[83] of which the NESAF v4.0 Framework Model and Controls[84] describes 11 key security and access areas relating to eHealth[85] and their associated security controls (which are based on a number of Australian and International Standards for information security management, and information security management in health).

The intended audience of the document is IT professionals responsible for overseeing information security, or involved in specifying, designing and building security controls for their healthcare organisation. It describes a standards-based model and relevant industry standards, including ISO27799 and ISO27001 (see Section 7.4).

Although it is aimed primarily at organisations directly collecting, and managing patient data within a clinical care framework, and there is only one specific reference to the use of health information in research (de-identification of health information)[86], the NESAF is potentially

[81] https://www.nehta.gov.au/implementation-resources/ehealth-foundations/national-ehealth-security-and-access-framework
[82] https://www.nehta.gov.au/
[83] Overview; Business Blueprint; Framework Model and Controls; Implementer Blueprint; Standards Mapping and Release Notes
[84] https://www.nehta.gov.au/implementation-resources/ehealth-foundations/EP-1544-2014/NEHTA-1549-2014
[85] Information Security Policy; Organising Information Security; Human Resources Security; Physical and Environmental Security; Communications and Operations Management; Asset Management; Access Control; Information systems acquisition, development and maintenance; Information security incident management; Information security aspects of business continuity management ; Compliance.
[86] Control C.2.3. De-identification of health information output. Health information systems should enable the de-identification of healthcare information output where such data are used for purposes other than the clinical care of patients. The privacy of personal or health information should be maintained when used for purposes other than clinical care, and in line with privacy law requirements, for instance, research or statistical purposes for public health or public safety. De-identification of personal health information is more than simply removing the patient's name. Whenever the information is in the form of individual data sets, there is a risk that the data set could be linked to a particular individual on the basis of details of age, postcode and medical condition. The more information included in the data set, the greater the risk of identification. Even where data is aggregated, care should be taken that the number of people in each "cell" or sub-group is sufficient to ensure that the privacy of the individuals involved is not compromised. If de-identification is not possible, and where it is impracticable to obtain the consent from the individual involved, the NHMRC Privacy Guidelines should be used. Control Sources: (a) Royal Australian College of General Practitioners Handbook for the Management of Health Information in Private Medical Practice (b) NHMRC Privacy Guidelines.

useful for determining the security best practices that should be employed when storing personal health information for any purpose (including research).

Note that also, an industry guide is currently being developed as part of NESAF v4.0 that is intended to address security for healthcare organisations looking at implementing cloud computing solutions[87] (which is potentially pertinent to efforts such as the med.data.edu.au project).

7.3 AUTHENTICATION FRAMEWORKS

Authentication is the process of determining whether someone is, in fact, who they declare they are. A number of national authentication frameworks have been developed to enable authentication to a desired level of assurance and confidence. The levels of assurance required for a system, is directly related to the severity of consequence if information is released to an inappropriately authenticated party.

Node Operators require expert guidance on the features of an Authentication framework that is suitable for providing appropriate levels of assurance for identifiable, re-identifiable and non-identifiable data.

Authentication is the process of determining whether someone is, in fact, who they declare they are. In this process, the credentials provided are compared to those on file in a database of authorised users' information within an authentication server. If the credentials match, the process is completed and the user is granted authorisation for access.

For shared data systems such as med.data, authentication is an important process, as it will enable users of the system to be identified with confidence and access to be subsequently granted to data held on the system only to qualified users.

A number of national authentication frameworks exist that are relevant to both the data intended to be held in med.data, and the intended audience of users. Node Operators will require expert guidance on the features of an Authentication framework that is suitable for controlling access to identifiable, re-identifiable and non-identifiable data.

7.3.1 National e-Authentication Framework

The National e-Authentication Framework (NeAF)[88] provides assistance to government agencies, jurisdictions and industry sectors in authenticating a party to a desired level of assurance and confidence. The NeAF encompasses the electronic authentication (e-authentication) of the identity of individuals and businesses dealing with the government, on one side of the transaction, as well as the authentication of government websites on the other side. While the Framework supports an agency-specific model where each agency develops its own separate technology solution, it recognises and accommodates broader sectoral and whole of government e-Authentication initiatives. These are supported through the re-use of existing authentication credentials and consideration of a variety of identity management frameworks.

Assurance levels are used to describe the level of importance of getting e-Authentication right and the resultant level of robustness of the required solution. The NeAF determines 5

[87] https://www.nehta.gov.au/implementation-resources/ehealth-foundations/national-ehealth-security-and-access-framework

[88] http://www.finance.gov.au/policy-guides-procurement/authentication-and-identity-management/national-e-authentication-framework/

assurance levels based upon the assessment of the threats and impacts to agencies and/or end-users of getting e-Authentication wrong:

		Minimal assurance	Low assurance	Moderate assurance	High assurance
	Level 0	Level 1	Level 2	Level 3	Level 4
	No confidence is required in the identity assertion.	Minimal confidence is required in the identity assertion.	Low confidence is required in the identity assertion.	Moderate confidence is required in the identity assertion.	High confidence is required in the identity assertion.

Illustrative consequences and severity are discussed for several aspects of the information held. For data types potentially to be held in med.data, the most relevant are personal or sensitive data:

Consequence	Insignificant	Minor	Moderate	Major	Severe
Release of personally or sensitive data to third parties without consent	No impact	Would have little impact	Measurable impact, breach of regulations or commitment to confidentiality	Release of information would have a significant impact	Would have severe consequences to a person, agency or business.

7.3.2 Australian Access Federation

NeAF levels of assurance 1-4 are consistent with levels 1-4 defined by the US National Institute of Standards and Technology (NIST)[89] in their Electronic Authentication Guideline – NIST SP 800-63-2[90] which is the basis of the Assurance Framework adopted by the Australian Access Federation (AAF)[91].

The AAF has been established to provide the means of allowing a participating institution (e.g. University or Medical Research Institute) and/or a service provider to trust the information it receives from another participating institution. This is to provide seamless access to resources and secure communication by removing most of the roadblocks to collaboration and sharing at both the institutional and end user levels. As the NIST guideline forms the basis of many assurance frameworks used internationally, it was selected by AAF with a view to being interoperable with other federations.

Currently AAF currently provide only Level 1 and Level 2 assurance (Minimal or Low confidence is required in the identity assertion)[92] it is unclear whether access control and identity management via the Australian Access Federation will be sufficient for sensitive data, since the consequences of inadvertently releasing personal or sensitive data to third parties without consent due to a lack of appropriate authentication mechanisms may have a major or severe consequence.

[89] http://aaf.edu.au/technical/levels-of-assurance/
[90] http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-63-2.pdf
[91] http://aaf.edu.au
[92] http://aaf.edu.au/technical/levels-of-assurance/

7.3.3 National Authentication Service for Health (NASH)

The National Authentication Service for Health (NASH)[93] makes it possible for healthcare providers and supporting organisations to securely access and exchange health information. NASH provides Public Key Infrastructure (PKI) Certificates that help an organisation to: access the Personally Controlled Electronic Health Record (eHealth record) system and to send and receive messages securely using software that meets the requirements of Secure Message Delivery[94]. NASH PKI Certificates can be issued to healthcare providers and supporting organisations that are registered in the Healthcare Identifiers Service[95].

7.4 INTERNATIONAL AND AUSTRALIAN INFORMATION SECURITY STANDARDS

A number of Information Security Standards have been identified as relevant to the collection, use, storage and disclosure of health and medical data, including data used for research.

It is envisaged that med.data will require compliance with the general information security standards listed. The Health Informatics standards may be used as a recommended standard or template for researchers when designing data collection or preparing data for deposit to med.data.

Use of agreed data standards and classifications by the Node Operators will enhance the linkage and shareability of data by Data Custodian/Stewards.

The following standards have been identified as relevant to the collection, use, storage and disclosure of health and medical data, including data used for research. This list is not exhaustive and will be refined based on continuing functional specification. It is envisaged that med.data will require compliance with the general information security standards listed and the Health Informatics standards may be used as a recommended standard or template for researchers when designing data collection or preparing data for deposit to med.data. Agreed data standards and classifications will enhance the linkage and shareability of data.

7.4.1 General Information Technology Standards

AS/NZS ISO/IEC 27001:2006 Information technology — Security techniques — Information security management systems — Requirements [96]
AS/NZS ISO/IEC 27002:2006 Information technology — Security techniques — Code of practice for information security management [97]
AS/NZS ISO/IEC 27005:2012 Information technology — Security techniques — Information Security Risk Management [98]
AS ISO 27799—2011 Information security management in health using ISO/IEC 27002 [99].
AS/NZS 8016:2013 Governance of IT enabled projects [100]

[93] http://www.nehta.gov.au/our-work/nash
[94] http://www.nehta.gov.au/implementation-resources/ehealth-foundations/secure-messaging
[95] http://www.nehta.gov.au/our-work/healthcare-identifiers-hi
[96] http://infostore.saiglobal.com/store/Details.aspx?DocN=AS0733774970AT
[97] http://infostore.saiglobal.com/store/Details.aspx?DocN=AS0733774921AT
[98] http://infostore.saiglobal.com/store/details.aspx?ProductID=1533756
[99] http://infostore.saiglobal.com/store/details.aspx?ProductID=1461737
[100] http://infostore.saiglobal.com/store/details.aspx?ProductID=1696546

AS/NZS 7799.2:2003 Information Security Management - Specification for Information Security Management Systems [101]

7.4.2 Health Informatics Standards

AS ISO 18308-2005 Health Informatics – Requirements for an electronic health record architecture. Includes standard for EHR privacy, consent and access levels [102]
AS 4700.6-2013 Implementation of Health Level 7 (HL7) [103]
AS 4846:2014 Person and Provider Identification in Healthcare [104]
ATS 516-2013 Digital Images for diagnostic and other clinical purposes [105]
ATS ISO 25237-2011 Pseudonymization [106]
SA HB 137-2013 Handbook E-health Interoperability Framework [107]
HB 291-2007 Handbook Health informatics—Guide to data development in health [108]
TR-2964-2010 Representing Archetyped Data in HL7 [109].
TR 4890-2008 HL7 messaging requirements for scheduling, bed availability, consent and eligibility[110]

7.5 NIH SECURITY BEST PRACTICES FOR CONTROLLED-DATA ACCESS

Whilst Australia does not yet have clear policy surrounding security best practices for controlling access to human-derived genomic data, the NIH Security Best Practices for Controlled-Access Data Subject to the NIH Genomic Data Sharing (GDS) Policy does provide clear and in-depth security guidelines for local and cloud computing infrastructure that is intended for storage and computation of these data in the US.

It is recommended that med.data Node Operators endeavour to demonstrate adherence to the NIH guidelines to provide assurance to Data Custodian/Stewards wishing to securely store and utilise human-derived genomic data.

The NIH Security Best Practices for Controlled-Access Data Subject to the NIH Genomic Data Sharing (GDS) Policy[111] provides guidance surrounding NIH's expectations for the management and protection of NIH controlled-access data transferred to and maintained by institutions whether in their own institutional data storage systems or in cloud-computing[112] systems. The NIH strongly recommends that investigators consult with institutional IT leaders, including the Chief Information Officer (CIO) and the institutional Information

[101] http://infostore.saiglobal.com/store/details.aspx?ProductID=391994
[102] http://infostore.saiglobal.com/store/details.aspx?ProductID=343008
[103] http://infostore.saiglobal.com/store/Details.aspx?ProductID=1625115
[104] http://infostore.saiglobal.com/store/details.aspx?ProductID=1753860
[105] http://infostore.saiglobal.com/store/details.aspx?ProductID=1693853
[106] http://infostore.saiglobal.com/store/details.aspx?ProductID=1466326
[107] http://infostore.saiglobal.com/store/details.aspx?ProductID=1635046
[108] http://infostore.saiglobal.com/store/Details.aspx?DocN=AS0733783554AT
[109] http://infostore.saiglobal.com/emea/details.aspx?ProductID=1385369
[110] http://infostore.saiglobal.com/store/details.aspx?ProductID=1085704
[111] http://www.ncbi.nlm.nih.gov/projects/gap/pdf/dbgap_2b_security_procedures.pdf

[112] Cloud computing, as defined by the National Institute for Standards and Technology (NIST), is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or cloud service provider interaction.

Systems Security Officer (ISSO) or equivalents to develop the formal information security plan prior to receipt of controlled access data from the NIH, and institutional signing officials should validate that an appropriate security plan is in place prior to accepting liability for data loss or breach on behalf of the institution.

The policy provides clear and in-depth security guidelines for local and cloud computing infrastructure (see Appendix Six).

8. METADATA STANDARDS

8.1 COLLECTION LEVEL METADATA

Functionality will be developed across med.data to allow Data Custodian/Stewards to create, publish and advertise the existence of data collections that are stored on the facility. Collection descriptions will broadly describe the contents of each data collection stored and also indicate the levels of access from open (publically available) through to private (e.g. where identifiable health information may be held) and clearly describe who may gain access to data, and under what circumstances. This is valuable to allow other researchers, organisations or policy makers to understand what data collections are stored in med.data and allows Data Custodian/Stewards to describe the content of a collection without disclosing sensitive information.

The metadata catalogue will be publically available and searchable to enable potential collaborators to assess the potential value of the data collections to their work. Descriptions will also be fed to Research Data Australia[113]; a national catalogue of research data collections that is indexed such that its contents can be searched via a standard web search engine.

As part of the data and collection technical analysis and ingestion process, Node Operators will work with Data Custodian/Stewards to identify what data descriptions should be publically available and searchable and ensure that these description clearly indicate the different levels of restricted access.

8.2 ITEM LEVEL METADATA

The format and standardisation of item-level metadata generated by health and medical research activities varies enormously with few globally accepted national and international standards. Where these do exist, the application of these is often inconsistent.

Certain standards such as HL7 Health Messaging[114], SNOMED Clinical Terms[115], or clinical trial protocols using SPIRIT[116] or Transcelerate[117] format are likely to be used for a proportion of data stored on med.data, as are metadata standards developed from a number of other national or international projects, however it is anticipated that the a significant amount of

[113] https://researchdata.ands.org.au/
[114] http://www.hl7.org.au
[115] http://www.nehta.gov.au/our-work/clinical-terminology/snomed-clinical-terms
[116] http://www.spirit-statement.org
[117] http://www.transceleratebiopharmainc.com/our-initiatives/clinical-data-standards/

item-level metadata in med.data will be undefined and/or project specific, providing the opportunity to explore options for standardisation.

9. SUMMARY AND CONCLUSIONS

The governance surrounding Health and Medical Research data (particularly that which includes personal or sensitive information derived from human participants) is a complex area and requires adherence to a wide range of policies, legislation and standards as outlined in this paper.

For the med.data facility, it is important to understand the breadth of the governance environment at the outset in order to plan for the development of services around Health and Medical Research data.

Current seed funding through the NCRIS RDS project is limited (totaling 4FTE for 18 months) and can set the groundwork to (a) design med.data as a fit-for purpose national facility for storing, managing and sharing all types of Health and Medical Research data (including sensitive personal information) and (b) to enable establishment of basic services to store data of these types. Considerable further investment is however likely to be required to enable the development of mature and accredited services that adhere to all in this area. As such, focusing available resources on prioritised requirements dictated by policy, legislation and security standards will be important for the establishment and increasing maturity of med.data over time and ongoing expert advice sourced from the med.data Advisory Board and elsewhere will be required to achieve this vision.

APPENDIX ONE

Summary of applicable Privacy legislation in all Australian State and Territories

Jurisdiction	Public Sector (including Public Health Organisation (PHO)s and State Health Agencies)	Private Sector (Health)	Private Sector (General)
ACT	Information Privacy Act 2014 (ACT) (ACT Public Sector Agencies) Health Records (Privacy and Access) Act 1997	Privacy Act 1988 (Clth) Health Records (Privacy and Access) Act 1997	Privacy Act 1988 (Clth)
NSW	Privacy and Personal information Protection Act 1998 Health Records and Information Privacy Act 2002 - Health records held by NSW Government agencies (including public hospitals)	Privacy Act 1988 (Clth)	Privacy Act 1988 (Clth)
NT Note: no health specific privacy legislation	Information Act (2002) (NT) – Applies to NT Government Organisations including PHOs.	Privacy Act 1988 (Clth)	Privacy Act 1988 (Clth)
QLD Note: no health specific privacy legislation	Information Privacy Act 2009 (Qld) Information Standards 42 (general) & 42A (health) Public Health Act 2005 Chapter 6, Part 4, Division 2, s281 – s284 (access to confidential information held by QLD Health	Privacy Act 1988 (Clth)	Privacy Act 1988 (Clth)
SA	There is no legislation that specifically addresses privacy in South Australia. The South Australian Department of the Premier and Cabinet, however, has issued an administrative instruction requiring its government agencies to comply with a set of Information Privacy Principles (IPPs) based on the IPPs in the Commonwealth Privacy Act[118]	Privacy Act 1988 (Clth)	Privacy Act 1988 (Clth)
TAS Note: no health specific privacy legislation	Personal Information and Protection Act 2004 (Tas) applies to the Tasmanian Public Sector, including the University of Tasmania	Privacy Act 1988 (Clth)	Privacy Act 1988 (Clth)
VIC	Privacy and Data Protection Act 2014 Health Records Act 2001 (Vic)	Privacy Act 1988 (Clth)	Privacy Act 1988 (Clth)
WA	There is no legislation that specifically addresses privacy in Western Australia	Privacy Act 1988 (Clth) Confidentiality of Health Information Committee	Privacy Act 1988 (Clth)

[118] Australian Law Reform Commission You’re your Information = Australian Privacy Law and Practice” (2008) http://www.alrc.gov.au/publications/2.%20Privacy%20Regulation%20in%20Australia/state-and-territory-regulation-privacy. Accessed 5th December 2014

APPENDIX TWO

International Privacy Legislation

United States of America

The US Federal **Health Insurance Portability and Accountability Act of 1996 (HIPAA) Public Law 104-191**, provides protections for individually identifiable health information held by covered entities and their business associates and gives patients an array of rights with respect to that information through a "Privacy Rule"[119]. The HIPAA Privacy Rule is balanced so that it permits the disclosure of health information needed for patient care and other important purposes, including for research purposes. Similar to the Australian procedures outlined above, for research projects, an Institutional Review Board (IRB, the US equivalent of a HREC) must review research projects for which explicit consent has NOT been received from study participants. An additional HIPAA "Security Rule"[120] specifies a series of administrative, physical, and technical safeguards for covered entities and their business associates to use to assure the confidentiality, integrity, and availability of electronic protected health information. The Security Rule operationalises the protections contained in the Privacy Rule by addressing the technical and non-technical safeguards that organisations called "covered entities" must put in place to secure an individual's "electronic protected health information". Note that the US National Institute of Standards and Technology (NIST) provide a HIPAA Security Toolkit Application[121] which is a self-assessment survey intended to help organisations better understand the requirements of the HIPAA Security Rule (HSR), implement those requirements, and assess those implementations in their operational environment.

The US Federal **Genetic Information Nondiscrimination Act (GINA)**[122] is designed to prohibit the use of genetic information to discriminate in health insurance and employment. GINA has implications for individuals participating in research studies. The US Department of Health and Human Services has issued guidance on integrating GINA into clinical research[123], and provides information on GINA's "research exemption" (which allows health insurers engaged in research to request (but not require) that an individual undergo a genetic test), considerations for Institutional Review Boards (i.e. HRECs), and integrating information on GINA into informed consents.

European Union

The EU Data Protection Directive (i.e. DIRECTIVE 95/46/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 24 October 1995[124]) is directed at EU Member States and is intended to protect the fundamental rights and freedoms of natural persons, and in particular their right to privacy with respect to the processing of personal data.

This legislation bears many similarities to the Australian Commonwealth Privacy Act (1988), and enshrines similar concepts:

- **Consent.** Personal data may be processed only if the data subject (i.e. a person) has unambiguously given his consent (Article 7).

[119] http://www.hhs.gov/ocr/privacy/hipaa/administrative/privacyrule/index.html
[120] http://www.hhs.gov/ocr/privacy/hipaa/administrative/securityrule/index.html
[121] http://scap.nist.gov/hipaa/
[122] https://www.govtrack.us/congress/bills/110/hr493/text
[123] http://www.hhs.gov/ohrp/policy/gina.pdf
[124] http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:31995L0046

- The prohibition of processing of personal health data - unless consent has been given or the data is required for the purposes of preventive medicine, medical diagnosis, the provision of care or treatment or the management of health-care services (Article 8).

- Informing data subjects of how their information is being used - “when the data have not been obtained from the data subject”, “providing the data subject with the identity of the data controller and the reasons for processing the data (unless the data processing is for statistical purposes or for the purposes of historical or scientific research, the provision of such information proves impossible or would involve a disproportionate effort or if recording or disclosure is expressly laid down by law)” (Article 11).

- Allowing the use of personal information in scientific research – “where there is clearly no risk of breaching the privacy of the data subject”, “when data are processed solely for purposes of scientific research or are kept in personal form for a period which does not exceed the period necessary for the sole purpose of creating statistics” (Article 13).

- Security of Processing – “Implementing appropriate technical and organizational measures to protect personal data against accidental or unlawful destruction or accidental loss, alteration, unauthorized disclosure or access, in particular where the processing involves the transmission of data over a network, and against all other unlawful forms of processing. The controller must, where processing is carried out on his behalf, choose a processor providing sufficient guarantees in respect of the technical security measures and organizational measures governing the processing to be carried out, and must ensure compliance with those measures” (Article 17).

- Cross-jurisdictional Data Transfer – “transfer to a third country of personal data which are undergoing processing or are intended for processing after transfer may take place only if, without prejudice to compliance with the national provisions adopted pursuant to the other provisions of this Directive, the third country in question ensures an adequate level of protection” (Article 25).

.

APPENDIX THREE

Roles and responsibilities of a Trial Sponsor through an Independent Data-Monitoring Committee (IDMC) as outlined in International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use – Good Clinical Practice (ICH-GCP)[125] Section 5.5[126]:

5.5.1. The sponsor should utilize appropriately qualified individuals to supervise the overall conduct of the trial, to handle the data, to verify the data, to conduct the statistical analyses, and to prepare the trial reports.

5.5.2. The sponsor may consider establishing an independent data-monitoring committee (IDMC) to assess the progress of a clinical trial, including the safety data and the critical efficacy endpoints at intervals, and to recommend to the sponsor whether to continue, modify, or stop a trial. The IDMC should have written operating procedures and maintain written records of all its meetings.

5.5.3. When using electronic trial data handling and/or remote electronic trial data systems, the sponsor should:

(a) Ensure and document that the electronic data processing system(s) conforms to the sponsor's established requirements for completeness, accuracy, reliability, and consistent intended performance (i.e. validation).

(b) Maintains SOPs for using these systems.

(c) Ensure that the systems are designed to permit data changes in such a way that the data changes are documented and that there is no deletion of entered data (i.e. maintain an audit trail, data trail, edit trail).

(d) Maintain a security system that prevents unauthorized access to the data.

(e) Maintain a list of the individuals who are authorized to make data changes.

(f) Maintain adequate backup of the data.

(g) Safeguard the blinding, if any (e.g. maintain the blinding during data entry and processing).

5.5.4. If data are transformed during processing, it should always be possible to compare the original data and observations with the processed data.

5.5.5. The sponsor should use an unambiguous subject identification code that allows identification of all the data reported for each subject.

5.5.6. The sponsor, or other owners of the data, should retain all of the sponsor-specific essential documents pertaining to the trial.

5.5.7. The sponsor should retain all sponsor-specific essential documents in conformance with the applicable regulatory requirement(s) of the country(ies) where the product is approved, and/or where the sponsor intends to apply for approval(s).

5.5.8. If the sponsor discontinues the clinical development of an investigational product (i.e. for any or all indications, routes of administration, or dosage forms), the sponsor should maintain all sponsor-specific essential documents for at least 2 years after formal discontinuation or in conformance with the applicable regulatory requirement(s).

5.5.9. If the sponsor discontinues the clinical development of an investigational product, the sponsor should notify all the trial investigators/institutions and all the regulatory authorities.

5.5.10. Any transfer of ownership of the data should be reported to the appropriate authority(ies), as required by the applicable regulatory requirement(s).

5.5.11. The sponsor specific essential documents should be retained until at least 2 years after the last approval of a marketing application in an ICH region and until there are no pending or contemplated marketing applications in an ICH region or at least 2 years have elapsed since the formal discontinuation of clinical development of the investigational product. These documents should be retained for a longer period however if required by the applicable regulatory requirement(s) or if needed by the sponsor. *TGA comment: The TGA requires records to be retained by the sponsor for 15 years following the completion of a clinical trial. However, in Australia, the overriding consideration for sponsors with respect to record retention is the issue of product liability and the potential need for sponsors of products to produce records at any time during, and possibly beyond, the life of a product in the event of a claim against the sponsor as a result of an adverse outcome associated with the use of the product.

[125] http://www.ich.org/products/guidelines/efficacy/efficacy-single/article/good-clinical-practice.html

[126] https://www.tga.gov.au/sites/default/files/ich13595an.pdf

5.5.12. The sponsor should inform the investigator(s)/institution(s) in writing of the need for record retention and should notify the investigator(s)/institution(s) in writing when the trial related records are no longer needed.

APPENDIX FOUR

Mandatory requirements for entities in undertaking Commonwealth Government Business as outlined in the Protective Security Policy Framework (PSPF):

GOV 1: Agencies must provide all staff, including contractors, with sufficient information and security awareness training to ensure they are aware of, and meet the requirements of the PSPF.

GOV 2: To fulfil their security obligations, agencies must appoint: (a) a member of the Senior Executive Service as the security executive, responsible for the agency protective security policy and oversight of protective security practices (b) an agency security adviser (ASA) responsible for the day-to-day performance of protective security functions, and (c) an information technology security adviser (ITSA) to advise senior management on the security of the agency's Information Communications Technology (ICT) systems.

GOV-3: Agencies must ensure that the ASA and ITSA have detailed knowledge of agency-specific protective security policy, protocols and mandatory protective security requirements in order to fulfil their protective security responsibilities.

GOV-4: Agencies must prepare a security plan to manage their security risks. The security plan must be updated or revised every two years or sooner when changes in risks and the agency's operating environment dictate.

GOV-5: Agencies must develop their own set of protective security policies and procedures to meet their specific business needs.

GOV-6: Agencies must adopt a risk management approach to cover all areas of protective security activity across their organisation, in accordance with the Australian Standards AS/NZS ISO 31000:2009 Risk management—Principles and guidelines and HB 167:2006 Security risk management.

GOV-7: For internal audit and reporting, agencies must: (a) undertake an annual security assessment against the mandatory requirements detailed within the PSPF, and (b) report their compliance with the mandatory requirements to the relevant portfolio Minister. The report must: contain a declaration of compliance by the agency head, and state any areas of non-compliance, including details on measures taken to lessen identified risks. In addition to their portfolio Minister, agencies must send a copy of their annual report on compliance with the mandatory requirements to: the Secretary, Attorney-General's Department, and the Auditor-General. Agencies must also advise any non-compliance with mandatory requirements to: (a) the Director, Australian Signals Directorate for matters relating to the Australian Government Information Security Manual (ISM) (b) the Director-General, Australian Security Intelligence Organisation for matters relating to national security, and (c) the heads of any agencies whose people, information or assets may be affected by the non- compliance.

GOV-8: Agencies must ensure investigators are appropriately trained and have in place procedures for reporting and investigating security incidents and taking corrective action, in accordance with the provisions of: (a) Australian Government protective security governance guidelines—Reporting incidents and conducting security investigations, and/or (b) The Australian Government Investigations Standards.

GOV- 9: Agencies must give all employees, including contractors, guidance on Sections 70 and 79 of the Crimes Act 1914, Section 91.1 of the Criminal Code 1995, the Freedom of Information Act 1982 and the Information Privacy Principles contained in the Privacy Act 1988 including how this legislation relates to their role.

GOV-10: Agencies must adhere to any provisions concerning the security of people, information and assets contained in multilateral or bilateral agreements and arrangements to which Australia is a party.

GOV-11: Agencies must establish a business continuity management program to provide for the continued availability of critical services and assets, and of other services and assets when warranted by a threat and risk assessment.

GOV-12: Agencies must ensure the contracted service provider complies with the requirements of this policy and any protective security protocols.

GOV-13: Agencies must comply with section 10 of the Public Governance, Performance and Accountability Rule 2014 and the Commonwealth Fraud Control Policy.

Personnel security

PERSEC 1: Agencies must ensure that their personnel who access Australian Government resources (people, information and assets): are eligible to have access, have had their identity established, are suitable to have access, and agree to comply with the Government's policies, standards, protocols and guidelines that safeguard the agency's resources from harm.
PERSEC 2: Agencies must have policies and procedures to assess and manage the ongoing suitability for employment of their personnel.
PERSEC 3: Agencies must identify, record and review positions that require a security clearance and the level of clearance required.
PERSEC 4: Agencies must ensure their personnel with ongoing access to Australian Government security classified resources hold a security clearance at the appropriate level, sponsored by an Australian Government agency.
PERSEC 5: Before issuing an eligibility waiver (citizenship or checkable background) and prior to requesting an Australian Government security clearance an agency must: justify an exceptional business requirement, conduct and document a risk assessment, define the period covered by the waiver (which cannot be open-ended), gain agreement from the clearance applicant to meet the conditions of the waiver, and consult with the vetting agency.
PERSEC 6: Agencies, other than authorised vetting agencies, must use the Australian Government Security Vetting Agency (AGSVA) to conduct initial vetting and reviews.
PERSEC 7: Agencies must establish, implement and maintain security clearance policies and procedures for clearance maintenance in their agencies.
PERSEC 8: Agencies and vetting agencies must share information that may impact on an individual's ongoing suitability to hold an Australian Government security clearance.
PERSEC 9: Agencies must have separation policies and procedures for departing clearance holders, which includes a requirement to: inform vetting agencies when a clearance holder leaves agency employment or contract engagement, and Advise vetting agencies of any security concerns.

Information security

INFOSEC 1: Agency heads must provide clear direction on information security through the development and implementation of an agency information security policy, and address agency information security requirements as part of the agency security plan.
INFOSEC 2: Each agency must establish a framework to provide direction and coordinated management of information security. Frameworks must be appropriate to the level of security risks to the agency's information environment.
INFOSEC 3: Agencies must implement policies and procedures for the security classification and protective control of information assets (in electronic and paper-based formats) which match their value, importance and sensitivity.
INFOSEC 4: Agencies must document and implement operational procedures and measures to ensure information, ICT systems and network tasks are managed securely and consistently, in accordance with the level of required security. This includes implementing the mandatory 'Strategies to Mitigate Targeted Cyber Intrusions' as detailed in the Australian Government Information Security Manual.
INFOSEC 5: Agencies must have in place control measures based on business owner requirements and assessed/accepted risks for controlling access to all information, ICT systems, networks (including remote access), infrastructures and applications. Agency access control rules must be consistent with agency business requirements and information classification as well as legal obligations.
INFOSEC 6: Agencies must have in place security measures during all stages of ICT system development, as well as when new ICT systems are implemented into the operational environment. Such measures must match the assessed security risk of the information holdings contained within, or passing across, ICT networks infrastructures and applications.
INFOSEC 7: Agencies must ensure that agency information security measures for all information processes, ICT systems and infrastructure adhere to any legislative or regulatory obligations under which the agency operates.

Physical security

PHYSEC 1: Agency heads must provide clear direction on physical security through the development and implementation of an agency physical security policy, and address agency physical security requirements as part of the agency security plan.
PHYSEC 2: Agencies must have in place policies and procedures to: identify, protect and support employees under threat of violence, based on a threat and risk assessment of specific situations. In certain cases, agencies may have to extend protection and support to family members and others,; report incidents to management, human resources, security and law enforcement authorities, as appropriate; provide information, training and counselling to employees, and; maintain thorough records and statements on reported incidents.
PHYSEC 3: Agencies must ensure they fully integrate protective security early in the process of planning, selecting, designing and modifying their facilities.
PHYSEC 4: Agencies must ensure that any proposed physical security measure or activity does not breach relevant employer work health and safety obligations.
PHYSEC 5: Agencies must show a duty of care for the physical safety of those members of the public interacting directly with the Australian Government. Where an agency's function involves providing services, the agency must ensure that clients can transact with the Australian Government with confidence about their physical wellbeing.
PHYSEC 6: Agencies must implement a level of physical security measures that minimises or removes the risk of information and ICT equipment being made inoperable or inaccessible, or being accessed, used or removed without appropriate authorisation.
PHYSEC 7: Agencies must develop plans and procedures to move up to heightened security levels in case of emergency and increased threat. The Australian Government may direct its agencies to implement heightened security levels.

APPENDIX FIVE

Controls as outlined in the National Security Directorate's Information Security Manual, which apply to outsourcing of data storage and processing of data from a relevant Australian Government agency in a public cloud, or offshore facilities.

Control:1376. Unclassified information that is not considered publicly releasable must not be stored or processed in public cloud or offshore ICT arrangements unless it meets the requirements outlined in the Attorney–General's Department's Australian Government Policy and Risk management guidelines for the storage and processing of Australian Government information in outsourced or offshore ICT arrangements.

Control:1377. Personal information as defined by the Privacy Act 1988 must not be stored or processed in public cloud or offshore ICT arrangements unless it meets the requirements outlined in the Attorney–General's Department's Australian Government Policy and Risk management guidelines for the storage and processing of Australian Government information in outsourced or offshore ICT arrangements.

Control:0873. Service providers' systems storing or processing Australian government information must be located in Australia.

Control:1378. Security classified information must not be stored or processed in a public cloud arrangement, unless the handling requirements have been appropriately downgraded as per the Cryptography and other related chapters of the ISM.

Control:1073. Agencies must ensure service providers seek their approval before allowing information to leave or be accessed from outside Australian borders.

Control:0872. Systems used by service providers for the provision of information technology services and functions must be accredited to the same minimum standard as the sponsoring agency's systems.

Control:1210. Agencies should assess the information security risks of using cloud computing services against ASD's Cloud Computing Security Considerations document.

Control:0744. Service providers should provide a single point of contact who will act as an equivalent to an ITSM.

APPENDIX SIX

Controls as outlined in the NIH Security Best Practices for Controlled-Access Data Subject to the NIH Genomic Data Sharing (GDS) Policy[127], which applies to data storage and processing of human derived genomic data arising from NIH-funded research.

Information for IT Professionals

1.Local Infrastructure Guidance

1.1 General Information Security Guidelines

- When using local infrastructure, make sure these files are never exposed to the Internet with the exception of such connections as are required to download data from source repositories. Infrastructure should be behind local and/or institutional firewalls that block access from outside of the institution. For cloud infrastructure, investigators must restrict external access to instances and storage under the investigator's control (see section on cloud computing for more details).
- Data must never be posted on servers in any fashion that will make them publically accessible, such as an investigator's (or institution's) website, because the files can be "discovered" by Internet search engines, e.g., Google, Bing.
- Institutions must not set up web or other electronic services that host data publicly, or that provide access to other individuals that are not listed on the Data Use Request even if those individuals have access to the same dbGaP data. Providing such access requires that an organization be an NIH Trusted Partner, with different requirements above and beyond those required for access to NIH controlled data.
- Utilize strong authentication technology for access control. Two factor authentication technologies (smart cards, hard or soft token, etc.) are preferred. When using single factor passwords, set policies that mandate the following requirements:

  o Minimum length of 12 characters
  o Does not contain user names, real names or company names
  o Does not contain a complete dictionary word
  o Contains characters from each of the following groups: lowercase letters, uppercase letters, numerals, and special characters
  o Passwords should expire every 120 days or at the rate required by institutional policies, whichever is more frequent.

- Avoid allowing users to place controlled access data on mobile devices (e.g. laptops, smartphones, tablets, mp3 players) or removable media such as USB thumb drives (except where such media are used as backups and follow appropriate physical security controls). If data must be placed on mobile devices, it must be encrypted. NIH recommends the use of NIST validated encryption technologies
- Keep all software patches up-to-date.

[127] http://www.ncbi.nlm.nih.gov/projects/gap/pdf/dbgap_2b_security_procedures.pdf

1.2 Physical Security Guidelines

- Data that are in hard copy or reside on portable media, e.g., on a USB stick, CD, flash drive or laptop should be treated as though it were cash, with appropriate controls in place. Such media must be encrypted and stored in a secured in a locked facility with access granted to the minimum number of individuals required to efficiently carry out research.
- Restrict physical access to all servers, network hardware, storage arrays, firewalls and backup media only to those that are required for efficient operations.
- Log access to secure facilities, ideally with electronic authentication.

1.3 Controls for Servers

- Keep servers from being accessible directly from the Internet, (i.e. must be behind a firewall or not connected to a larger network) and disable unnecessary services. It is better to begin with a server image that disables all non-essential services and restore those that are needed than to start with a full-featured image and disable unnecessary services.
- Enforce principle of Least Privilege to ensure that individuals and/or processes grant only the rights and permissions to perform their assigned tasks and functions, but no more.
- Secure controlled-access genomic and phenotypic data on the systems from other users (restrict directory permissions to only the owner and group) and if exported via file sharing, ensure limited access to remote systems.
- If accessing systems remotely, use encrypted data access (such as Secure Shell (SSH) or Virtual Private Network (VPN)). It is preferred to use a tool such as Remote Desktop (RDP), X-windows or Virtual Network Computing (VNC) that does not permit copying of data and provides "View only" support.
- If data is used on multiple systems (such as a compute cluster), ensure that data access policies are retained throughout the processing of the data on all the other systems. If data is cached on local systems, directory protection must be kept, and data must be removed when processing is complete. Requesting investigators must meet the spirit and intent of these protection requirements to ensure a secure environment 24 hours a day for the period of the agreement.

1.4 Source Data and Control of Copies of Data

- Approved users must retain the original version of the encrypted data, track all copies or extracts and ensure that the information is not divulged to anyone except authorized staff members at the institution. NIH therefore recommends ensuring careful control of physical copies of data and providing appropriate logging on machines where such data is resident.
- As collaborating investigators from other institutions must submit an independent DAR and be approved by NIH to access to the data, restrict outbound access from devices that host controlled access data.

1.5 Destruction of Data

- Data downloaded from NIH-designated data repositories must be destroyed if they are no longer needed or used, or if the project is to be terminated and closed-out in the dbGaP Authorized Access System. Delete all data for the project from storage, virtual and physical machines, databases, and random access archives (i.e., archival technology that allows for deletion of specified records within the context of media containing multiple records).
- Investigators and Institutions may retain only encrypted copies of the minimum data necessary at their institution to comply with institutional scientific data retention policy and any data stored on temporary backup media as are required to maintain the integrity of the institution's data protection program. Ideally, the data will exist on backup media that is not used by other

projects and can therefore be destroyed or erased without impacting other users/tenants. If retaining the data on separate backup media is not possible, as will be the case with many users, the media may be retained for the standard media retention period but may not be recovered for any purpose without a new Data Access Request approved by the NIH. Retained data should be deleted at the appropriate time, according to institutional policies.

- Shred hard copies and CD ROMs or other non-reusable physical media.
- Delete electronic files securely. For personal computers, the minimum would involve deleting files and emptying the recycle bin or equivalent with equivalent procedures for servers. Optimally, use a secure method that performs a delete and overwrite of the physical media that was used to store the files.
- Ensure that backups are reused (data deleted) and any archive copies are also destroyed.
- Destroy media according to (NIST) Guidelines for Information Media Sanitization (http://csrc.nist.gov/publications/PubsSPs.html)

2. Additional Guidance for Cloud Computing

Institutions that wish to use cloud computing must work with their cloud service provider to devise an appropriate security plan that meets the general dbGaP Information Security Best Practices as well as these additional requirements that derive from the nature of multi-tenant clouds with default access to the internet. Please refer to the specific cloud service provider for methods, processes and procedures for working with controlled-access data subject to the GDS Policy in the cloud.

2.1 General Cloud Computing Guidelines

- Whenever possible, use end-to-end encryption for network traffic. For example, use Hypertext Transfer Protocol (HTTPS) sessions between you and your instance. Ensure that your service uses only valid and up-to-date certificates.
- Encrypt data at rest with a user's own keys. SRA-toolkit includes this feature; other software providers offer tools to meet this requirement.
- Use security groups and firewalls to control inbound traffic access to your instance. Ensure that your security profile is configured to allow access only to the minimum set of ports required to provide necessary functionality for your services and limit access to specific networks or hosts. In addition, allow administrative access only to the minimum set of ports and source IP address ranges necessary.
- Be aware of the top 10 vulnerabilities for web applications and build your applications accordingly. To learn more, visit Open Web Application Security Project (OWASP) - Top 10 Web Application Security Risks. When new Internet vulnerabilities are discovered, promptly update any web applications included in your Virtual Machine (VM) images. Examples of resources that include this information are SecurityFocus and the NIST National Vulnerability Database.
- Review the Access Control Lists (ACLs), permissions, and security perimeter to ensure consistent definition.

2.2 Audit and Accountability

- Ensure that account access is logged along with access controls and file access and this information is reviewed by the investigator on regular basis to ensure continued secure access.

- Ensure that data is accessible only to those approved for access, and controls for changing that access are retained by the investigator who submitted the DAR and the appropriate IT staff. A mechanism for monitoring and notification needs to be in place to monitor changes in permission changes.

2.3 Image Specific Security

- Ensure images do not contain any known vulnerabilities, malware, or viruses. A number of tools are available for scanning the software, such as Chkrootkit, rkhunter, OpenVAS and Nessus.
- Ensure that Linux-based Images lock/disable root login and allow only sudo access. Additionally, root password must not be null or blank.
- Ensure that images allow end-users with OS-level administration capabilities to allow for compliance requirements, vulnerability updates, and log file access. For Linux-based Images, this is normally through SSH, and for Windows-based virtual machine images, this is normally through RDP.

3. Best Practices for Specific Cloud Service Providers:

Examples of cloud service provider best practices are provided in the links below, links to the best practices of additional cloud service providers will be periodically appended to this document when they become available. Please be aware that these are provided for convenience only, and do not imply endorsement by the NIH or the United States Government for any of these services, nor does the government guarantee that these links lead to the most current version of these best practices. NIH recommends that investigators consult with their cloud service provider to ensure that they are using the most up to date best practice documents.

Amazon Web Services:

- http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/AMIs.html
- http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ec2-best-practices.html
- http://aws.amazon.com/documentation/ec2/

Google Cloud Platforms:

- https://cloud.google.com/developers/articles/best-practices-for-configuring-permissions-on-gcp

4. Others Sources of Information for Cloud Best Practices:

Examples of cloud best practices from organizations that leverage the cloud are provided in the link below. Links to additional documentation will be periodically appended to this document when they become available. Please be aware that these are provided for convenience only, and do not imply endorsement by the NIH or the United States Government for any of these services, nor does the government guarantee that these links lead to the most current version of these best practices. NIH recommends that investigators consult with these organizations to ensure that they are using the most up to date best practice documents.

DNAnexus: https://dnanexus.com/papers/Compliance_White_Paper.pdf