

Submission to the Consultation on
Development of a Framework on Secondary Use of My
Health Record Data



FUTURE
WISE

Introduction

Thank you for the invitation to make a submission to the consultation on secondary use of myHealthRecord data and the development of a framework to support this.

About Future Wise

This submission is authored by Future Wise. We are a group of Australian professionals of varied backgrounds who seek to promote ideas which improve the long-term direction of Australia, particularly in the areas of technology and health. More information about Future Wise is available on our website.¹ We are happy to provide further clarification of any of the points in the submission, or for one of the authors to attend any further hearings in person if required. Please contact Dr Yarwood via email: trent@futurewise.org.au or info@futurewise.org.au

Dr Yarwood has been contracted as a medical advisor to the Australian Commission on Safety and Quality in Healthcare's AURA surveillance system. He has not worked with the Commission's e-Health team. His contribution to this submission is in his personal capacity as a member of Future Wise and as an independent medical specialist, and should not be considered to reflect the position of the Commission or the Commonwealth Public Service.

¹ <https://futurewise.org.au/>

Summary of Submission

Future Wise:

- Supports the secondary use of myHealthRecord data for epidemiologic purposes, including disease statistics, hospital casemix funding and health service planning
- Recommends that a “privacy first” model of secondary use be applied to the framework’s development
- Recommends that the Framework include an example of a patient consent form for secondary use of healthcare data, to support clinicians in explaining to patients what this secondary use entails and the potential risks to their privacy
- Highlights that de-identification of data cannot be absolute without significantly degrading the quality of the data and its utility for scientific use
- Highlights that the greatest risk to individual privacy is through the linkage of data from myHealthRecord with other personal information (for example, obtained through other privacy breaches or through commercial data gathering)
- Rejects that secondary use of myHealthData should be allowed in any way by commercial entities, including (but not limited to pharmaceutical companies and health insurance companies); or by intelligence or security services or the non-health-related branches of the Australian Public Service

Discussion

Future Wise acknowledges the uptake of digital healthcare across the primary and hospital sectors, and recognises the important advances in clinical care and patient safety that can come about through better access to health data by healthcare professionals. Increased patient agency in healthcare has been shown to improve the management of some chronic diseases, and if patients take an active interest in their myHealthRecord, then it has the potential to be a valuable tool in improving the quality of clinical care.

Future Wise is also strongly supportive of Open Data, and applauds the Government's commitment to the principles of the Open Government Partnership, including the "open by default" approach to Government datasets.

However, it is imperative to understand that data aggregated from myHealthRecord is *not* Government data. Legally, medical records remain the property of the clinician or institution which has created them, and myHealthRecord aggregates and makes these records accessible. This aggregation of data does not transfer "ownership" of the data to the government. Ethically, the individuals who are the subject of the healthcare information contained in these records should also have certain rights over this data. Further, as the information in the myHealthRecord is all health information, it is considered *sensitive information* under the *Privacy Act (1998)* and therefore "open by default" is clearly not the relevant default position for the data contained in myHealthRecords.

In the release of the *National Digital Health Strategy* ("the Strategy"), the Australian Digital Health Agency ("ADHA") notes the high value Australians place on the confidentiality of their healthcare data, and the importance of "protect[ion]...from any unauthorised access".² This context is an important reminder that it is critical to acknowledge the risks to patient privacy and

² Australian Digital Health Agency. Australia's National Digital Health Strategy. 2017. Available [online](#).

confidentiality of the myHealthRecord system generally, as well as from secondary use of the data.

The greatest risk to privacy comes from the collection of data in the first place. If data is not reasonably necessary for one of the clinician's functions or activities, it should not be collected.

Government collection of data should proceed from a position of assuming that data security issues are inevitable, and take steps by design to minimise the risk of unauthorised access or disclosure.

It is imperative that the Secondary Use Framework ("the Framework") take patient confidentiality as its most important principle. Once the privacy of an individual's sensitive healthcare information has been compromised it is impossible for that confidentiality to be regained. The nature of healthcare information means that the consequences of the inevitable data breach may be much more severe for an individual than a breach involving non-healthcare related data. For example, data about sexually transmitted infections, domestic violence, or pregnancy outcomes.

There have been too many instances of data breaches of Government-collected data, which highlights that a risk-reduction approach prior to any secondary use is essential.

Of significant concern in this context is the way that a dataset of Medicare Benefits Scheme (MBS) and Pharmaceutical Benefits Scheme (PBS) data published by the Government on data.gov.au was able to be partially re-identified by a team of researchers from the University of Melbourne.^{3,4} Although the re-identification in this instance was of medical provider numbers, rather than patient identifiers, this incident highlighted the challenges in successfully de-identifying such a large dataset. Combined with the leak and sale of Medicare

³ <http://www.abc.net.au/news/2016-09-29/medicare-pbs-dataset-pulled-over-encryption-concerns/7888686>

⁴ <https://pursuit.unimelb.edu.au/articles/understanding-the-maths-is-crucial-for-protecting-privacy>

numbers reported this year,⁵ the potential privacy implications of an inadequately de-identified healthcare dataset are significant.

It is also salutary to consider the lessons we can learn from examples of big data in healthcare which have occurred overseas. The Royal Free NHS trust in the United Kingdom partnered with [DeepMind](https://www.deepmind.com),⁶ a British artificial intelligence research company, to try to improve the clinical care of patients in Royal Free Hospitals. This trial was found by the UK Information Commissioner's Office to have failed to have complied with the UK's *Data Protection Act*.⁷ It is worth noting that the same protections do not all apply in Australia (although some of the issues will be covered by the provisions of the *Privacy Act*) so the protections to patients under Australian law are significantly weaker.

It is a fundamental principle of public health research that questions of data structure should be decided before data collection begins. This process enables databases and data collection tools to be designed to be efficiently used to collect and then make use of data for research purposes. In contrast, with a "data aggregation platform" like myHealthRecord, this process is clearly not possible, as the input data will be in the form created by the creator of the medical record. This does not mean, however, that a specification for the sort of derived data from myHR to be shared with third parties under the framework should not be considered.

In addition to the privacy issues, it is also worth considering the potential utility of this aggregated healthcare data. The consultation paper describes a number of historic examples of the successful application of big data to health. While it is certainly likely (and even probable) that secondary use of myHR data will lead to

⁵ <https://www.theguardian.com/australia-news/2017/jul/04/the-medicare-machine-patient-details-of-any-australian-for-sale-on-darknet>

⁶ <https://www.deepmind.com>

⁷ <https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2017/07/royal-free-google-deepmind-trial-failed-to-comply-with-data-protection-law/>

public health benefits, as well as its use in service planning, it is impossible to accurately predict the benefits which will flow from such secondary use.

This uncertainty about use and benefits means it is not possible for members of the public to give proper informed consent for the use of their data. Informed consent for medical research is a key part of the *Ethical Principles for Medical Research Involving Human Subjects* (also known as the Declaration of Helsinki).

Articles 8 and 9 of the Declaration state:

8. While the primary purpose of medical research is to generate new knowledge, this goal can never take precedence over the rights and interests of individual research subjects.
9. It is the duty of physicians who are involved in medical research to protect the life, health, dignity, integrity, right to self-determination, privacy, and confidentiality of personal information of research subjects. The responsibility for the protection of research subjects must always rest with the physician or other health care professionals and never with the research subjects, even though they have given consent.

Further, Article 32 states:

32. For medical research using identifiable human material or data, such as research on material or data contained in biobanks or similar repositories, physicians must seek informed consent for its collection, storage and/or reuse. There may be exceptional situations where consent would be impossible or impracticable to obtain for such research. In such situations the research may be done only after consideration and approval of a research ethics committee.

Future Wise believes that the wholesale secondary use of data in myHealthRecord is too broad a scope to be permitted. The risks to privacy are too great for the secondary use of this data to fall under the “exceptional situations” exemption of the usual requirements for consent.

Even if there was a mechanism by which patients could provide consent to the reuse of their data (for example, as part of the framework), it is very likely that the majority of clinicians tasked with gaining informed consent from patients would have an insufficient working knowledge of the data security risks associated with Big Data analysis to provide a truly informed assessment of the risks, which invalidates the informed consent.

These real and significant risks should be considered in the context of the benefits of such data reuse. It would be proportionate to the risk and reasonable to expect that secondary use of data could be used for descriptive epidemiology and health service planning; as clinical coding data generated in hospitals is already used in these ways.

Population health statistics are also an obvious and reasonable use for aggregated myHealthRecord data. It is important to recognise, however, that there may be significant differences between statistics generated from big data analysis of secondary use data and rates of diseases defined using accepted clinical definitions.

The reuse of data derived from clinical notes in many cases does not provide accurate epidemiological data. For example, diagnoses may be under-reported due to failure to record in the notes, may be over-reported due to incorrect diagnostic coding.^{8,9,10}

It is therefore important not to overstate the potential benefits to public health that can be easily obtained from the analysis of this data without significant input from content area experts.

⁸ Jhung & Banerjee. *Clin Infect Dis* 2009. doi: 10.1086/605086

⁹ Burns et al. *J Pub Health* 2011. doi: 10.1093/pubmed/fdr054

¹⁰ Mitchell & Ferguson. *Infect Dis Health* 2016. doi: 10.1016/j.idh.2016.03.002

In addition, there are already multiple rich datasets available for use in this area - data from the Medicare Benefits Scheme and Pharmaceutical Benefits Scheme, the Australian Institutes of Health and Welfare datasets, including the Hospital Morbidity database, the Public Hospital Establishments database and the various Non-admitted patient care databases. While aggregates of myHealthRecord may add additional data to these existing datasets, a wealth of data exists already, which may be better utilised for health benefits.



Response to Consultation Questions

Question 1: What secondary purposes, if any, should My Health Record data be used for?

Secondary use of myHR data should be with the express purpose of improving the care provided to patients.

This can be through health-service planning, population epidemiology and descriptive statistics and aggregate data where there is no risk to patient privacy or confidentiality.

Research on linked data should only be done on an opt-in basis, where there is adequate deidentification (noting that there is a trade-off between the quality of the linked data and the privacy protection afforded by reidentification) and within a framework of informed consent for this sort of research. It would be within the scope of the framework to develop a model consent form which could be used / modified by clinicians at the time of enrolling patients in myHealthRecord.

Question 2: What secondary purposes should My Health Record data not be used for?

Secondary use of the data in myHealthRecord should be explicitly forbidden for organisations whose primary purpose is commercial - specifically pharmaceutical companies and health insurance companies.

The consultation paper also makes mention in passing of secondary use for "security" without elaborating on what this means. Future Wise rejects categorically the use of health data for law enforcement, security or other related purposes and calls on the Digital Health Agency to clarify what the meaning of this mention in the consultation paper involves.

Question 3: What types of organisations/individuals should be able to access My Health Record data for secondary purposes?

- Healthcare organisations recognised by the Australian Institute of Health and Welfare and their employees
- Registered healthcare practitioners, where they are trained in, or work with people trained in population health epidemiology as well as the privacy implications of secondary data use and data linkage
- Academics and higher education facilities
- Statutory agencies (for example, The Commission for Safety and Quality in Healthcare)
- Public sector agencies directly involved with health service planning, the delivery of healthcare and/or the financing thereof (but limited to those areas directly relevant to these services - for example the Medicare team of Department of Human Services, but not Centrelink)

Question 4: Should access to My Health Record data for secondary uses be restricted to Australian users only or could overseas users be allowed access?

Given the global nature of academic research and the mobility of health services researchers (especially the movement of these researchers from Australia to positions overseas), restricting access to Australia would likely limit the utility of the aggregated data. It is also possible that, if not done correctly, this sort of geoblocking would unreasonably restrict Australian researchers who use virtual private networks (VPNs) or cloud services.

Restricting data access to Australia only would also reduce the potential for collaboration with countries overseas already undertaking research on large linked datasets.

Question 5: What principles, if any, should be included in the Framework to guide the release of data for secondary purposes from the My Health Record system?

1. Secondary use of health data is considered medical research under the principles of the Declaration of Helsinki and therefore all research should be under the auspices of a Human Research Ethics Committee
2. The framework should be developed using the guiding principle of “privacy first” - whereby all practical steps are taken to preserve the privacy of the patients whose data is being used
3. Truly informed consent regarding the privacy implications of secondary use requires greater understanding of the risks of healthcare information technology than is possessed by most clinicians who will be providing records to myHealthRecord, and the data custodian should therefore take an active role in the consent process.
4. It would be a valuable role of the Framework to provide support to clinicians in providing informed consent to individuals in a meaningful way

Question 6: Which of the governance models described above should be adopted to oversee the secondary use of My Health Record data?

Future Wise supports the AIHW model, whereby a committee experienced in the handling of healthcare data is involved with the request, and appropriate, independent ethical review is assured.

Question 7: What principles, if any, should be adopted to enable organisations/researchers to request and gain approval for de-identified data from the My Health Record system to be provided for secondary purposes?

- Informed consent of patients consenting to their healthcare data being made available for secondary use
- Best-practice de-identification (for example, using the CSIRO De-Identification Decision-making Framework¹¹)
- Approval by Human Research Ethics Committee
- Genuine health benefit (or expected benefit) of research
- Requirement for best-practice data-handling techniques, including requirement for secure storage and deletion once research is complete

Question 8: What principles, if any, should be adopted to enable organisations/researchers to request and gain approval for identified data from the My Health Record system to be provided for secondary purposes?

Identified data should not be provided for secondary purposes, as the risk to individual privacy is too great. De-identification and linkage should be done (using best-practice de-identification practices) by AIHW or another agency prior to the release of any data.

Question 9: Should there be specific requirements if researchers/organisations make a request that needs the My Health Record data to be linked to another dataset? If so, what should these requirements be?

The secondary dataset should also be de-identified using best-practice principles (for example the CSIRO guidelines) and should retain only a linkage key that cannot be re-identified or potentially re-identified. Specifically, SLK3 is not acceptable as a linkage key, given the potential for aggregating this with other publically accessible data to allow re-identification.

¹¹ <https://publications.csiro.au/rpr/download?pid=csiro:EP173122&dsid=DS2>

Question 10: What processes should be used to ensure that the data released for secondary purposes protects the privacy of an individual?

See above for discussion.

- Privacy-first design of secondary use framework
- Restriction of disclosure of secondary use data to authorised researchers / health services
- Mandatory training for people accessing the data in principles of data security and the risks to privacy of inadequately de-identified data
- Use of best-practice data de-identification techniques, and avoidance of potentially re-identifiable linkage keys (eg SLK3)
- Mandatory requirements for secure data storage and deletion of data when it is no longer required / after a defined period of time

Question 11: What precautions should be taken to reduce the risk of de-identified data from the My Health Record system being re-identified after release?

- Limitation of access to data which has not been de-identified
- Mandatory requirements for secured data storage and deletion as soon as practicable
- "Black box" data linkage - provision of linked datasets by an external agency without using a potentially re-identifiable linkage key

Question 12: What arrangements should be considered for the preparation and release of My Health Record data and who should be responsible for undertaking and overseeing these arrangements?

In an ideal situation, the data in myHealthRecord would be entered in such a way that minimal processing of aggregate data would be required. The nature of the system as a central store of a variety of records from multiple sources mean that this is unlikely to be possible.

Future Wise considers it essential that a team containing expertise in epidemiology, clinical coding and data security be involved in preparing and reviewing any secondary use data prior to its release to other researchers.

Question 13: Whose responsibility should it be to make a quality statement about the My Health Record data and to ensure the data are of high quality?

If a team is involved with data review prior to its release (see Q12 above), then they should be well positioned to make a statement on data quality.

Question 14: What monitoring and assurance processes, if any, should be considered to ensure My Health Record data secondary users comply with the Framework?

Monitoring of data use once the data has been released “into the wild” is extremely difficult, and relies on goodwill of researchers, unless secure data access portals are used. The use of a “walled garden” for researchers to access the data in a controlled way may limit the ease with which researchers may make use of the data and decrease the utility of the data for secondary use.

Future Wise believes that it would be appropriate for the Department of Health to require that researchers accessing the data have done formal training in good clinical practice for research (these training courses are available for free online).

Question 15: What risk mitigation strategies should be included in the Framework?

Future Wise believes that the Framework should be developed using a “privacy first” model, whereby the primary focus is maintaining the confidentiality of individual’s health data, given the seriousness of health data breaches and the importance that patients place on their healthcare privacy.

The Framework should support as the primary risk mitigation strategy the introduction of a tool for informed consent for use of a patient’s data for secondary use. It would be an essential part of this to develop resources to train clinicians in data security, so that they have a better understanding of the privacy implications of this sort of secondary use.

The Framework should mandate best-practice data de-identification procedures, acknowledging the effects that this may have on data quality.

Question 16: Should there be a public register which shows which organisations/researchers have requested data, the purpose, the status of their data request, what they have found by using the data; and any publications that have resulted from using the data?

Yes, this sort of publicly-accessible register would increase the transparency of secondary use, and provide reassurance to the public their data is being used appropriately.

While noting that myHealthRecord data does not fit under the Open Data framework, Future Wise supports making available lists of publications resulting from the data, and furthermore would support a requirement that these publications themselves be made available in Open Access journals.

Question 17: Are the existing penalties under the My Health Record Act sufficient?

Future Wise supports strong penalties for misuse of health data, while noting that a deterrence-based regime does not provide any meaningful protection to privacy. Penalties for misuse act as a deterrent, and sanction users or organisations who use or release this data improperly, but these methods cannot regain the privacy and confidentiality of those affected by a data breach.

Future Wise supports all breaches of myHealthRecord data be considered as eligible data breaches under the *Privacy Amendment (Notifiable Data Breaches) Act 2017*, due to the potentially severe impacts on patient privacy. We believe that this should be the case even with de-identified data, given the potential for data to be linked with datasets from other publicly-accessible datasets.

Future Wise does not support the “knowing or reckless” exemption for data breaches. The security of health data and the privacy implications of lapses are sufficient that lack of malicious intent should not be grounds for ignoring a data breach.

Question 18: What policy changes, if any, need to be considered to support the release of de-identified data for secondary uses from the My Health Record system?

Future Wise believes it is important to highlight that myHealthRecord data is not under the remit of the Open Government Principles. Despite our strong commitment to the principles of open data, it is critical that secondary use not be rushed through citing OGP as a rationale. This also applies to the recent Productivity Commission report on data availability; the risks to privacy of increased access to healthcare data must always be balanced against any potential benefits.

We welcome the development of CSIRO guidelines on data de-identification, and support its introduction as a mandatory component of government data practice.

Future Wise further believes that a “privacy first” approach should be introduced by Government in all matters dealing with healthcare information.

Conclusion

Future Wise supports appropriate secondary use of healthcare data, and recognises the potential for health benefits to arise from this use. However, we believe that the clinicians primarily responsible for providing data to myHR, and also for discussing with patients their participation in myHealthRecord are poorly equipped to discuss the risk this use poses to patient privacy and confidentiality.

The Framework must provide these clinicians with support to provide patients information on these risks, so that a process of truly informed consent can occur.

Data linkage always poses a risk to privacy, and mandatory safeguards must be put in place to minimise the harms arising from an (inevitable) data breach of this data. Deidentification must necessarily either be incomplete, or so great that it risks degradation of the data. The need for this trade-off means it is essential that experts in both population health data and information security are involved with its processing prior to release.