

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

Department of Health

Independent Review of 10% Sample MBS and PBS Data Decryption

October 2016

Australian Government
Department of Health

protiviti®
Risk & Business Consulting.
Internal Audit.

Australian Government
Department of Health

protiviti
Risk & Business Consulting.
Internal Audit.

Review Sign-off

Timeline

Milestone	Date Completed
Planning	12 September 2016
Agreed Review Plan	12 September 2016
Fieldwork Complete	23 September 2016
Exit meeting	23 September 2016
Draft Review Report to Management	28 September 2016
Management Comments	28 October 2016
Final Review Report	2 November 2016
Report Tabled at Audit Committee	

Sign-off

I acknowledge the findings and recommendations enclosed in this report and undertake to have the agreed actions completed within the identified time frame

.....	.....
October 2016	October 2016

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

Australian Government
Department of Health

protiviti
Risk & Business Consulting.
Internal Audit.

Table of Contents

1. Executive Summary 4
2. Background, Scope and Approach 6
3. Chronology 9
4. Risk Analysis 13
5. Positive Practices 15
5.1. Consistency with the Government's espoused principles for data access 15
6. Findings and Recommendations 16
6.1. Encryption method applied to the MBS data in the dataset had not been tested for its purpose 16
6.2. Data governance arrangements at Health are in a state of change, reducing clarity of requirements and expectations in data management and release 18
6.3. Formalities for approval of release of the dataset were not appropriately robust 20
6.4. Project and risk management disciplines associated with the development and release of the dataset would have benefited from increased rigour 21
7. Observations 23
7.1. Consistency of standards and testing in central data release forums 23
Appendix 1: Personnel Consulted 24
Appendix 2: Documents Reviewed 25

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

1. Executive Summary

1.1. Background

On 1 August 2016, a sample of de-identified and encrypted Medicare Benefits Scheme (MBS) and Pharmaceutical Benefits Scheme (PBS) patient level claims data was publicly released onto the website www.data.gov.au. This followed over a year of work to assemble, test and confirm the dataset for release. A suite of confidentiality measures had been applied to the data to protect the identity of patients and providers.

On 8 September 2016, the Australian Bureau of Statistics (ABS) was advised by researchers from the University of Melbourne that they had been able to decrypt the Service Provider Number from the 2014 MBS dataset contained within the broader dataset. After being notified, the Department of Health confirmed that the sample data had in fact been decrypted.

1.2. Overall Assessment

This independent review found that the Department's initiatives to publicly release this data were consistent with the principles of the Government regarding enabling broader access to information for research, policy-making and decision-making.

Scope Item 1: Understand and Define Risk Exposure to the Department from the Incident

The likelihood of public re-identification of the data in the dataset is possible given for some components of the dataset it has been proven to be possible, and this will likely become public. There are varying levels of likelihood depending on attributes of the dataset.

The primary consequence of this public re-identification is reputational, and the reputational consequence is likely to be high, given the sensitivity of the data and the level of public expectation. There is also some risk of legislative breach being an outcome if there is public reidentification; however this is likely to be contested.

Scope Item 2: Define root cause of the unexpected disclosure of sensitive information to the Department

This independent review has found that there were four primary contributing factors that led to the unexpected re-identification of sensitive information from the publicly released linkable 10% MBS and PBS Sample dataset:

- The encryption methods utilised for the Service Provider Number had not been tested to confirm fitness for purpose given the unique nature of data and the more public release of the data.
- Data governance arrangements at Health were in a state of change, to a more controlled model of management, reducing clarity at the time of requirements and expectations in data management and release.
- Formalities for approval of release of the dataset were not appropriately robust.
- Project and risk management disciplines associated with the development and release of the dataset would have benefited from increased rigour.

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

1.3. Summary of Findings and Recommendations

Report Section	Finding	Recommendation	Risk Rating
6.1	Encryption method applied to the MBS data in the dataset had not been tested for its purpose	Data release policies for the Department should include a requirement for specific testing of confidentialisation measures for all public releases of unit-level or patient level data to confirm that they are appropriate to meet the Department's risk appetite for data release. Such testing should be undertaken by qualified personnel with relevant skills including (as required) cryptography and data analysis.	High
6.2	Data governance arrangements at Health are in a state of change, reducing clarity of requirements and expectations in data management and release	The Department should continue and accelerate its implementation of centralisation of certain aspects of data governance in order to increase consistency of approach and decision making for data access. Further, the central governance body should develop a comprehensive set of guidance for line areas that hold data to ensure consistent approaches.	High
6.3	Formalities for approval of release of the dataset were not appropriately robust	Data release policies for the Department should include a requirement for all public releases of unit-level or patient level data to be reviewed and approved by the responsible Deputy Secretary.	Medium
6.4	Project and risk management disciplines associated with the development and release of the dataset would have benefited from increased rigour	Data release policies for the Department should include a requirement for all public releases of unit-level or patient level data to apply Departmental project management disciplines and risk assessment and management disciplines consistent with those applied for data integration projects.	Medium

1.4. Overall Management Comments

.

1.5. Limitation

This report is intended solely for the use by the Department of Health, and should not be distributed to any third party without the consent of Protiviti, which will not be unreasonably withheld.

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

2. Background, Scope and Approach

2.1. Background

The Department of Health is the custodian of extensive data holdings relating to health care and health care costs in Australia. This is an important resource to inform decision making by Governments, providers, pharmaceutical companies and other relevant entities.

The *Public Sector Data Management* report issued by the Department of the Prime Minister and Cabinet in July 2015, noted the following:

"Currently, the Commonwealth's capacity to fully derive value from public sector data is constrained by competing priorities and the lack of an overarching strategy.

- There is no clear mandate for the Commonwealth to use and release public sector data.
- There are barriers (perceived and real) to sharing data within the Commonwealth and with jurisdictions to improve policy and service delivery.
- The APS lacks sufficient incentives, skills and organisational arrangements to capitalise on its data.
- The Commonwealth does not have a strong culture of publishing data to foster economic opportunities."

As a consequence, a series of recommendations were made, which the Government has committed to implement. In total, 15 recommendations were made, which can be grouped as follows:

1. Build confidence and momentum (six recommendations) to be implemented over 1-6 months.
2. Systematise the use and release of public sector data (nine recommendations) to be implemented over 1-18 months.

On 1 August 2016, a sample of de-identified and encrypted Medicare Benefits Scheme (MBS) and Pharmaceutical Benefits Scheme (PBS) patient level claims data was publicly released onto the website www.data.gov.au. This followed over a year of work to assemble, test and confirm the dataset for release. A suite of confidentiality measures had been applied to the data to protect the identity of patients and providers.

The dataset comprised linkable MBS and PBS claims data for a random selection of 10% of the Australian patients, including MBS data from 1984 and PBS data from 2003 up to 2014. The data release represents approximately 3 billion lines of data relating to approximately 2.5 million Australians. The intention of the Department is to update this dataset annually as additional data becomes available.

The release of the data is expected to provide researchers with the opportunity to undertake research without having to bear the costs and delays associated with requests to government for unit record linked data. It would facilitate longitudinal analysis of patient progression through the Australian health system and provide the basis for cohort analysis.

The release of the dataset was an extension of a long-standing arrangement that had been established by the Department of Human Services (DHS) whereby a 10% sample of de-identified PBS (only) claims data had been released to a selection of approved research organisations on a contracted basis, with certain restrictions on how the data could be used by those approved research organisations. This arrangement had been in place since 2005. The structure and methodology of the sampling and confidentialisation of the data for the linkable MBS and PBS claims data that was publicly released was based on the approach that had been taken for the PBS (only) data.

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

On 8 September 2016, the Australian Bureau of Statistics (ABS) was advised by researchers from the University of Melbourne that they had been able to decrypt the Service Provider Number from the 2014 MBS dataset contained within the broader dataset. After being notified, the Department of Health confirmed that the sample data had in fact been decrypted. The Department immediately requested the suspension of access to the dataset on www.data.gov.au.

It is understood that approximately 1,500 downloads of all or part of the dataset had been made prior to the suspension of access.

2.2. Objective

Following the identification of this issue, the Department of Health commissioned an independent review to assess and document the circumstances associated with the recent event of decryption and re-identification of certain MBS and PBS data and assess and document the controls that were undertaken to manage the risk of privacy breach or unexpected disclosure of the sensitive information with a view to:

1. Understand and define exposure to the Department;
2. Define root cause of the unexpected disclosure of sensitive information to the Department; and
3. Identify recommendations for the future to ensure the risk is managed consistent with the Department's risk appetite.

2.3. Scope

All aspects of the selection, assembly, management (including encryption and perturbation) and subsequent re-identification of data associated with the reported incident are included within the scope of the independent review.

Scope Limitations

The independent review does not:

- assess response strategies once the incident was identified;
- assess the Department's data management and publication strategy, but rather has assessed the processes and decision making in the case of the data release in question (consistent with that strategy);
- test the encryption and perturbation mechanisms;
- provide legal opinion in connection with compliance with legislation, although it did consider legal advice sought by the Department;
- make any judgement on the accuracy or otherwise of the content of the dataset.

The assessments made during this independent review have been provided in good faith and in the belief that such statements and opinions are not false or misleading. Due to the limited duration of the independent review, Protiviti has relied on information that was provided by Health and other stakeholders. (A full list of officers consulted in the independent review is included in Appendix 2 to this report.) Protiviti does not express an opinion as to whether the information supplied is accurate and no warranty of accuracy or reliability will be given. Furthermore, we have not implied and it should not be construed that we have verified the information provided to us, or that our enquiries could reveal any matter that a more extensive examination might disclose.

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

2.4. Methodology

The approach to the independent review is defined below.

- Through consultation with data owners, stakeholders including DHS and ABS, and personnel who identified the ability to decrypt the data, and consideration of existing risk assessment documentation associated with the released data, document the risk associated with the potential decryption of the released data (considering both likelihood and consequence of decryption and identification);
- Through consultation with the University of Melbourne, identify how the decryption was identified and implemented, with a view to assisting in identification of root cause;
- Identify all aspects of the management of the project to select, assemble and, manage (including decryption and perturbation) the data, with a focus on identification of key controls to manage the risk. Key controls may relate to (but not be limited by): due diligence regarding release project, selection of encryption and perturbation techniques, testing of encryption and perturbation, information authorisation, quality assurance and risk assurance processes, and decision making and authorisation for release;
- Identify the root causes of the unexpected decryption of information;
- Through consideration of root cause and controls, identify:
    - Whether any weakness in control contributed to the occurrence of the event, and whether this is outside the Department's documented risk appetite.
    - What (if any) recommendations should be implemented in the future to manage the risk to a level consistent with the Department's risk appetite.
- Make comment on records retained to support the overall project and management of risk associated with the data release.

2.5. Review Team

Team Member	Position
s47F	Managing Director

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

3. Chronology

The following table presents a chronology of relevant events leading up to the identification of decryption of the data

Date	Event
1990's	DHS records indicate that the method of scrambling/encryption of data used for the 10% PBS Sample dataset was potentially being used by the then Health Insurance Commission and was later adopted for use in the 10% PBS Sample dataset. (Based on advice from DHS.)
Mid 2005	DHS was approached by a research organisation requesting a PBS sample data. From this the 10% PBS sample dataset was created and the research organisation contracted on a fee-for-service basis to provide quarterly 10% sample datasets. This arrangement has continued to operate ever since, with new approved research organisations being included in the arrangement.
16 July 2010	Minister for Finance makes *Declaration of Open Government*.
May 2011	Office of the Australian Information Commissioner releases *Principles on open public sector information*.
19 July 2011	*Enterprise Data Management – High Level Principles* released by the Chief Information and Knowledge Officer of the Department of Health.
March 2012	DHS, in consultation with Health, proposed a review of the existing 10% PBS sample dataset provision arrangements (at that time, to six private sector organisations), including review of the contract and the arrangements. It was agreed to utilise the assistance of an outposted officer from ABS in Health. Review was part of a broader workplan which included a "review of the 10% PBS dataset in its current form", including "is there an ability to identify consumers, providers or pharmacies within a three month period or over a greater length of time". The workplan also included broader work in development of guidelines and review of fee-for-service contract." The intent was the work was to be completed within one year.
25 May 2012	Terms of Reference of Department of Health Data Governance Council is approved.
14 June 2012	*Enterprise Data Governance Framework* re-released (prior version April 2012), which provides guidance on data governance and management. (The framework includes limited detailed instruction regarding data release procedures.)
30 August 2012	Meeting between representatives of Health and ABS to progress ABS review. It was agreed at this meeting that the project would be split into two phases.
27 November 2012	ABS (Methodology and Data Management Division) issued quote for conduct of the review
23 May 2013	Memorandum of Understanding between ABS, Health and DHS in agreeing the *Review of the 10% PBS Sample File*, including definition of scope which includes two main components (summary only): (1) Is there an ability to identify consumers, providers or pharmacies in the current file?; and (2) Is the 10% PBS sample file fit for purpose?.
28 August 2013	Australian Bureau of Statistics issues *Review of the 10% PBS Sample File – Phase 1 Report* following their review of the previously released 10% PBS Sample dataset.
11 November 2013	Health accepts findings of *Review of the 10% PBS Sample File – Phase 1 Report* and confirms requirement to continue to *Review of the 10% PBS Sample File – Phase 2 Report.*
29 July 2014	Australian Bureau of Statistics issues *Review of the 10% PBS Sample File – Phase 2*

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

Date	Event
	Report following their review of the previously released 10% PBS Sample dataset.
29 January 2015	Meeting between Health (s47E(c), s47F) and ABS representatives to discuss and scope a potential review of the disclosure risks associated with the 10% MBS Sample dataset (based on similar review done for 10% PBS Sample dataset). This review was never commissioned by the Department of Health.
July 2015	Discussion Paper *Data – getting it, governing it, making it available, strategic analytics and research* issued by s47E(c), s47F Special Adviser proposing new arrangements for data governance and management at Health. These new arrangements involve a centralisation of certain aspects of data governance.
24 July 2015	LSB issues advice (LEX 23633) to Therapeutic Goods Administration *Linking MBS and PBS data*. Advice responds to the following questions: (1) Can you advice how the TGA can link, work with and store MBS and PBS data for the Project?; (2) Is a Public Interest Certificate required?; (3) can TGA staff or their contractors link the data – is the Secretary's approval required for this activity?; (4) For how long can the linked file be stored on TGA systems"; and (5) How should the separate MBS and PBS files be stored?
25 August 2015	Issue of latest version of Department's *Data Access and Release Policy*. The policy defines principles for access and release of "identifiable" and "de-identified" data but does not make statement on what constitutes de-identified.
14 October 2015	Data Governance Council (DGC) meeting at which the proposed release of the 10% MBS/PBS Sample dataset. Minutes of the meeting note: • "Data will be put through a confidentialisation process that the department is confident will make re-identification of the owner of the data a practical impossibility." • "Health is confident the re-identification is near impossible as the information held is of an administrative nature and doesn't include diagnostic data." • "The DGC agreed to the public release of linkable MBS, PBS and public hospital 10% sample data set, noting the release proposed will be presented to the Executive for final approval."
28 October 2015	*Bilateral Head Agreement 2015-18* between Secretary of DHS and Secretary of Health signed. This is an umbrella agreement on broad ranging working arrangements.
3 November 2015	Advice from LSB to MBD: "We have reviewed the methods being proposed to ensure no individual is able to be identified as set out in the Data Governance Council agenda paper. Based on this review and your statement that you are satisfied that once those actions are taken no individual will be able to be re-identified, we are of the opinion that the linkable 10% sample of MBS and PBS data will be de-identified and therefore not information 'with respect to the affairs' of a person. Therefore the secrecy provisions in s130 of the Health Insurance Act 1972 and s135A of the National Health Act 1953 will not apply to prohibit disclosure. The information will also not be 'personal information' for the purposes of s6 of the Privacy Act 1988."
November 2015	Draft[1] Information Brief to Deputy Secretary advising of planned MBS data releases, including reference to10% linkable sample. Notes that "careful attention to confidentiality requirements and other data sensitivities".
1 December 2015	Australian Government Solicitor issues advice (15194245) *De-identified MBS and PBS data* to the Department of Health (LSB). Advice responds to the following questions: (1) In what circumstances will a court find that a 'de-identified" MBS and PBS data is

[1] The original and final Information Brief has not been able to be obtained by this independent review.

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

<table>
<tr><th>Date</th><th>Event</th></tr>
<tr><td></td><td>not ‘information relating to the affairs of [a] person’ such that the prohibitions in s135(1) of the National Health Act 1953 and s130(1) of the Health Insurance Act 1973 will not apply?; (2) In what circumstances, will a court find that this data is not information relating to an individual and therefore not subject to the guidelines made under s135AA of the National Health Act 1953?; (3) If a court finds that the MBS and PBS data is information relation to the affairs of [a] person … what sanctions could apply to the officer in handling the data in breach of the prohibitions in s135A(1) of the National Health Act 1953 and s130(1) of the Health Insurance Act 1973”; and (4) Is ‘information relating to the affairs of [a] person’ the same as ‘personal information’ for the purposes of the Privacy Act 1988?</td></tr>
<tr><td>3 December 2015</td><td>Department of Prime Minister and Cabinet issue the Public Sector Data Management Report advocating better utilisation of data. Report recommendations include:
<table>
<tr><th>1-6 months: Build confidence and momentum</th><th>1-18 months: Systematise the use and release of public sector data</th></tr>
<tr><td>1. Secretary of PM&C signals data is a priority
2. Commission several high value projects
3. Build external partnerships
4. Publish readily available non-sensitive datasets
5. Build data and analytics capability
6. PM&C coordinates projects and progresses data policies and governance</td><td>7. Implement a data policy framework
8. Build and maintain trust
9. Establish a trusted-access model for sharing integrated data
10. Create and publish a searchable whole-of-government data catalogue
11. Develop a Commonwealth Government high-value dataset framework
12. Publish data management standards
13. Establish a consistent and transparent approach to user charging.
14. Create a legislative environment that supports data use while maintaining privacy.
15. Promote innovation in public administration.</td></tr>
</table></td></tr>
<tr><td>7 December 2015</td><td>Prime Minister issues Australian Government Public Data Policy Statement. (Statement notes that Australian Government entities will “uphold the highest standards of security and privacy for the individual”.</td></tr>
<tr><td>15 April 2016</td><td>Deputy Secretary acknowledges information brief noting pre-launch release of subset of 10% MBS/PBS Sample dataset at PharmHack 2016.
Brief notes: “The design of the datasets has been rigorously tested both within the Department and vetted by the Australian Bureau of Statistics. The datasets have also undergone extensive internal testing. A suite of confidentiality measures including encryption, perturbation and exclusion of rare events has been applied to safeguard personal health information and ensure that patients and providers cannot be identified.”</td></tr>
<tr><td>16-17 April 2016</td><td>Pre-launch release of subset of 10% MBS/PBS Sample dataset at PharmHack 2016.</td></tr>
<tr><td>May 2016</td><td>Draft High Level Principles for the Release of Health Information and Data released.</td></tr>
<tr><td>May 2016</td><td>Draft Protocol for the External Release of Health Information and Data released.</td></tr>
<tr><td>25 May 2016</td><td>Half-day workshop between Health and DHS to establish a “Protocol for the External</td></tr>
</table>

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

Date	Event
	Release of Health Information and Data". (Draft Protocol at this time was draft with considerable gaps.).
2 June 2016	Feedback received on a subset of the 10% MBS/PBS Sample dataset, which was circulated to some contracted recipients of the previous 10% PBS Sample dataset. Comments focussed on usability, not on security.
July 2016	Information brief released to Minister Ley, Assistant Minister Wyatt and Assistant Minister Gillespie informing them of the release of linkable de-identified 10% sample of Medicare and PBS claims data (MB16-001178). (Cleared by FAS and AS in MBD and PBD on 26 July 2016). Brief notes: "A suite of confidentiality measures including encryption, perturbation and exclusion of rare events has been applied to safeguard personal health information and ensure that patients and providers cannot be re-identified."
July 2016	Department of Prime Minister and Cabinet issues an Implementation Report arising from the Public Sector Data Governance Report advocating better utilisation of data.
1 August 2016	Linkable de-identified 10% sample of MBS and PBS data, including description of dataset, explanation of confidentiality measures and data adjustments and high level definition of encryption is published on www.data.gov.au.
8 September 2016	Draft Terms of Reference of proposed Data Governance and Analytics Committee is issued for comment.
8 September 2016	University of Melbourne representatives draft and deliver a report to the ABS describing the problem in the encryption of the Service Provider ID from the MBS 10% Sample dataset.
8 September 2016	ABS representative provides a copy of the University of Melbourne report to Health.

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

4. Risk Analysis

The data set includes two types of personal de-identified data relating to Australian citizens who have received medical services or pharmaceuticals (patients), and medical service providers who have received medical benefits under the MBS.

In the context of this review, the primary risk relates to the broader (public) access to information that can identify either of those types of personal information.

Re-identification can occur broadly through one of two means: (a) decryption of patient or service provider identifiers that have been encrypted using a variety of means; or (b) identifying one or several known records and using them as identifiers, or other known information outside the dataset, to re-identify other records relating to an individual or the entire population.

Likelihood of Decryption of Service Provider Identifier

The event that triggered this independent review indicates that the re-identification of service provider identifiers is at least possible. While it requires some knowledge of encryption, there is sufficient capability in society to suggest that re-identification is **possible**. Further, there are tools and information that can assist people to link a service provider identifier to personal or practice information about the service provider.

This possibility is reinforced by the fact that there are approximately 1,500 downloads of the dataset (in part or in full) that are now in circulation and uncontrolled.

Likelihood of Decryption of Patient Identifier

The Department has sought the assistance of an Australian Government agency with deep expertise in cryptography to see if they can re-identify the patient data from the dataset. At this time of this independent review, that agency has been unable to re-identify patient information from the encrypted identifier. While not absolute, this suggests that the likelihood of reidentification of patients through decryption is **low**.

Likelihood of Re-identification Through Data Linkage with Other Known or External Data

The Department has sought the assistance of Data 61 to quantify the likelihood of re-identification from the dataset. The results of this analysis identified that:

(a) Re-identification likelihood varies from very low (<0.1%) for attributes classified as commonly available, to significantly higher risks (>10%) (which can be classed as **possible** using the Department's risk assessment methodology for selected moderately available data attributes, and then to severe risk for rarely available attributes.
(b) The combination of attributes does increase the risk of re-identification; and
(c) The impact of high-identification risks is limited to a small population in the dataset (1.2% of individuals).

Consequence of Reidentification

In the context of the specific event that prompted this independent review, the re-identification of service providers from the data has a **moderate** consequence to the Department. The data set provides the opportunity, for identified service providers, to observe billing practices (including billing in excess of MBS rates) and to observe services provided. For some service providers that may provide visibility of provision of sensitive services which some service providers will not wish to be publicly available.

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

Alternatively, should it transpire that patients are re-identified from the dataset, this is likely to have a higher consequence to the Department as the patient confidentiality is the very reason that structures are in place to protect health record data.

From a legal perspective, reidentification of service provider or patient data raises a risk of breach of any of the Privacy Act, the National Health Act and the Health Insurance Act. The organisation has received legal advice that if data was released in a form that the Department genuinely believed it had been de-identified, and then is subsequently re-identified, this would not represent a breach. However, this independent review noted that there is some contention on this matter, as it is untested. It is likely that this legal principle may be challenged in the event of publication of re-identification of data. It is further noted that the Department's Legal Services Branch is aware of this risk and is preparing accordingly.

Ultimately the main immediate consequence of re-identification of this data is reputational damage for the Department. The Department is tasked to hold this data and is expected to do so carefully. The very re-identification of the data, and some of the circumstances surrounding the event, are likely to attract considerable criticism.

The independent review has been made aware that other jurisdictions are (or are planning to) release similar data for similar purposes. This may mitigate some of the consequence of the risk.

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

5. Positive Practices

5.1. Consistency with the Government's espoused principles for data access

The Department's initiatives to publicly release this data are consistent with the principles of the Government regarding enabling broader access to information for research, policy-making and decision-making. These principles are noted in:

- Minister for Finance's *Declaration of Open Government* (2010)
- Office of the Australian Information Commissioner *Principles on open public sector information* (May 2011)
- Department of the Prime Minister and Cabinet's *Public Sector Data Management Report* (2015); and
- Australian Government *Public Sector Data Policy Statement* (2015).

Further, feedback from researchers and representatives of the Department has noted that the information is useful and valuable to researchers to inform Government and marketplace activity in the provision of healthcare.

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

6. Findings and Recommendations

6.1. Encryption method applied to the MBS data in the dataset had not been tested for its purpose

Finding

While the review was advised that reliance on pre-existing methodologies for encryption and de-identification, and previous reviews of existing methodologies under different circumstances was anticipated at the time to be appropriate for the public release of the linkable MBS and PBS dataset, the extent of testing of fitness for purpose was insufficient to obtain comprehensive assurance given the unique nature of the data release.

Discussion

The confidentiality measures applied to patient and provider identifiers is outlined below:

s47E(d)

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

s47E(d)



Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

Risk Exposure

The absence of effective testing and confirmation of the strength of de-identification measures can be directly linked to the de-cryption by the University of Melbourne. Further, the absence of such testing and confirmation indicates poor risk management which may be noted when the event is analysed by stakeholders.

Recommendation 1	
Risk Rating	High
Data release policies for the Department should include a requirement for specific testing of confidentialisation measures for all public releases of unit-level or patient level data to confirm that they are appropriate to meet the Department's risk appetite for data release. Such testing should be undertaken by qualified personnel with relevant skills including (as required) cryptography and data analysis.	
Management Comments	
Agree	

[2] It is noted that this recommendation was implemented by Health.

New, more centralised arrangements are being implemented for the management of access to the department's data. This is to enable greatest possible use of the data whilst ensuring that privacy and confidentiality of data are preserved, and that consistent decisions are made with respect to data access requests. The new arrangements mean that all data requests will be centrally coordinated via a "single front door". This will ensure that all data requests are assessed using standardised criteria, and that a record is kept of who has received health data and the uses to which they have put it.

As part of this new process, a centralised data release assessment committee will be established to assess data requests to ensure that confidentiality is maintained and standards consistently applied across the Department. The assessment committee will be advised by a group of technical experts – including experts from external agencies with expertise in cryptography and data analysis – who will advise on data confidentiality and security. The data release assessment committee will report to the Department's new centralised Chief Data Steward. The committee will work under policies, procedures and standards agreed by the Department's new Data Governance and Analytics Committee which will oversee data management, analytics and release approaches across the Department.

Action Officer	Estimated Completion Date
	The centralised data governance and release framework will be submitted to the Data Governance and Analytics Committee and then to Executive for consideration for endorsement in December and implementation from January 2017.

6.2. Data governance arrangements at Health are in a state of change, reducing clarity of requirements and expectations in data management and release

Finding

At the time of the incident, the Department was transitioning to more controlled governance arrangements for data custody and release which are expected to drive improvements in control and consistency in decision making on data management and release. These were not in place to inform the management arrangements for the 10% MBS and PBS Sample data initiative.

Discussion

Over the past four years, there has been considerable change in the data management arrangements at Department of Health.

In June 2012, the Department issued its first *Enterprise Data Governance Framework*, as a component of the then Department of Health National Alignment Project. This framework was designed to be consistent with the globally recognised Data Management Body of Knowledge. It is a set of data governance principles, responsibilities and structures designed to ensure data is managed effectively and securely.

In August 2015, the Department issued its latest *Data Access and Release Policy* which defined a policy statement on data release and certain guidelines, including guidelines relating to "low-risk de-identified, confidentialised or non re-identifiable data". The guidelines note a focus on public accessibility of data, and that public accessibility and release should "maintain the security of data holdings and individual and organisational policy".

However there is no further guidance on Department standards on how to maintain security and confidentiality.

Since mid-2015, and the release of a discussion paper entitled *Data – getting it, governing it, making it available, strategic analysis and research*, there has been consideration of changes in structure regarding the responsibility for data management in the Department. These considerations remain ongoing at the time of this independent review.

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

This environment requires data custodians to make judgements on how and when to take action to comply with the *Enterprise Data Governance Framework* and related policies, in the absence of clarity on limits and tolerances or expectations on how to apply the policies. The data custodians for the MBS and PBS data in effect exercised their judgements in that context in the release of the 10% MBS and PBS Sample dataset.

The following represents a description of the anticipated arrangements (which were not in place at the time of the release of the 10% MBS and PBS Sample dataset:

- New, more centralised arrangements are being implemented for the management of access to the department's data. This is to enable the greatest possible use of the data whilst ensuring that privacy and confidentiality of data are preserved, and that consistent decisions are made with respect to data access requests.
- The new arrangements mean that all data requests will be centrally coordinated via a "single front door". This will make it easier for researchers and others to know how to apply for access, as well as allow us to monitor progress and ensure requests are assessed in an appropriate timeframe relative to the complexity of the request.
- The arrangements will also ensure that all data requests are assessed using standardised criteria, and that a record is kept of who has received health data and the uses to which they have put it.
- In addition, a new senior oversight committee has been formed, chaired at deputy level, which advances the work already completed by the department in implementing a data governance framework and risk management framework to the department's data management, including the management of data requests.
- These arrangements are part of a broader strategy to continuously improve the management of the department's data and analytics capability. The new "hub and spoke" model for data and analysis was implemented at the beginning of September, following Executive agreement in August.

It is understood that as part of the new arrangements, consistent central approaches to certain elements of data management and access are being considered. One of these is the "five safes" framework for access to sensitive data.

Risk Exposure

The absence of consistent data management disciplines likely contributed to the de-cryption by the University of Melbourne. Further, the absence of such disciplines indicates poor risk management which may be noted when the event is analysed by stakeholders.

Recommendation 2	
Risk Rating	High
The Department should continue and accelerate its implementation of centralisation of certain aspects of data governance in order to increase consistency of approach and decision making for data access. Further, the central governance body should develop a comprehensive set of guidance for line areas that hold data to ensure consistent approaches.	
Management Comments	
Action Officer	**Estimated Completion Date**

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

6.3. Formalities for approval of release of the dataset were not appropriately robust

Finding

No Executive or Deputy Secretary approval was obtained prior to the release of the 10% MBS and PBS Sample dataset to the public. While this may not have been required in a formal delegation context, it is not unreasonable to suggest it would have been appropriate given the unprecedented nature of the data release and the commitments made that such approval would be sought.

Discussion

The Secretary of the Department of Health issues privacy guidelines for MBS and PBS data, consistent with the *Privacy Guidelines for the Medicare Benefits and Pharmaceutical Benefits Programs* issued by the Privacy Commissioner under s135AA of the *National Health Act 1953*.

The Secretary's privacy guidelines (and delegations) do not specifically cater for the use and release of de-identified linkable data, which therefore provides some lack of clarity on where approval to release the data should be sought.

The proposal to release to 10% MBS and PBS Sample data was considered by the Department's Data Governance Council in October 2015. The minutes of that meeting record the following:

> "The DGC agreed to the public release of linkable MBS, PBS and public hospital 10% sample data set, noting the release proposed will be presented to the Executive for final approval."

Following that consideration, this independent review has noted the following briefs regarding the release of the dataset:

- November 2015: Draft Information Brief to Deputy Secretary advising of planned MBS data releases, including reference to10% linkable sample.
- 15 April 2016: Information brief noting pre-launch release of subset of 10% MBS/PBS Sample dataset at PharmHack 2016.
- July 2016: Information brief released to Minister Ley, Assistant Minister Wyatt and Assistant Minister Gillespie informing them of the release of linkable de-identified 10% sample of Medicare and PBS claims data (MB16-001178). (Cleared by FAS and AS in PBD on 26 July 2016).

Each of the April 2016 and July 2016 Briefs note the following: "A suite of confidentiality measures including encryption, perturbation and exclusion of rare events has been applied to safeguard personal health information and ensure that patients and providers cannot be re-identified."

The independent review has not identified any documented approval from the Deputy Secretary or the Departmental Executive to release the dataset. Given the guidance from the Data Governance Council and the fact that this data release was unprecedented, it is reasonable to expect that such an approval would have been obtained prior to release.

Risk Exposure

While Executive review and sign-off may or may not have identified the weakness in encryption, the failure to obtain such review and sign-off represents a process failure which will likely be criticised by stakeholders when this event is analysed.

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

Recommendation 3	
Risk Rating	Medium
Data release policies for the Department should include a requirement for all public releases of unit-level or patient level data to be reviewed and approved by the responsible Deputy Secretary.	
Management Comments	
Action Officer	**Estimated Completion Date**
	The centralised data governance and release framework will be submitted to the Data Governance and Analytics Committee and then to Executive for consideration for endorsement in December and implementation from January 2017.

6.4. Project and risk management disciplines associated with the development and release of the dataset would have benefited from increased rigour

Finding

The 10% MBS and PBS Sample dataset development and release was not conducted as a formal project with associated disciplines regarding business cases, approvals, budget and schedule management and formal testing and clearance processes.

Further, although there was clearly a deep consideration of the risks associated with the development and release of the dataset, no formal risk assessment was prepared, reviewed and approved by a project owner or senior executive.

Discussion

The independent review was advised that the 10% MBS and PBS Sample dataset development and release was not run as a formal project, but rather was "fitted in with other obligations". This means that many of the disciplines associated with good project management (business cases, approvals, budget and schedule management and formal testing and clearance processes) were not applied.

This is not to suggest it was not well-accepted and known that the initiative was underway at Senior Executive and broader Departmental contexts. There was reasonable visibility of the initiative and relevant staff were keeping their Branch and Divisional leaders informed.

However, the absence of business case and certain gate approvals likely missed an opportunity for consideration and guidance on issues such as confidentialising of the data in the dataset. It is impossible to know if such disciplines would have resulted in a different course of action; however the opportunity was not available.

In the context of data integration processes, the Department has a comprehensive risk assessment method and model which is applied to ensure risks are identified and mitigations approved. As this initiative was not a data integration process (it is a data linkage process), such a risk assessment was not undertaken.

It is clear that the officers responsible were highly attuned to the risks to be managed and were using their judgement and best endeavours to manage the risks accordingly. However,

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

the absence of a documented and rigorous assessment missed an opportunity for assurance of completeness of the identification of risks, and the confirmation that approaches being taken to manage risks were appropriate and within the risk appetite of the Department.

Risk Exposure

The absence of effective project and risk management disciplines likely contributed to the de-cryption by the University of Melbourne. Further, the absence of such disciplines indicates poor risk management which may be noted when the event is analysed by stakeholders.

Recommendation 4	
Risk Rating	Medium
Data release policies for the Department should include a requirement for all public releases of unit-level or patient level data to apply Departmental project management disciplines and risk assessment and management disciplines consistent with those applied for data integration projects.	
Management Comments	
Agree The new centralised data governance and release framework will include a formal risk assessment framework which will be applied to all releases of data external to the department. This framework will categorise data releases into low, medium and high risk releases, with standardised assessment, risk mitigation and approval processes to deal with each. This framework will be consistent with that used for data linkage projects, as well as any new whole-of-government processes introduced to deal with unit record data releases. The framework will be designed such that the release approval process will be auditable, reportable and consistently applied across the Department. Project management disciplines consistent with those applied to data integration projects will be applied to data releases generally.	
Action Officer	**Estimated Completion Date**
	The centralised data governance and release framework will be submitted to the Data Governance and Analytics Committee and then to Executive for consideration for endorsement in December and implementation from January 2017.

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

7. Observations

The independent review identified the following area for improvement in knowledge management. This does not relate to a specific control weakness and therefore is not presented as a finding. However, it is suggested that this represents an opportunity for more effective and efficient practices.

7.1. Consistency of standards and testing in central data release forums

It is understood by the Department that there is limited central guidance on de-identification, or data structures, from the administrators of www.data.gov.au. In order to achieve a centralised consistency around the nature, structure and confidentiality of data being loaded on that central release site, there is an opportunity to impose certain standards and/or undertaken certain pre-loading checks to ensure only data that meets the Government's data release standards are released.

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

Australian Government
Department of Health

protiviti
Risk & Business Consulting.
Internal Audit.

Appendix 1: Personnel Consulted

The following personnel were consulted as part of this independent review. We appreciate their assistance and candour throughout:

Name	Title/Role
Mr Andrew Stuart	Deputy Secretary Health Benefits
Mr Paul Madden	Special Adviser Strategic Health Systems and Information Management
Ms Maria Jolly	First Assistant Secretary Medical Benefits
Ms Penny Shakespeare	First Assistant Secretary Pharmaceutical Benefits
Ms Alanna Foster	First Assistant Secretary Research, Data and Evaluation
Mr Matt Yannopoulos	First Assistant Secretary Portfolio Investment
Ms April Purry	Assistant Secretary Legal Services
Mr Jaye Smith	Assistant Secretary Primary Care and Diagnostics
Mr Nick Henderson	Assistant Secretary Pharmaceutical Policy
Mr Ian Crettenden	Assistant Secretary Health Analytics
s47E(c), s47F	Director Information and Knowledge Management
	Director PBS Information Management
	Director MBS Analytics
	Director Information Management and Relationships
	Assistant Director PBS Information Management
Ms Malisa Golightly	Deputy Secretary, Department of Human Services
Mr Mark Whithnell	First Assistant Secretary, Department of Human Services
Ms Joanne Mason	Assistant Secretary, Department of Human Services
s47E(c), s47F	General Manager, Australian Bureau of Statistics
	Director, Australian Bureau of Statistics
	University of Melbourne
	University of Melbourne

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

Australian Government
Department of Health

protiviti
Risk & Business Consulting.
Internal Audit.

Appendix 2: Documents Reviewed

The following documents were considered as part of this independent review:

Documentation Title
Question Time Brief "10% MBS/PBS Sample Decryption", 12 September 2016
File 2013/017801, Pharmaceutical Benefits & Access: Agreements – Review of 10 percent PBS Sample
File 2015/000477, Pharmaceutical Benefits & Access: Agreements – Review of 10 percent PBS Sample – Part 2
Talking Points for Minister's Office "Public Release of Linkable 10% sample of Medicare Benefits Scheme (Medicare) and Pharmaceutical Benefits Scheme (PBS) Date, clearance by FAS PBD
Information Brief, "Release of linkable de-identified 10% sample of Medicare and PBS claims data", various drafts, including comments from PBD Executive.
Various emails relating to release of the data, from PBD staff.
Protocol for the External Release of Health Information and Data (Draft v 0.9), May 2016
High Level Principles for the Release of Health Information and Data (various drafts), May 2016
Information Management and Release Protocol, November 2012 (extract)
Bilateral Head Agreement 2015-2018 between The Secretary of the Department of Human Services and the Secretary of the Department of Health (signed, October 2015)
Department of Health Data Access and Release Policy (August 2015)
Frequently asked questions: further information about the Data Access and Release Policy
Australian Government Public Data Policy Statement (December 2015)
Minute to Deputy Secretary Andrew Stuart, "Pre-launch release of five years of linkable, line-by-line Medicare (MBS) and Pharmaceutical Benefits (PBS) data (April 2016) (final and various drafts)
Administrator of the National Health Funding Pool, "Data Linkage: Governance Framework Considerations" (October 2015) (including request for comment)
Undated document "MBS, PBS Open Source Data Release 2016"
Email exchange February 2016 "MI5073 – Linking MBS & PBS data to CHeReL dataset"
Email February 2015 "Notes from ABS/Health meeting on the 10% PBS and MBS samples"
Email exchange January 2016 "10% MBS/PBS Sample"
Minute to Deputy Secretary Andrew Stuart "For Information: MBS Data Release Plans" (November 2015) (draft)
Minute to Deputy Secretary Andrew Stuart "Approval to Link MBS and PBS Data to Analyse Whether Drugs Used to Treat Multiple Sclerosis are Related to the Incidence of Melanoma" (3 November 2015)
Email exchange November 2015 "MBS data release – update"
Email exchange November 2015 "Response from legals re release of linkable 10% sample of MBS, PBS"

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

Email exchange September 2015 “For comment: draft papers for DGC – release of 10% MBS, PBS and hospital data”
Minute from David Cullen, Strategic Policy Unit “Providing the Strategic Policy Unit with Access to PBS Data”, 1 April 2015
Minute from Paul Creech, Pharmaceutical Policy Branch “Providing the Strategic Policy Unit with Access to PBS Data”, 20 February 2015
Minute from David Cullen, Strategic Policy Unit “Providing the Strategic Policy Unit with Access to PBS Data”, 2 February 2015
Email 6 February 2015 “rules for creating 10% sample”
“Review of the 10% PBS Sample File – Phase 1 Report”, Australian Bureau of Statistics, 28 August 2013
“Review of the 10% PBS Sample File – Phase 2 Report”, Australian Bureau of Statistics, 29 July 2014
Paper “Key events in preparation of 10% sample for release”
Australian Government Solicitor advice “De-identified MBS and PBS data” (Ref: 15194245), 1 December 2015
Australian Government Solicitor “Executive Summary: De-identified MBS and PBS data” (Draft 7 December 2015)
Email exchange March 2016 “Analytics Strategy – some question to assist the development of the strategy”
Title: Linkable De-identified 10% Sample of Medicare Benefits Scheme (Medicare) and Pharmaceutical Benefits Scheme (PBS) (undated)
“Deanonymizing the Medicare Benefits Schedule (MBS) and Pharmaceutical Benefits Schedule (PBS) dataset”, University of Melbourne, 8 September 2016
Email exchange March 2016 “Request for PBS data for PharmHack”
List of PBS 10% Sample Licensees (including addresses)
Letter from Paul Storey “PBS/MBS 10% Data Sample” dated 11 May 2016 (advised that sent to PBS 10% Sample Licensees (with data release)
Minute to Deputy Secretary Andrew Stuart (draft) “To seek your in-principal agreement to publicly release de-identified, linkable 10% sample of the MBS and PBS via data.gov.au” (October 2015)
“Confidentiality practices in DHS”, undated paper.
Legal Services Branch Advice 24 July 2015 “Linking MBS and PBS data”.

Released Under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

Documentation Title
Deputy Secretaries Meeting Discussion Paper "Data – getting it, governing it, making it available, strategic analytics and research", July 2015
Download from www.data.gov.au (dated 14 September 2016) "Linkable de-identified 10% sample of MBS and PBS".
"Public Sector Data Management Report", Department of Prime Minister and Cabinet, December 2015
"Public Sector Data Management Implementation Report", Department of Prime Minister and Cabinet, July 2016
Data Governance Council Terms of Reference, version 1.0
"Enterprise Data Management High Level Principles", version 1.0
Database Alignment Project Enterprise Data Governance Framework, Version 1.01
Commonwealth Health Data Integration Risk Assessment Tool
Data Governance and Analytics Committee Terms of Reference
Five Safes: designing data access for research, University of the West of England, 2016
Data Access and Release Plan – Pharmaceutical Benefits Division v0.3, 31 July 2015
Pharmaceutical Benefits Division 10% MBS and PBS Sample Data Release Chronology
"How to confidentialise data: the basic principles:, National Statistical Service
"Principles on open public sector information", Office of the Australian Information Commissioner, May 2011
"Declaration of Open Government", Minister for Finance, July 2010
Quantitative analysis of re-identification risk from MBS PBS data, Data 61, September 2016
MBS/PBS re-identification scenarios, Data 61, September 2016

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing

Terms of Reference: Audit of the 10% MBS/PBS Data Encryption

Objective: To assess and document the circumstances associated with the recent event of decryption and re-identification of certain MBS/PBS data and then assess and document the controls that were undertaken to manage the risk of privacy breach or unexpected disclosure of the sensitive information with a view to:

(a) Understand and define exposure to the Department
(b) Define root cause of the unexpected disclosure of sensitive information to the Department
(c) Identify recommendations for the future to ensure the risk is managed consistent with the Department's risk appetite.

Scope: All aspects of the selection, assembly, management (including encryption and perturbation) and subsequent re-identification of data associated with the reported incident.

Scope exclusions:

- The audit is not assessing response strategies once the incident as identified.
- The audit is not assessing the Department's data management and publication strategy, but rather will assess the processes and decision making in the case of the data release in question (consistent with that strategy).
- The audit will assess the appropriateness of the encryption and perturbation mechanisms applied, but will not test the encryption and perturbation mechanisms.

Process:

- Through consultation with data owners, stakeholders in DHS, ABS and the Privacy Commissioner, and personnel who identified the ability to decrypt the data, and consideration of existing risk assessment documentation associated with the released data, document the risk associated with the potential decryption of the released data (considering both likelihood and consequence of decryption and identification).
- Through consultation with the University of Melbourne, identify and document how the decryption was identified and implemented, with a view to assisting in identification of root cause.
- Document all aspects of the management of the project to select, assemble and, manage (including decryption and perturbation) the data, with a focus on identification of key controls to manage the risk. Key controls may relate to (but not be limited by): due diligence regarding release project, selection of encryption and perturbation techniques, testing of encryption and perturbation, information authorisation, quality assurance and risk assurance processes, and decision making and authorisation for release.
- Identify the root causes of the unexpected decryption of information.
- Through consideration of root cause and controls, identify:
    - Whether any weakness in control contributed to the occurrence of the event, and whether this is outside the Department's documented risk appetite.
    - What (if any) recommendations should be implemented in the future to manage the risk to a level consistent with the Department's risk appetite.
- Make comment on records retained to support the overall project and management of risk associated with the data release.

DRAFT

Released under the Freedom of Information Act 1982 by the Department of Health, Disability and Ageing