Report: 05 December 2022

Australian Digital Health Agency

Commercial-in-Confidence

PRIVACY IMPACT ASSESSMENT – DATA ANALYTICS INFRASTRUCTURE

IIS Partners
INFORMATION INTEGRITY SOLUTIONS

Contents

Glossary

Abbreviation or term	Expansion or definition
Agency	Australian Digital Health Agency
AIHW	Australian Institute of Health and Welfare
APP	Australian Privacy Principles
DAIW project	Data analytics infrastructure workstream project
DGB	Data Governance Board
Disclosure	When an entity makes personal information accessible to others outside the entity and releases the subsequent handling of the information from its effective control
Function Creep	The gradual widening of the use of a technology or personal information beyond the reasonable expectations of individuals (to the detriment of individual privacy). A hallmark of function creep is when personal information is used for a purpose that is not the original specified purpose
Health	The Department of Health and Aged Care
IIS	Information Integrity Solutions Pty Ltd
IWG	Implementation Working Group
MHR	My Health Record
MHR Act	*My Health Records Act 2012*
MVOT	Multiple Versions of Truth
PbD	Privacy by Design
PIA	Privacy Impact Assessment
PoC	Proof of concept
RACI	Responsible, Accountable, Consulted, Informed
SSOT	Single Source of Truth

1. Executive summary

The Australian Digital Health Agency (the Agency) engaged IIS Partners (IIS) to conduct a Privacy Impact Assessment (PIA) for the research and public health data analytics infrastructure to be established as part of a proof of concept (PoC) project. IIS refers to this aspect of the wider PoC as the data analytics infrastructure workstream (DAIW) project.

The Agency in collaboration with the Department of Health and Aged Care (Health) and the Australian Institute of Health and Welfare (AIHW) is working on a PoC for the use of My Health Record (MHR) data for research or public health purposes. The Agency is responsible for establishing the data analytics infrastructure, which will enable data preparation and secure sharing of MHR data. The Agency, the AIHW and Health are jointly working to establish and test a pilot of the technical infrastructure, the end-to-end governance, processes and protocols.

This PIA – which focuses on the DAIW project's **development and testing of the data analytics infrastructure** – has been conducted at a point in time where building the technical infrastructure to enable secure data sharing between the Agency and the AIHW has not yet commenced. Rather than undertaking a compliance check against the privacy principles or a detailed security analysis, the PIA takes a conceptual approach to ensure appropriate consideration of:

- Community awareness and social licence
- Good privacy decision-making
- De-identification
- Privacy management and governance.

1.1 IIS's overall view

Overall, IIS considers that the DAIW project has a MEDIUM level of privacy risk.

The project has a HIGH level of inherent privacy risk because of factors including, but not limited to:

- The storage and subsequent use of live MHR data in the testing environment
- The testing environment will store personal information for an extended period of time as the Agency establishes processes for preparing and de-identifying the data
- Individuals may have a perception that their MHR data is being used for broader health system purposes, even if this is not actually intended by project participants
- There are not yet mechanisms in place for data deletion, nor to verify de-identification processes.

These privacy risks tend to be mitigated by the following positive privacy aspects:

- The project is limited in scope (to develop and test the capability of the data analytics infrastructure), rather than actual use for research or public health purposes

- There will be business rules and technical safeguards to limit the extraction of data in certain respects and to apply de-identification
- The data is only used for testing and will be deleted once the testing is complete
- Ongoing attention to information security imperatives
- Strong commitment from key stakeholders, including Health to ensure privacy is considered throughout the DAIW project and wider PoC, as well as to promote community awareness and participation.

IIS has identified more work that needs to be done to address privacy risk – for example, across areas of openness and transparency, data minimisation, de-identification, and privacy governance – and has made recommendations in this regard.

1.2 Recommendations

IIS has made **13 recommendations** which address key issues and areas of privacy risk identified in Sections 4-6 of the report.

Recommendations	Who	When
Recommendation 1 – Formalise community consultation plans in relation to the development of the data analytics infrastructure	Agency, AIHW & Health	As soon as practicable
Recommendation 2 – Inform the community about the PoC	Agency & Health	During the DAIW project
Recommendation 3 – Avoid 'use case' to describe the Agency's testing of the data analytics infrastructure	Agency	As soon as practicable
Recommendation 4 – Be clear in public communications of de-identification of MHR data on the one hand and downstream uses on the other hand	Agency	During the DAIW project
Recommendation 5 – Limit collection from the MHR system to what is necessary (as far as possible) for the purpose of developing and testing the data analytics infrastructure	Agency	During the DAIW project
Recommendation 6 – Formalise de-identification approach for the data analytics infrastructure	Agency & AIHW	During the DAIW project
Recommendation 7 – Clarify the AIHW's role in de-identification within the Agency's data analytics infrastructure	Agency & AIHW	As soon as practicable

Recommendations	Who	When
Recommendation 8 – Determine level of resourcing needed for data governance and de-identification functions	Agency	During the DAIW project
Recommendation 9 – Establish mechanisms to determine the effectiveness of de-identification controls	Agency	During the DAIW project
Recommendation 10 – Undertake shared risk analysis and document in a shared risk register	Agency, AIHW & Health	As soon as practicable
Recommendation 11 – Develop a RACI matrix for data governance and de-identification	Agency & AIHW	During the DAIW project
Recommendation 12 – Develop a plan for responding to privacy and security risk events arising from the data analytics process flow	Agency & AIHW	During the DAIW project
Recommendation 13 – Formalise process for and conduct further PIAs on the DAIW project	Agency	During the DAIW project

1.3 Reading the PIA

How to read this PIA report:

- The Glossary explains key privacy and project-related abbreviations and terms used in the report.
- Section 1 outlines IIS's overall view on the DAIW project and provides a summary of recommendations.
- Section 2 describes the DAIW project, including its relevance to the wider PoC project, objectives and scope, project status, project participants, data flows, and key applicable legislation.
- Section 3 provides an overview of IIS's approach to the privacy risk assessment and our assessment of the DAIW project's inherent privacy risks, positive privacy aspects and residual privacy risk level.
- Section 4 sets out IIS's findings with respect to privacy decision-making.
- Section 5 sets out IIS's findings with respect to de-identification.
- Section 6 sets out IIS's findings with respect to governance of privacy risk.
- Appendix A sets out the scope and methodology for the PIA.

2. Project description

2.1 Context

My Health Record (MHR) is an online summary of individuals' key health information, including medical conditions and treatments, medicine details, allergies, and test or scan results. Having an MHR allows individuals to have a personally controlled and easily accessible record of their health, and enables better connected care and access to key health information in an emergency.

Over the past several years, the Australian Government has undertaken legal and policy changes with the goal of using information collected by the MHR system to guide health services planning, policy development and research, and to further improve Australia's health care system.

Following wide community and stakeholder consultation, the Department of Health and Aged Care (Health) published a framework in May 2018 to guide the secondary use of MHR system data (the Framework).[1] The Framework outlines how data in the MHR system may be used for research and public health while maintaining privacy and security.

In December 2018, the *My Health Record Act 2012* (MHR Act) was amended to enact certain principles in the Framework. In particular, it prescribed the AIHW as the 'Data Custodian', provided the authority to establish an MHR Data Governance Board (DGB), and amended the System Operator functions in s 15 of the MHR Act to enable the Agency, in accordance with guidance and direction of DGB, to prepare and provide de-identified data (and, with the consent of the healthcare recipient, health information) for research or public health purposes.

Since July 2021, the Agency, the AIHW and Health (the tripartite group) have been working together to progress the PoC. The agency is implementing the technical infrastructure component. The PoC will also explore the benefits of using MHR data for research and public health purposes. It will investigate establishing a DGB as part of the ongoing governance arrangements required to oversee future MHR data research projects.

Health has created the role of Interim Chair of the DGB to provide expert advice for the PoC work. The Interim Chair will provide strategic support for:

- A refresh of the 'Framework'
- The PoC, including establishing the DGB and developing a legislative rule that will impose requirements on people handling MHR information for research and public health purposes
- Conveying the benefits of using MHR data for research and public health purposes
- Consulting with the broader public and key stakeholders on their views, including what they see as benefits from using the MHR data in scope.

[1] https://www.health.gov.au/sites/default/files/documents/2021/12/framework-to-guide-the-secondary-use-of-my-health-record-system-data.pdf.

2.2 Enabling the PoC – establishing and testing a technical infrastructure

One of the primary responsibilities of the Agency for the PoC is to establish the data analytics infrastructure, which will enable data separation, de-identification, and preparation for sharing of de-identified MHR clinical data to a secure protected environment within the AIHW. De-identification will be conducted according to AIHW standards and practices.

The data analytics infrastructure workstream (DAIW) project is a component of the wider PoC project outlined above. This workstream aims to establish the basic capability, including technical solutions as well as know-how to extract relevant data from the MHR system and prepare and de-identify these data for testing, to make sure that future sharing of de-identified data for research or public health uses can occur once the legislation is in place.

The Agency, the AIHW and Health are jointly working to establish and test a pilot of the technical infrastructure, the end-to-end governance, processes and protocols. The pilot will focus on MHR data relating to two 'use cases'.[2]

Data that is extracted from the MHR system for the project will be for the purposes of testing the technical solutions to be established in the data analytics infrastructure. The use of any MHR data by the System Operator must be authorised by law and must meet the reasonable expectation of healthcare recipients.

This PIA focuses on the DAIW project's **development and testing of the data analytics infrastructure**. It has been conducted at a point in time where the project team has yet to start building the technical infrastructure (which would enable secure data sharing between the Agency and the AIHW), to support consideration and remediation of privacy risk from the outset.

2.3 DAIW project status

The Agency is currently evaluating responses to the RFQ for building the data analytics infrastructure. The Agency has also conducted the first stage of the Threat and Risk Assessment. It is currently developing supporting documents for the data analytics infrastructure, including the business rules for MHR data extraction and the Standard Operating Procedures.

[2] Note that IIS is referring to 'use cases' for consistency with existing Agency information. It does not refer to use for actual research and public health or for any broader health system outcomes; rather, the use is to test the technical capability of the data analytics infrastructure. We discuss the term 'use cases' Section 4.1.3 below.

2.4 DAIW project participants

The key participants in the DAIW project are as follows:

2.4.1 Australian Digital Health Agency (the Agency)

The Agency is the Australian Government statutory agency that delivers digital innovation, health systems and services in accordance with the National Digital Health Strategy. It plays a lead role in developing the MHR system and is the System Operator of the MHR system.

The role of the Agency for the DAIW project includes:

- Leading the technical infrastructure build
- Determining the data extraction business rules and methodology
- Developing de-identification framework and rules for the data analytics infrastructure
- Developing data governance (drawing on AIHW guidance).

2.4.2 Australian Institute of Health and Welfare (AIHW)

The AIHW is the independent statutory Australian Government agency that provides meaningful information and statistics for the benefit of Australian people by working with health and welfare data. The AIHW is the MHR Data Custodian for research and public health purposes. It is responsible for ensuring the security, quality and usefulness of data before it is released for research projects.

The role of the AIHW for the DAIW project includes:

- Co-design and expert advice to the Agency for the project
- Assisting the Agency in developing de-identification controls, data governance and user requirements
- Developing specifications for delivery of data to the AIHW.

In the absence of a MHR Rule and a DGB, the AIHW is assessing the suitability of MHR data under the extended authorisation of the System Operator. This is done through a legally binding contract between the AIHW and the Agency, where the MHR data remains under the effective control of the System Operator.

2.4.3 Department of Health and Aged Care (Health)

Health develops and delivers policies and programs and advises the Australian Government on health, aged care and sport. It is responsible for digital health policy and legislation, including for the MHR system.

The role of Health is sponsoring of and guidance for the DAIW project in the context of the overall PoC.

2.5 Nature of systems and information flows

2.5.1 Kinds of information involved

As part of testing the data analytics infrastructure and the de-identification process, the Agency as System Operator will extract personal information (including sensitive and health information) from the MHR system, as it relates to healthcare recipients.[3]

Over the course of the PIA the project has continued to evolve; IIS understands that at a high level the following types of personal information will be extracted:

- Given name, family name and initials
- Date of birth
- Demographic information
- Individual Healthcare Identifier (IHI)
- Personal and shared health summary
- Clinical documents
- Prescriptions records
- Information relating to the Medicare Benefits Schedule (MBS), Pharmaceutical Benefits Scheme (PBS) and Australian Immunisation Register (AIR).

IIS understands the Agency is still working on the business rules for the extraction of data from the MHR system. The information has to be sufficient to identify potential records of interest for the cohort of individuals that belong to the two use cases,[4] but also considering the importance of data minimisation. The Agency is seeking to future-proof the methodology so it can be applied to cohorts in the future state.

The Agency has informed IIS of the rules for extraction:

- Not including people who have opted out of de-identified data use for research or public health purposes.
- Not including people's MHRs or documents that have access restrictions turned on
- Check that it is an existing MHR that has not been cancelled (although MHRs are included if the person has passed away)
- Check that the MHR is created in the relevant specified time period
- Check MHRs for cohort of interest (e.g., querying the MHR system's HDR database with key words, fuzzy logic).

[3] The extract will not include personal information of healthcare provider individuals and employees, nor nominated and authorised representatives.

[4] The two cohorts are: (i) individuals receiving treatment for osteoarthritis of the knee and hip, and (ii) individuals with dementia who need or will be needing high care management.

The MHR system has both structured and narrative data (i.e., free text information). For the testing of the data analytics infrastructure, IIS understands the Agency will be extracting and storing both structured and narrative data, and seeking to test de-identification techniques on both kinds of data.

2.5.2 Overview of process flow – MHR system to the Agency

The scope of this PIA involves pulling live data from MHR system to the Agency's secure data analytics infrastructure.

The analytics infrastructure will be built within the National Infrastructure that holds the MHR system and will be hosted in Australia. It will be in the same tenancy but will exist as two separate environments. The data will be stored in the National Infrastructure at the PROTECTED level and the de-identified data that will be shared with the AIHW will be classified as OFFICIAL SENSITIVE.

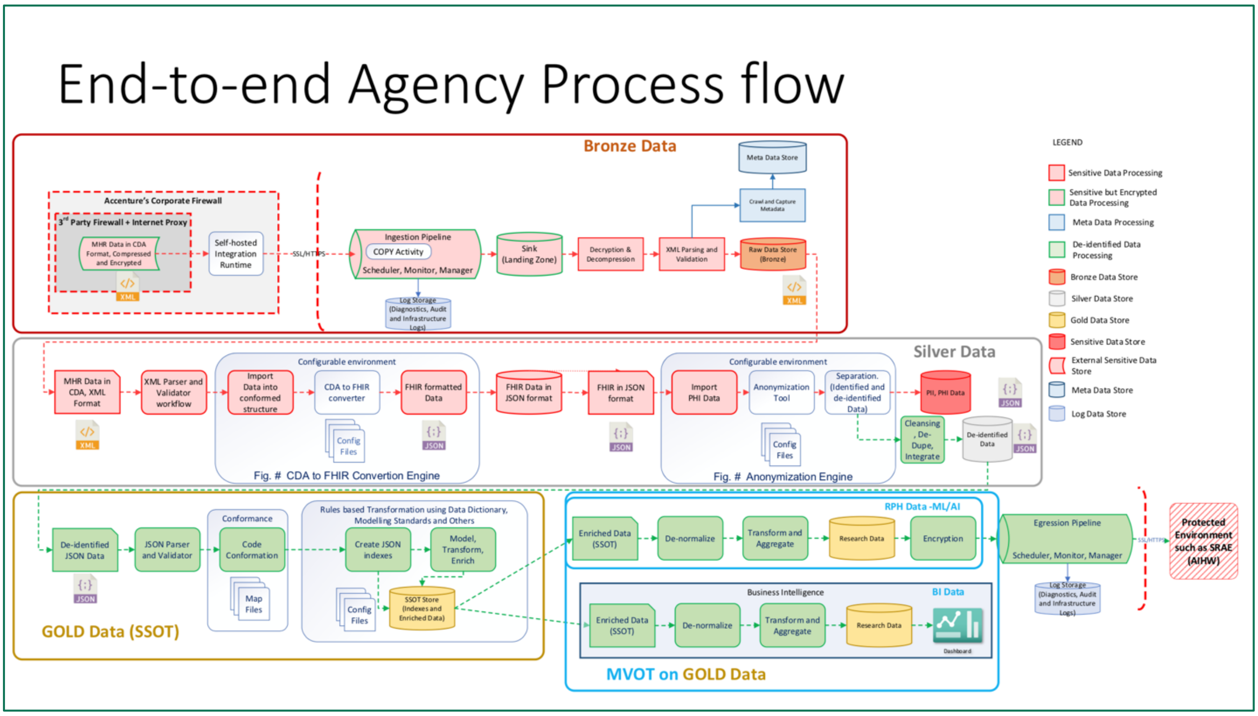
For testing the data analytics infrastructure, only data pertaining to the two 'use cases' will be sought (in accordance with DAIW project business rules) and extracted from the MHR system. IIS understands that the Agency intends to destroy the data that is copied into the analytics infrastructure once testing is finalised. However, at this point it does not have a clear timeline or formalised process for doing so.

IIS has been advised that the information flows for the DAIW project have been developed to test the data analytics infrastructure and should not be construed as final BAU flows. At a high level, the information flows to be tested for the data analytics infrastructure are described below:

Data layer	Description
Bronze Data	The Agency will ingest MHR data from the MHR system that is located in the National Infrastructure and store it in the bronze data layer. Information in this layer is refreshed from the MHR system on a 24-hour basis. The Bronze layer serves as the source for any subsequent data processing.
Silver Data	This is the stage where the data is de-identified. The data is also converted, curated, formatted, integrated, cleansed and deduplicated.
Gold Data	Once the data has gone through the de-identification and preparation process it will be organised and stored in the Gold Data Layer as a single source of truth for the de-identified data. The data can then be further transformed and enriched to cater to researcher requests.

IIS understands that the Agency will not provide real MHR data to the AIHW as part of testing the data analytics infrastructure's integration with the AIHW's system. Rather, it will provide dummy data.

Figure 1. End-to-end information flow within the Agency's network (Source: Australian Digital Health Agency)


End-to-end Agency Process flow
Bronze Data
Accenture's Corporate Firewall
3rd Party Firewall + Internet Proxy
MHR Data in CDA Format, Compressed and Encrypted
XML
Self-hosted Integration Runtime
SSL/HTTPS
Ingestion Pipeline
COPY Activity
Scheduler, Monitor, Manager
Log Storage (Diagnostics, Audit and Infrastructure Logs)
Sink (Landing Zone)
Decryption & Decompression
XML Parsing and Validation
Crawl and Capture Metadata
Meta Data Store
Raw Data Store (Bronze)
XML
Silver Data
MHR Data in CDA, XML Format
XML
XML Parser and Validator workflow
Configurable environment
Import Data into conformed structure
CDA to FHIR converter
Config Files
FHIR formatted Data
JSON
Fig. # CDA to FHIR Convertion Engine
FHIR Data in JSON format
FHIR in JSON format
JSON
Configurable environment
Import PHI Data
Anonymization Tool
Separation. (Identified and de-identified Data)
Config Files
Fig. # Anonymization Engine
PII, PHI Data
JSON
Cleansing , De-Dupe, Integrate
De-identified Data
JSON
GOLD Data (SSOT)
De-identified JSON Data
JSON
JSON Parser and Validator
Conformance
Code Conformation
Map Files
Rules based Transformation using Data Dictionary, Modelling Standards and Others
Create JSON indexes
Model, Transform, Enrich
Config Files
SSOT Store (Indexes and Enriched Data)
RPH Data -ML/AI
Enriched Data (SSOT)
De-normalize
Transform and Aggregate
Research Data
Encryption
Business Intelligence
BI Data
Enriched Data (SSOT)
De-normalize
Transform and Aggregate
Research Data
Dashboard
MVOT on GOLD Data
Egression Pipeline
Scheduler, Monitor, Manager
Log Storage (Diagnostics, Audit and Infrastructure Logs)
SSL/HTTPS
Protected Environment such as SRAE (AIHW)
LEGEND
Sensitive Data Processing
Sensitive but Encrypted Data Processing
Meta Data Processing
De-identified Data Processing
Bronze Data Store
Silver Data Store
Gold Data Store
Sensitive Data Store
External Sensitive Data Store
Meta Data Store
Log Data Store

2.6 Legal framework

The establishment of the data analytics infrastructure must comply with the following relevant laws:

2.6.1 My Health Records Act 2012

The *My Health Records Act 2012* (MHR Act) limits when and how information included in an MHR can be collected, used and disclosed. The Agency is the System Operator of the MHR system and must comply with the MHR Act, including in relation to the collection, use and disclosure of information regulated by the MHR Act. A failure to do so is both a breach of the MHR Act and an interference with privacy, and subject to enforcement action of the Information Commissioner set out in the My Health Records (Information Commissioner Enforcement Powers) Guidelines 2016.

The MHR Act also requires mandatory data breach notification involving data breaches where:[5]

- A person has or may have contravened the MHR Act in a manner involving an unauthorised collection, use or disclosure of health information included in a healthcare recipient's MHR
- Any event that has, or may have, occurred (whether or not involving a contravention of the MHR Act) that compromises, may compromise, has compromised or may have compromised the security or integrity of the MHR system

[5] See OAIC, 'Guide to mandatory data breach notification in the My Health Record system' (6 October 2017).

- Any circumstances that have, or may have arisen (whether or not involving a contravention of the MHR Act), that compromise, may compromise, have compromised or may have compromised the security or integrity of the MHR system.

As the System Operator, the Agency must report a notifiable data breach to the OAIC. It is also responsible for notifying affected healthcare recipients of a breach where this is required by the MHR Act.

IIS notes that MHR information and its uses are explicitly excluded from the *Data Availability and Transparency Act 2022* by the Data Availability and Transparency Regulations 2022.

It is not within scope of this PIA to canvas the extent to which the Agency has established its legal authority for extracting MHR data for the purposes of the DAIW project (or, more broadly, the PoC). The Agency has advised IIS that for the DAIW project it is relying on s 15(o) of the MHR Act, which enables it to conduct work that is incidental to or conductive to the work of s 15(ma), namely, preparing and providing de-identified data (and, with the consent of the healthcare recipient, health information) for research or public health purposes. Further exploration of legal authority by the Agency and Health will be commensurate with advancement of the broader PoC.

2.6.2 The Privacy Act 1988

The *Privacy Act 1988* (Privacy Act) regulates the handling and protection of personal information by Australian Government agencies. Agencies must comply with the Privacy Act and the Australian Privacy Principles (APPs) that contains requirements across the information lifecycle.

2.6.2.1 Australian Government Agencies Privacy Code

The *Australian Government Agencies Privacy Code* sets out specific requirements and key practical steps that agencies must take as part of complying with APP 1.2. It requires agencies to move towards a best practice approach to privacy governance to help build a consistent, high standard of personal information management across all Australian Government agencies. The completion of PIAs for high privacy impact projects is a requirement of the Privacy Code.

3. Approach to our analysis

This section assesses the DAIW project's residual privacy risk level, by weighing the inherent privacy risks against the existing privacy positive aspects.

Sections 4, 5 and 6 then discuss the project's privacy issues and risks identified in detail and make recommendations to mitigate these.

IIS notes that we have not considered IT security as part of our privacy risk assessment, as the technical build has not yet commenced at the time of writing.

While IIS has been conscious of assessing privacy risk for the DAIW project specifically, at times it was necessary to consider the contextual relevance of the larger PoC project in our analysis.

3.1 Inherent privacy risks

IIS's risk analysis approach begins with identifying the inherent privacy risks. Generally, inherent privacy risks arise from:

- The nature of the personal information to be collected and managed – for example, its quantity, sensitivity, and the potential for, and consequences of, misuse
- The range of people from whom personal information may be collected
- The contexts in which personal information is handled in the organisation, broadly, and within the project being assessed – for example, senior management commitment to privacy, employee privacy skills and specific training/experience, the technical systems involved and the nature of the project
- The likelihood of function creep or other privacy-detracting factors – such as those arising from the availability, accessibility and perceived utility of personal information
- The extent to which information is accessed or handled by third parties – such as by vendors and contracted service providers – and the level of privacy sophistication of those third parties
- The likely community and/or media interest in the privacy aspects of the project.

Considering these points, IIS considers the DAIW project has a HIGH degree of inherent privacy risk because of factors including (but not limited to):

- The testing environment will not use 'dummy data'; rather, it will pull live data (i.e., personal and health information) from the MHR system and prepare it via a complex de-identification process.
- The Agency is still determining the boundary of the 'MHR system' as it pertains to where the de-identification of personal information is to take place (although this is likely to be separate to the 'production environment' that supports the provision of healthcare for the MHR system).

- The Agency's testing environment contains – at the Bronze Layer – a reproduction of MHR data, and presents an additional vector for cyber-attack.
- The testing environment will store personal information for an extended period of time as the Agency establishes processes for preparing and de-identifying the data.
- There may be a *perception* risk among the public, regardless of the project participants' actual intentions, that healthcare recipients' personal information is being pulled from the MHR system and used for broader health system purposes that are unrelated to providing them with health outcomes.
- A possible consequence of the perception of privacy invasion is 'decisional interference', where individuals change or alter their usual decisions and behaviours. In the context of the MHR system, individuals who currently participate may decide to opt out of the MHR system due to concerns arising from this project (and their anticipation of similar initiatives in future).
- There are not yet mechanisms in place for data deletion, nor to verify de-identification processes.

3.2 Positive privacy aspects

IIS has also identified positive privacy aspects of the DAIW project, including:

- The project (including its 'use cases') is a limited study to develop and test the capability of the data analytics infrastructure, rather than the use of de-identified MHR data for research or public health purposes
- There will be business rules to limit the data to be extracted from the MHR system, such as excluding people who have opted out of de-identified data use for research or public health purposes and excluding MHRs and documents that have access restrictions turned on.
- There is a focus on testing the de-identification personal information (including transforming it within a secure environment), so that the Agency can increase its experience and capability in this risk mitigation strategy.
- The data is only being used for testing and will be deleted once the testing is complete.
- Ongoing attention to information security imperatives, including technical security reviews and remediation where required.
- The project team within the Agency are privacy conscious and have access to guidance from the Agency's in-house Privacy Officer.
- Strong commitment from Health to ensure privacy is considered throughout the DAIW project and wider PoC, as well as to promote community awareness and participation
- Health and the Interim Chair aim to ensure that the design of the DAIW project is based on potential future uses that are in line with benefits that are elucidated from the public.

3.3 Residual privacy risk level

The privacy positive aspects of the DAIW project mitigate the HIGH inherent risks to some degree. IIS considers that this imbues the DAIW project with positive privacy momentum and decreases the privacy risk to a MEDIUM level.

IIS has identified more that work needs to be done to address privacy risk – for example, across areas of openness and transparency, good privacy decision-making, de-identification, and privacy governance – and has made recommendations in this regard.

Given that implementation of the recommendations in this report will rely on a multi-agency approach with a multiplicity of non-privacy-related variables to work through, IIS is unable to consider the extent to which implementation of the recommendations in this report will have the effect of creating a residual risk for the DAIW project that is less than MEDIUM.

4. Findings – Privacy decision-making

Early in project development is the right time to combine privacy knowledge (i.e., awareness of privacy risks) with good decision-making (i.e., decisions to minimise or eliminate privacy risks, thereby reducing the possibility of privacy harm to individuals).

There is a tendency in government to focus on risk in an inward-looking manner, which is about risks to the organisation – for example, risk of non-compliance with legislation or risk of non-conformity with key governance measures. Viewing risk in this manner can help government to understand (and take steps to address) matters that can harm their ability to operate, be free from regulatory scrutiny and enjoy behavioural conformity.

However, good decision-making in respect of privacy requires an expansion of risk mindset to include consideration of what risk looks like in terms of potential harm to the individual.

For example, knowing that their personal information may be extracted from the MHR system for the purposes of the DAIW project could (if there is a perception of privacy intrusion) cause a person to self-select out of continued participation in the MHR system. This privacy harm is called 'decisional interference'. Similarly, where personal information is relevant to the project's 'use cases', knowing that such data is extracted from the MHR system and used in the testing environment may cause a person to feel that details of their physical health and ailments are exposed. This privacy harm of 'exposure' is relevant even where the information is not made visible to the world at large.[6] The harms lens (focused on the community we serve) is deeply relevant to the project and its ability to create and maintain a 'social licence to operate'.

There are several high-level principles that apply to good decision-making at the commencement of a project, or when developing and deploying technologies that involve the collection and management of personal information in some way. The following principles, which have informed the creation of statutory privacy protections globally since being enshrined in Organisation for Economic Co-Operation and Development (OECD) guidance in 1980[7] and are foundational to Australia's APPs[8] should be considered:

- Openness and transparency
- Respect for people
- Purpose limitation
- Necessity and proportionality.

[6] Dan Solove discusses privacy harm in detail in his seminal article (which is cited in foundational privacy course curricula, including IAPP textbook 'Strategic Privacy by Design' by R Jason Cronk). See: Solove, Daniel J., A Taxonomy of Privacy. University of Pennsylvania Law Review, Vol. 154, No. 3, p. 477, January 2006, GWU Law School Public Law Research Paper No. 129, Available at SSRN: https://ssrn.com/abstract=667622.

[7] OECD Basic Principles of National Application.

[8] IIS expects the APPs will be considered to a greater degree of granularity at future stages of the PoC project.

The analysis in this section provides a description of each of these good decision-making principles and relevance to the project. It discusses what privacy risks are associated with failing to apply each principle, the extent to which mitigations have been explored by the Agency and IIS recommendations in this regard.

4.1 Openness and transparency

Openness and transparency are about ensuring visibility and accountability of government in its practices and decisions. These features are characterised by sufficiently detailed and readily available information (whether via digital or analogue channels), that is communicated using plain-language and easily accessed by the community. Complementary to APP requirements for transparency of personal information handling practices, this principle when taken more broadly speaks to the respect a government has for the community it serves, and empowering, supporting and guiding that community with knowledge.

For the DAIW project, adherence to this principle strengthens the Agency’s social licence; that is, community support for a project it believes is in its interest. IIS considers that there are opportunities to improve in this context.

4.1.1 Respect for people

Trust of the community – and demonstrating trustworthiness as a government – are key factors for success of projects involving personal information.

The Agency and Health pointed to past community engagement on the MHR, from its launch in 2016, to consultation about the Framework for MHR data use for research and public health purposes, and further during the review of the MHR Act in 2020. Public consultation has resulted in important changes, including allowing individuals to opt out from de-identified data use, as well as updating the MHR Act to establish the data custodian, DGB and a specific provision for limited use of health information in certain circumstances.

On the other hand, community trust for the Agency to undertake the DAIW project is not assured, and may, indeed, be inherently low due to historical community sentiment and media representations in relation to the roll-out of the MHR system. Relevant considerations in the current climate include concerns (articulated in the news media, social networks, and by civil society groups and prominent academics) about increased social surveillance perpetrated on Australian citizens by their government, as well as high profile data breaches (e.g., Optus, Medibank) that raise concerns about sensitive data stores.

It will be important to demonstrate to the community that the DAIW project is a test environment to establish the efficacy of using the data analytics infrastructure in a limited context, and that the extent to which it leads to further stages of the PoC project is not yet a fait accompli. However, framing the DAIW project as relevant to the success of the wider PoC is a meaningful opportunity to demonstrate to the community that there is a big picture goal – focused on community health outcomes (not community surveillance) – and to engage them early in the process.

Stakeholder consultations revealed that key executive supporting the PoC project, including the Interim Chair, have taken a strong position on co-designing data use with the community – using the following mantras, “Nothing about me without me” and “Research with, not about, groups”.

IIS understands that community consultations about the PoC are in early stages, and notes that (while it would be imprudent – and potentially a security risk – to consult in relation to granular details, such as schematics and technical controls, for the data analytics infrastructure) it will be relevant to inform the community that the DAIW project is underway and indicate its role in the larger context of the PoC.

Recommendation 1 – Formalise community consultation plans in relation to the development of the data analytics infrastructure
Formalise plans to raise community awareness about the development of the data analytics infrastructure, noting the context of the wider PoC. **Who**: Agency, AIHW and Health **When**: As soon as practicable

4.1.2 Informing the community about the PoC

In addition to being consulted about the testing and capabilities of the data analytics infrastructure, it is important that the community is fully informed about the PoC (which comprises the DAIW project, among other initiatives). Individuals may wish to opt out of participating in the MHR altogether if they know their personal information will be used to create de-identified data sets (for the PoC, and potentially in a more routinised way in the future).

Importantly, consideration must be given to the potential harm – in addition to the previously mentioned privacy harm of ‘decisional interference’ – to individuals should they decide opt-out; that is, the extent to which their health outcomes are at risk (noting that co-location of health records, no matter where in Australia an individual accesses health services, is a key selling feature of the MHR).

At present, a preponderance of government messaging about the MHR focuses on enhancing the ability for health providers to provide timely care based on up-to-date information. It is possible that, even where individuals are deeply concerned about the PoC, they will feel unable to exercise any choice at all for fear of losing the benefits they believe are associated with participating in the MHR. Likewise, opting out may jeopardise the care of frequent health system users whose practitioners rely on their MHR to provide a complete patient profile.

IIS acknowledges that there is some information in the public domain about the PoC project, of which the DAIW project is a part:[9]

> *Since July 2021, these partners* [i.e., the Agency, AIHW and Health] *have been working together to establish the technical infrastructure needed to run a proof of concept. The proof of concept will also explore the benefits of using My Health Record data for research and public health purposes.*
>
> *It will investigate establishing a Data Governance Board as part of the ongoing governance arrangements required to oversee future My Health Record data research projects*.

IIS considers that the information supplied is insufficient to be meaningful to the average community member (and may even be alarming without further context, using language like *'exploring benefits'* and *'future […] research projects'*). Progress on the DAIW project provides an opportunity to take the next step of informing the community about the PoC in a more detailed manner. Details should include information about privacy protection.

Recommendation 2 – Inform the community about the PoC

Publish a plain-language summary of the PoC that includes:

- Details of the project and its health system purposes
- Information about how privacy is being protected
- A reminder of the current pathway for those wishing to opt-out of the MHR due to privacy concerns (whether or not those concerns relate to the PoC)
- Information about the potential consequences of opting-out of the MHR.

Who: Agency and Health

When: During the DAIW project

4.1.3 Using helpful terminology

When describing the DAIW project to the community, it is important to communicate that it uses two hypothetical case studies to test the capabilities of data analytics infrastructure. At present, internal documentation describes the hypothetical case studies as 'use cases'. This language – with the focus on 'use' – may imply that real-world decision-making using MHR data is occurring and create a perception that routinised use of MHR data for broader health system purposes is not 'conceptual' but is, rather, already a fait accompli.

[9] Excerpt from Department of Health and Aged Care webpage entitled 'Use of My Health Record data'.

IIS canvassed this issue with the Agency during our engagement. Possible alternatives to ‘use case’ could include ‘themes’ or ‘approaches’ to preparing the MHR data for testing.

Recommendation 3 – Avoid ‘use case’ to describe the Agency’s testing of the data analytics infrastructure

Ensure that all internal and external references to the testing of data analytics infrastructure avoids the term ‘use case’ and instead contains accurate terminology that is easily understood by the community.

Who: Agency

When: As soon as practicable

4.2 Purpose limitation

Purpose limitation seeks to ensure there is a clear and proper purpose for the work undertaken by government, rather than a vague or indeterminate purpose. From a privacy perspective, this principle requires that personal information collected for a specified purpose is not used for a new, or incompatible, purpose.

The DAIW project has an identified purpose of testing that the data analytics infrastructure has the capability to support data preparation relating to two unique hypothetical case studies (the ‘use cases’). Future iterations of the project will seek to do the same across a wider variety of health outcomes. It is unclear, however, if the purpose for which MHR data was collected from the community in the first place can be neatly ‘mapped’ to the purposes of the project.[10]

4.2.1 Link to reasonable expectations

When the community was asked to opt-in to the MHR system, the communications explaining purpose were centred around continuity of healthcare – for example, better connected care, personally controlled, access to your key health information in an emergency, and your health information in one place.[11] The relevant communications did not focus on other health system purposes, such as accessing the MHR system to derive de-identified data sets that support the project and other initiatives.

[10] See also, Legal Authority.

[11] See, e.g., https://www.myhealthrecord.gov.au/for-you-your-family/my-health-record-benefits.

The OAIC comments on the meaning of reasonable expectations in its guidance of use and disclosure of personal information, and offers the following:[12]

> The 'reasonably expects' test is an objective one that has regard to what a reasonable person, who is properly informed, would expect in the circumstances. This is a question of fact in each individual case. It is the responsibility of the APP entity to be able to justify its conduct.

It may not be within the reasonable expectations of the community that personal information collected and stored by health practitioners while providing them with a health service – which they have opted to have stored in the MHR system – would be used for other government purposes, such as health system planning, resource allocation, community intervention, etc.

In order to strengthen the public's reasonable expectations for the use of their MHR data for the DAIW project and the wider PoC (IIS considers it would be challenging to artificially separate them in the public's mind), the Agency should ensure that Australians are properly informed about:

- How the DAIW project and the PoC relate to the MHR system, especially from the perspective of benefits to the healthcare recipient (consistent with the original intent of the MHR system)
- That the act of extracting, preparing and de-identifying the data for testing purposes is the limited 'use' to which the MHR data is put (as opposed to broader research and public health uses).

IIS emphasises that such communication should be timely, accessible and easy to understand (i.e., does not require the person to have a sophisticated understanding of the healthcare system or machinery of government).

Recommendation 4 – Be clear in public communications of de-identification of MHR data on the one hand and downstream uses on the other hand

In project documentation and public-facing communications about the DAIW project:

- Explain how the DAIW project and the PoC relate to the MHR system, especially from the perspective of benefits to the healthcare recipient
- Ensure that the 'use' to which personal information from the MHR system is being put (i.e., testing of de-identification) is described accurately and not conflated with the broader health system uses of de-identified data sets.

Who: Agency

When: During the DAIW project

[12] OAIC, 'Chapter 6: Australian Privacy Principle 6 – use or disclosure of person information', para 6.20.

4.3 Necessity and proportionality

The principles of necessity and proportionality work together to ensure that:

- Only the personal information that is necessary to fill a defined purpose is collected and used
- The amount and sensitivity of the personal information is proportionate to the need.

This helps to minimise both real and perceived privacy invasions at the outset. Concerns of community and privacy advocates about the potential for MHR data to be misused elevate the importance of these principles in decision-making for the project.

4.3.1 Evidence of necessity

It will be important to provide evidence that using data derived from the MHR system is necessary to develop and test the data analytics infrastructure, tailored to suit the purposes, and likely to be effective (in terms of confirming the capabilities of the infrastructure via the 'use case').

IIS recognises that the Agency is still in the early stages of determining the most effective way to test the data analytics infrastructure. This makes it inherently difficult to definitively rule certain kinds of information in or out at the outset, let alone developing the technical means of doing so.

Presently, the Agency has developed a set of extraction rules (see Section 2.5.1 above) that seek to place limits on the MHR data that the infrastructure will collect from the MHR system. Broadly speaking, the rules set limits with respect to:

- Business need – only MHRs pertaining to the two use cases and within applicable time periods will be extracted
- Privacy – not extracting MHRs of people who have opted out of de-identified data use for research or public health purposes and/or with access restrictions turned on.

IIS accepts that the Agency has taken reasonable steps in the circumstances and at this point in time to limit the collection to what is necessary. However, we do note that the quantity of information extracted for the in-scope MHR data is substantial and includes both structured and narrative data contained in the MHR. This is a privacy risk that requires ongoing attention and management.

Recommendation 5 – Limit collection from the MHR system to what is necessary (as far as possible) for the purpose of developing and testing the data analytics infrastructure

Evaluate the extent to which the data sets copied from the MHR system are necessary for achieving the purpose of developing and testing the data analytics infrastructure, and report on this.

Who: Agency

When: During the DAIW project

5. Findings – De-identification

In this section, IIS makes findings about de-identification based on our understanding of good practice and information we have gathered about the DAIW project and the anticipated future state.

5.1 De-identification – specific controls

The DAIW project is about developing effective and appropriate de-identification treatments to support the wider PoC's goal of exploring the benefits of using MHR data for research and public health purposes. Section 15(ma) of the MHR Act specifically mentions the role of the Agency as System Operator to 'prepare and provide de-identified data' for such purposes.

According to the Office of the Australian Information Commissioner (OAIC), whether information is personal or de-identified depends on the context:[13]

> Information will be de-identified where the <u>risk of an individual being re-identified in the data is very low</u> in the relevant release context (or data access environment). Put another way, information will be de-identified where there is <u>no reasonable likelihood of re-identification occurring</u>. [Emphasis added]

The OAIC considers de-identification to involve two steps:

- Firstly, the entity should remove direct identifiers (such as name and address)
- Secondly, the entity should take one or both additional steps:
  - Removal or alteration of other information ('quasi-identifiers') that could potentially be used to re-identify an individual
  - The use of controls and safeguards in the data access environment to prevent re-identification.

The AIHW, along with the Australian Bureau of Statistics (ABS), has adopted the Fives Safes framework as an approach to thinking about, assessing and managing re-identification risks associated with the use and release of data. IIS understands that this framework will also be applied to the research and public health use of MHR data involving the Agency and the AIHW.

IIS anticipates that the Five Safes dimensions will be applied in the following ways (we have also highlighted relevance to the DAIW project):[14]

[13] https://www.oaic.gov.au/privacy/guidance-and-advice/de-identification-and-the-privacy-act/.

[14] IIS has adapted the table from the AIHW's explanatory material on the Five Safes framework, available here: https://www.aihw.gov.au/about-our-data/data-governance/the-five-safes-framework

Dimension	Meaning	Application to MHR data
Projects	Use of the data is legal, ethical and the project is expected to deliver public benefit.	This dimension will be considered in the future as part of AIHW's initial interaction with the researcher and the approval process involving the Data Governance Board, consistent with purposes that are enshrined in legislation.
People	Data users have the knowledge, skills and incentives to act in accordance with required standards of behaviour.	Relevant Agency people may require access to the MHR data in order to prepare, de-identify and provide the data to the AIHW. The DAIW project should establish controls including: ● Defining who within the Agency should have access to the MHR data ● The kinds of knowledge and skills they should have (including training where appropriate) ● Confidentiality undertakings ● Role-based access controls.
Data	Data has been treated appropriately to minimise the potential for identification of individuals or organisations.	The MHR data will be copied into the analytics infrastructure at the Bronze layer. Various levels of treatment will be applied at the Silver and Gold layers to de-identify the data. The DAIW project should define what the specific data treatment techniques and controls are throughout the data lifecycle, from the MHR system to the AIHW secure research access environment.
Settings	There are practical controls on the way the data is accessed – both from a technology perspective and considering the physical environment.	The data extracted from the MHR system will be held in a series of logical environments in the Bronze, Silver and Gold layers of the analytics infrastructure. The DAIW project should ensure that the underlying infrastructure is secure, and that there are business and technical controls to restrict access to the logical environments where the data is held.

Dimension	Meaning	Application to MHR data
Output	A final check can be required to minimise risk when releasing the findings of the project.	There are two points at which checks will need to be conducted to minimise the risk of involuntary disclosure of identifiable data: (i) before data is provided from the Agency environment to the AIHW environment, and (ii) before the AIHW makes the data available to any internal or relevant stakeholders who may require visibility of the output as part of the POC (e.g., internal AIHW, Agency or Health personnel, executives; researchers) (future state). The DAIW project should ensure there is a clear process and delineation of responsibilities for checking the output as data moves from the Agency environment to the AIHW environment.

The Agency has committed to following de-identification guidance currently being prepared by the AIHW, in terms of applying elements of the Five Safes within its data analytics infrastructure. The end goal is that the Agency will only provide de-identified MHR data to the AIHW, with any processing of identifiable data occurring in the analytics infrastructure subject to strict controls. IIS understands that the Agency anticipates MHR data will be provided to the AIHW as part of testing the data analytics infrastructure's integration with the AIHW's system.

At the time of writing, the AIHW is in the process of finalising its 'Approach to De-identification' report that outlines the approach AIHW recommends in applying its policies to future releases of de-identified MHR data, including where and how de-identification would be incorporated in data flows between the Agency and the AIHW. IIS has reviewed a draft version of the report (dated 1 September), which focuses on the AIHW's approach to de-identification generally and to the MHR data more specifically. There are useful concepts and practical de-identification steps mapped against each of the Five Safes. However, more work is required to translate these concepts and steps into the Agency's implementation of the data analytics infrastructure.

Recommendation 6 – Formalise de-identification approach for the data analytics infrastructure

Develop and formalise the de-identification approach for the data analytics infrastructure, drawing upon the AIHW's guidance. The approach should clearly set how the Agency will apply the Five Safes.

Who: Agency and AIHW

When: During the DAIW project

5.2 De-identification – broader considerations

In addition to the kinds of de-identification controls to be implemented by the Agency as flagged above, IIS has identified three other considerations:

- The roles and responsibilities for de-identification within the Agency's environment
- The Agency's organisational commitment to implementing de-identification controls
- Ongoing assurance of the efficacy of de-identification controls.

5.2.1 Roles and responsibilities

The AIHW will play a pivotal role in helping the Agency set up the internal data governance and management structures for the analytics infrastructure, including on de-identification. In discussions with the two parties, IIS recognised that there are areas of responsibility that are still to be defined between the two parties, especially the extent to which the AIHW will have a de-identification role in the Agency environment.

One example is the development and deployment of specific de-identification tools. IIS understands that the AIHW is currently exploring the use of machine learning (ML) to assist with the de-identification of narrative data that contain a lot of free text information. The Agency anticipates that it will trial the de-identification of narrative data after June 2023. It is conceivable that such an ML tool developed by the AIHW could be deployed within the Agency's data analytics infrastructure. The circumstances and terms by which such a tool is deployed (e.g., the extent to which the AIHW can access data from the Agency environment to refine the tool) will need to be agreed between the Agency and the AIHW.

Another area that needs to be clarified in a timely manner is whether and how the AIHW will play a role in checking the de-identification of data within the Agency environment before it is provided to the AIHW environment (which would be at a later stage in the project). Given that the Agency intends for the DAIW project to reflect as much as possible the future state, the role that the AIHW will play in this area should be clarified as a matter of priority.

Recommendation 7 – Clarify the AIHW's role in de-identification within the Agency's data analytics infrastructure
Clarify whether and how the AIHW will play a role in checking the de-identification of data within the Agency environment before it is provided to the AIHW environment. **Who**: Agency and AIHW **When**: As soon as practicable

5.2.2 Organisational commitment

IIS observes that the DAIW project is notable not only for the brand-new data analytics infrastructure, but that it marks a significant strategic move in terms of the Agency seeking to establish data analytics capabilities. It is envisaged that these capabilities will first revolve around MHR data, with the possibility of expanding to other datasets in the future.

There are varying levels of data analytics capability and experience within the Agency. In addition to building the technical infrastructure, the Agency will need to establish a new team and 'ways of working', including on de-identification.

This requires a serious organisational commitment, by which IIS means that the Agency must be willing and able to:

- Acquire expert advice on how to set up its data governance and de-identification controls for the analytics infrastructure
- Establish processes and hire people required to carry out the data analytics and implement the appropriate controls
- Commit to ongoing training, oversight and improvement
- Foster a culture that remains cognisant of the community at the heart of data initiatives
- Dedicate sufficient resources to support the above.

In discussions, the Agency informed IIS that it is confident that the organisational commitment is present. Those IIS spoke with pointed to:

- The recent formation of the Data Strategy and Governance team
- The alignment of this project with one of the key Agency priorities of using digital information to support vital research to benefit public health planning and resourcing
- The buy-in from the Agency's leadership for this project
- The ongoing working relationship with the AIHW and Health
- The current oversight and input from the Interim Chair of the DGB on all governance structures and processes, including de-identification that will be established
- The guidance and direction of the DGB (once established).

The Project Initiation Document indicates that the 2021-22 Federal budget includes approximately $7 million for the Agency to implement the POC over two financial years. It also identifies a number of key project roles, including the hiring of a Data Governance Lead and Data Governance Support Officer, as well as a Senior Data Engineer and Senior Data Analyst. IIS understands that these roles have been established, along with three additional Senior Data Policy Leads that are part of the Data Strategy and Governance team.

IIS is not in a position yet to comment on the appropriateness of this level of resourcing. We encourage the Agency to scope out the resources required as the data governance and de-identification requirements are formalised, with a view to establishing the data analytics infrastructure as well as 'future proofing' its role in supporting the broader PoC.

Recommendation 8 – Determine level of resourcing needed for data governance and de-identification functions

Determine the level of resourcing the Agency requires to support its data governance and de-identification functions as they are formalised for the DAIW project. Ensure that appropriate resourcing is available for the functions to continue into the broader PoC project and as the Agency's data analytics capabilities increase in sophistication.

Who: Agency

When: During the DAIW project

5.2.3 Ongoing assurance

IIS heard repeatedly from the Agency during the engagement that it wants to carry out the DAIW project and any future plans in a way that maintains community trust and confidence. This includes taking a Privacy by Design approach, being transparent with what it is doing, undertaking stakeholder engagement (including via Health channels), and implementing strong de-identification controls.

To 'close the loop' on the good practices that the Agency is embarking on, IIS considers that it should also commit to demonstrating proof of performance. Firstly, this involves determining that its de-identification controls are in fact working. This is a way for the Agency to remain accountable for what it has promised to do. Secondly, by demonstrating this to stakeholders, the Agency can enhance its trustworthiness. This can include both areas where the controls are working as intended, as well as areas where deficiencies were found and actions the Agency will take for continuous improvement.

While at this stage it is too early to assess the Agency's de-identification controls, it should be developing them with a view to determining their effectiveness down the line. For example, this can include setting up appropriate logs of user access and system activity, and committing to a cadence of internal review. As a matter of best practice, Agency should also consider expert and independent verification of its de-identification controls. This can be combined with publishing the results of such assessments to enhance transparency and give the public confidence.

Recommendation 9 – Establish mechanisms to determine the effectiveness of de-identification controls
As part of implementing de-identification controls for the DAIW project, ensure there are mechanisms for determining that controls are working and there is a cadence for regular review of effectiveness. **Who**: Agency **When**: During the DAIW project

6. Findings – Governance

In this section, IIS makes findings about privacy governance aimed at helping the Agency and other key stakeholders manage privacy risks for the DAIW project and beyond.

6.1 Management of shared risk

According to the Federal Department of Finance, a shared risk 'extends beyond a single entity. It is a risk that emerges from a single source and impacts interrelated objectives of entities.'[15]

In our increasingly complex and connected world, shared risk is a common feature of projects and initiatives. For example, many worthwhile initiatives – including the MHR – require the participation of multiple parties from different sectors and different jurisdictions. The Australian National Audit Office (ANAO) specifically raised shared risk as a consideration for the MHR system.[16]

The management of shared risk presents challenges, as outlined in the table below:

Distinguishing features of shared risk	Implications for the DAIW project
A shared risk may have no naturally apparent owner and no one entity may be able to manage the risk on their own.	It is important to identify who has (shared) ownership and responsibility for managing each identified privacy risk (as the risk could be unstated, confused or not assigned at all)
Shared risks can have complex causes, and can be influenced by the actions and inaction of a range of participants in different ways.	More than one party (i.e., the Agency and AIHW) may cause the risk (in terms of contributing to its occurrence) and/or mitigate the risk (in terms of existing controls and potential treatments)
Should a shared risk be realised, it can affect different parties in different ways.	The consequences of privacy risk may impact (i) individuals (e.g., personal and financial), (ii) key stakeholders (the Agency, AIHW, Health) (e.g., reputation and business goals), and (iii) the overall MHR system (e.g., public confidence).

[15] https://www.finance.gov.au/sites/default/files/2019-11/comcover-information-sheet-understanding-and-managing-shared-risk.pdf

[16] https://www.anao.gov.au/work/performance-audit/implementation-the-my-health-record-system

Sources of shared risk

In the context of the DAIW project, shared privacy risks may arise throughout the data lifecycle as data is collected from the MHR system, prepared and de-identified by the Agency in its data analytics infrastructure, and provided to the AIHW. Privacy risks that result in unauthorised access, use, disclosure or loss of MHR data may arise from discrete events such as:

- Unauthorised user accessing the data via shared or compromised login credentials
- Loss or theft of data during transit
- Failure to exclude data of people who have opted out or placed restrictions on their record
- Inadequate application of de-identification controls leading to the disclosure of identifiable data.

At a higher level, privacy risk is influenced by 'macro' factors that have a downstream effect on the extent to which the MHR data is properly used, protected and governed. These factors may include:

- The growing scale and complexity of the shared privacy risk environment over time – for example, as data collection and processing grow beyond the DAIW project and into the broader PoC, and at BAU with more researchers accessing the data for research and analysis, etc.
- Unclear lines of responsibility and accountability for overall shared privacy risk governance
- Increasing appetite to leverage the MHR data, including more use cases and greater volume of use overall
- Lack of resources for implementing protective operational and governance arrangements, and for conducting assurance activities.

Current findings

Based on information gathered from the Agency and AIHW, there are some important building blocks for managing shared risk that are already in place:

- The Project Initiation Document has mapped internal and external stakeholders, and presents an interagency view of project implementation (across the Agency, AIHW and Health)
- Project oversight is carried out by an interagency Implementation Working Group and a Senior Responsible Officer Forum; there is shared project governance and understanding of project risks
- There are regular interagency project officer meetings to discuss risks and timelines
- The Agency maintains a Project Risks and Issues Register, which tracks risks that are owned by other parties.

IIS considers that the approach to shared risk management can be further strengthened by:

- Thinking of risk management not just in parallel (i.e., each party responsible for its own risks) but also as interdependent

- Going beyond DAIW project delivery and considering what the (shared) risks are and what each parties' role is in managing them moving through the PoC and into BAU (e.g., how data governance and de-identification will work in practice)
- Going beyond an organisational view of privacy risk and considering it from the perspective of, and consequences for, individuals.

Practically speaking, strengthening shared risk management could involve undertaking a formal joint exercise in identifying shared risks, their mitigations and recording them in a shared risk register. Furthermore, the Agency and the AIHW should explore the development of a Responsible, Accountable, Consulted, Informed (RACI) matrix that will delineate responsibilities for data governance and de-identification both for the DAIW project as well as for the future state.

Recommendation 10 – Undertake shared risk analysis and document in a shared risk register

Undertake a shared risk analysis to identify and document shared risks (including privacy risks) and their mitigations for the DAIW project. The shared risk register should track:

- (Co-)owners of the risk
- (Co-)contributors to the risk
- The consequences of the risk
- Risk treatments and the part(ies) responsible for them

Who: Agency, AIHW and Health

When: As soon as practicable

Recommendation 11 – Develop a RACI matrix for data governance and de-identification

Develop a RACI matrix that outlines data governance and de-identification responsibilities between the Agency and the AIHW.

Align the RACI matrix with the de-identification guide to be developed by the AIHW and ongoing work to establish a data governance framework for the Agency's emerging data analytics function.

Who: Agency and AIHW

When: During the DAIW project

6.2 Responding to failure

Closely related to the issue of managing privacy risk is responding when a risk event materialises. When done well, an organisation's response to failure can enhance stakeholder confidence and trust. The testing of the data analytics infrastructure provides a good opportunity to consider what could go wrong, and make preparations accordingly

Following the end-to-end data flow, it is clear that there are multiple points of potential failure, with different kinds of risk events that could eventuate:

Process flow	Risk scenario
Ingestion of MHR data into the Bronze Layer within the analytics infrastructure	Unauthorised access to MHR data (identified) Failure to properly exclude records
Separating MHR data and applying de-identification controls in the Silver Layer	Unauthorised access to MHR data (identified or de-identified) Failure to properly apply de-identification controls
Preparing de-identified data for sharing in the Gold Layer	Unauthorised access to MHR data (de-identified) Failure to properly apply de-identification controls
Ingestion of de-identified data into AIHW environment	Unauthorised access to MHR data (de-identified) Failure to properly check de-identification
(*Future state*) Access and use by researcher	Inappropriate use of data beyond approved project specifications
At every stage of the process flow	Compromise of data by external attacker

Where the risk event involves personal or health information, it could become an eligible data breach under the Privacy Act and/or a data breach under the MHR Act. In such cases, the Agency would presumably follow its existing data breach response plan.

However, as the table outlines there could be risk events downstream of the Bronze/Silver Layer where a compromise of MHR data occurs that does not meet the Privacy Act or MHR Act's definition but nevertheless could have consequences for the Agency and other stakeholders. The Agency should proactively address such events that fall below the legislative threshold of a 'data breach'.

A good reference point is the Agency's existing contract with the AIHW, whereby the Institute is authorised to access health information to assess the suitability, quality and de-identification strategy of MHR data for research or public health purposes. IIS reviewed the specific provisions on responding to failure and we note two important aspects:

- In addition to outlining the steps in the event of an eligible data breach (within the definition of the Privacy Act), the contract also addresses the handling of 'MHR data',[17] including what happens if the AIHW becomes aware of its unauthorised access, use or loss
- The Agency is framed as the accountable authority, and the AIHW must notify the Agency and comply with its instructions.

The next step will be for the Agency to develop a plan for what to do in risk events involving MHR data (both identified and de-identified), if and when they arise. IIS considers that this would not entail creating something brand new, as the Agency could leverage its existing data breach and security incident response plans.

Recommendation 12 – Develop a plan for responding to privacy and security risk events arising from the data analytics process flow

Develop a plan for responding to privacy and security risk events arising from the data analytics process flow. The plan should:

- Consider what could go wrong at multiple points in the process flow
- Specifically address risk events that fall below the legislative threshold of a data breach
- Address how the Agency will work together with key stakeholders, such as the AIHW and Health
- Be aligned as much as possible to the Agency's existing data breach response procedures

Who: Agency and AIHW

When: During the DAIW project

[17] Defined broadly as 'any data from the [MHR] system and any data (whether identified or de-identified) derived from the data'.

6.3 Governance of change

A key privacy protection for the DAIW project will be ensuring that any changes with privacy implications happen only with systematic consideration of privacy risks. Without a commitment to such a formal risk assessment process, there is a risk that ad hoc or incremental changes could lead to significant privacy impacts.

IIS considers that the most essential element for proper privacy governance of change is to conduct PIAs at key junctures of the project. These could include, for example:

- Any change to the way the data analytics infrastructure collects, stores or uses personal or health information
- Integration of additional source systems
- New internal uses of the data
- Data sharing with external parties
- Changing the internal stance on retaining versus deleting the data.

The Agency should leverage its internal processes for conducting privacy threshold assessments and PIAs, as well as consult relevant OAIC guidance on this topic. Such assessments do not have to be 'one-size-fits-all', but rather be appropriately rigorous considering the size and scope of the proposed changes.

Recommendation 13 – Formalise process for and commit to further PIAs on the DAIW project
Ensure that consideration of privacy changes is part of ongoing project governance and management. Conduct further PIAs where proposed changes to the project could affect the handling of personal or health information. **Who**: Agency **When**: During the DAIW project

Appendix A. Scope and methodology

A.1 Scope

The Australian Digital Health Agency (the Agency) engaged IIS Partners (IIS) to conduct a Privacy Impact Assessment (PIA) for the research and public health data analytics infrastructure to be established as part of a proof of concept (PoC) project, adapted to two themes.

The scope of the work was to guide the Agency during the design and development phase of the data analytics infrastructure by applying a Privacy by Design (PbD) approach. This is to ensure that the Agency meets compliance as well as community expectations. The PIA documents IIS's PbD advice and was conducted in a point in time during the build for the PoC.

The assessment was made against the APPs, the Five Safes Framework, the PbD principles and IIS's understanding of privacy best practice and community expectations.

In providing this report, IIS makes the following qualifications:

- The PIA considers possible security or technical issues for the project, but it did not undertake detailed investigations or reviews of technical or security features
- The PIA is based on information gathered from the Agency and AIHW
- IIS conducted stakeholder consultations for the project and the PIA took these into account
- IIS does not provide legal advice; rather we provide strategic privacy and security advice.

A.2 Methodology

IIS took the following steps to carry out the PIA:

- *Planning* with the Agency to confirm the approach and deliverable
- *Gathering information* by reading documents and meeting with relevant stakeholders from the Agency and AIHW
- *Analysing* the information against privacy obligations and taking account of possible broader privacy issues, regulator guidance, and privacy best practice
- *Identifying privacy risks* and developing ways to mitigate those risks
- *Drafting Draft 1 of PIA report* and providing this to the Agency and Health for comments and feedback
- *Addressing initial feedback and providing Draft 2 of PIA report* to the Agency for comments and feedback,
- *Addressing final feedback*, *finalising the PIA report and providing it* to the Agency.

A.3 Documents reviewed

Documents reviewed
1. Implementation Working Group (IWG) approved Interagency
2. Research and Public Health Use of MHR Data – Project Initiation Document 1.0(1)
3. 202204 IWG Brief - Prioritised Use Cases
4. FINAL 2019 MHR Data Quality Assessment PIA
5. Research and Public Health Use of MHR data - Project Initiation Document 1.0
6. Research and Public Health use of MHR data Privacy Impact Assessment DRAFT v007
7. 20220726 Process flow diagrams for IIS
8. MyHR Logical ER
9. MHR research or public health data application process end-to-end (session 2)
10. Approach to deidentification report Final Draft 1 Sept 2022
11. Data Developer Procurement Risk Assessment- Data Analytics Platform
12. MHR Data Extraction for PoC Use Case 1
13. Standard Operation Procedure
14. CDA Structure Complexity Narrative Section

A.4 Meetings held

Meetings held (in-person and remote, excludes phone calls)	Date
Kick-off meeting ● IIS consultants ● Agency personnel	1 July 2022
Information gathering – Workshop ● IIS consultants ● Agency personnel	13 July 2022

Meetings held (in-person and remote, excludes phone calls)	Date
Information gathering – PoC • IIS consultants • Agency personnel	26 July 2022
Information gathering – AIHW • IIS consultants • Agency personnel • AIHW personnel	27 July 2022
Information gathering – Cybersecurity • IIS consultants • Agency personnel	27 July 2022
Information gathering – Follow-up • IIS consultants • Agency personnel	2 August 2022
Information gathering – Further follow-up • IIS consultants • Agency personnel	5 August 2022
Feedback on PIA draft #1 • IIS consultants • Agency personnel • Health personnel	16 September 2022
Feedback on PIA draft #1 – Further follow-up • IIS consultants • Agency personnel • Health personnel	29 September 2022
Feedback on PIA draft #2 • IIS consultants • Health personnel	25 November 2022

INFORMATION INTEGRITY SOLUTIONS PTY LTD
PO Box 978, Strawberry Hills NSW 2012, Australia

P: +61 2 8303 2438
F: +61 2 9319 5754
E: contact@iispartners.com
www.iispartners.com

ABN 78 107 611 898
ACN 107 611 898

IIS Partners
INFORMATION INTEGRITY SOLUTIONS