DATA SHARING AND RELEASE
Legislative Reforms Discussion Paper, September 2019
15 October 2019

The Australian & New Zealand Real-World Data Network (RADiANT) appreciates the opportunity to comment on the Legislative Reforms Discussion Paper.

We offer the following unsolicited feedback:

1. We note that information provided under the **My Health Record** (MHR) scheme is listed as being one of two main classes of information that will likely be exempted from the scope of the legislation. Justification for this proposed exemption is required, given that a framework to guide the secondary use of MHR system data is in place, and the MHR Act (section 15(ma)) specifically defines one of the functions of the System Operator as 'to prepare and provide de-identified data for research or public health purposes'. It is not clear why MHR should be treated differently from other potentially sensitive health data. For innovative projects looking to harness national electronic medical record (EMR) data for rare diseases or iteritive association analysis in complex disease, the secondary use of MHR data may prove critical given the multiplicity of data systems and processes nationally, and the significant approval and technological challenges in integrating these for research.

2. **Waiver of consent**. We support the availability of a waiver of consent for data sharing. We suggest the data sharing principles could align with the long-standing, effective National Health & Medical Research Council (NHMRC) conditions for a waiver of consent for public good health and medical research (see National Statement on Ethical Conduct in Human Research, 2.3.10).

3. **Re-consent**. We would like to see national harmonisation and Human Research Ethics Committee (HREC) guidance surrounding the re-consent of child research participants as adult research participants (in the same project). At present, neither legislation nor the NHMRC National Statement on Ethical Conduct in Human Research provide clarity regarding this matter. While most HRECs will ask for the child to be re-consented at the age of 18, the Department of Human Services (DHS) now asks that children provide their own consent at the age of 14 years for the release of their DHS-held data. The latter effectively suggests that children 14 years and older are mature enough to consent to research participation. If so, can these children consent to full research participation at 14 years to alleviate the challenges of a second re-consent at 18 years? Or can another harmonised mechanism be identified?

4. **Privacy positive measures**. Applications to human research ethics committees under a waiver of consent require that researchers only receive data that is reasonably necessary for their research – this is framed as 'data minimisation' in the discussion document (p31) and Privacy Impact Assessment. Modern 'big-data' approaches, including deep learning and random forests, require all available data about individuals. Will the legislation allow the use of these approaches?

5. **Public interest test**. Galexia recommended an additional public interest test. This appears to be inherent in the NHMRC National Statement on Ethical Conduct in Human Research and the conditions for waiving consent.

We offer the following responses to the questions posed by the Office of the National Data Commissioner (ONDC):

1. **Do you think the distinction between data sharing and data release is clear? How could this distinction be clearer?**

   We think the distinction between data sharing and data release is clear.

2. **What are the challenges for open release of public sector data?**

   Ensuring that any release of anonymised unit record data cannot be combined with other data, or otherwise be manipulated, to reveal the identity of an individual or organisation.

   We see merit in establishing a working group with stakeholders to identify variables suitable for open release so that data interoperability and utility for future sharing and integration projects can be assessed.

3. **Do you think the Data Sharing and Release legislative framework will achieve more streamlined and safer data sharing?**

   The success of the framework will be contingent upon:
   (i) Uptake by State and Territory governments and their data custodians,
   (ii) The datasets included in the exemption list,
   (iii) How the framework is operationalised and the relevant entities are funded. Without additional funds, the currently experienced lengthy delays at the AIHW (12-18 months+) will only be exacerbated by the increase in demand for data sharing,
   (iv) Whether the legislation is applied in a consistent manner for government and non-government researchers. There are currently major inequities.
   (v) The production and availability of enduring National Interest Datasets,
   (vi) The availability of secure remote access laboratories with sufficient processing power, and a reasonable cost structure.

   However, as described in Figure 4 and the related text, it is highly likely the draft data sharing Framework will not improve data sharing and may make it worse. Figure 4 outlines a generic pathway for data sharing that would be unworkable in practice. For example, requiring researchers to have a conversation with a Data Custodian will generate a huge burden on both parties, and will slow down the process. Such a conversation is not necessary for every proposed use, and it places an excessive responsibility on the data custodian. A one-size-fits all approach (as represented by

Figure 4) is potentially highly limiting and, if taken literally as the data sharing framework, may have unintended consequences.

We recommend that the data sharing role of the primary Data Custodian could be devolved to a body/intermediary that acquires the responsibility for the protection of data that are shared for secondary use and integration (e.g. an Accredited Data Service Provider), particularly for health sector data. We believe that data stewardship for the sharing of anonymised records should not rest solely with the primary data custodian, as this approach quickly becomes unworkable once multiple datasets, with different primary data custodians, are integrated. Likewise, for enduring integrated datasets (e.g. National Interest Datasets) we believe consideration should be given to devolving governance of the integrated asset to a single body, rather than requiring decisions regarding access to the integrated asset to be escalated to individual data contributors.

Consideration could be given to designating a "Data Steward" in each Accredited Data Service Provider who has delegated responsibility for decisions to share data, and for setting the conditions of the Data Sharing Agreement.

4. What do you think about the name, Data Sharing and Release Act?

We agree that "Release" should be removed from the name of the Act, given the focus on shared data, which is distinct from open data.

5. Do the purposes for sharing data meet your expectations? What about precluded purposes?

The purposes for sharing data, and the precluded purposes, meet our expectations.

In addition to "*Increasing the transparency … which improves the community's trust in the government's handling of data*", the Act should also promote greater community awareness and education of the significant potential of data sharing for research innovation and continuous improvement in service delivery (particularly in healthcare), against concerns regarding privacy and security breaches. Public co-design is critical in such endeavours.

6. What are your expectations for commercial uses? Do we need to preclude a purpose, or do the Data Sharing Principles and existing legislative protections work?

We recommend that rigorous and evidence-based methods be used to ascertain the community expectations for commercial uses (the social license). This could include deliberative democracy methods such as citizen juries. International evidence generated through such methods suggests an informed citizenry is generally supportive of data sharing that may result in commercial profits if there is also a strong potential for public benefit.

The approach taken by the *Institute for Clinical Evaluative Sciences (ICES)* in Ontario, Canada may be instructive. ICES held a number of public consultation sessions when determining whether to make ICES data available to commercial entities. They resolved that for-profit entities cannot access the data directly, they can only purchase analytic services from ICES. Commercial questions, for example direct product comparisons, are not permitted; the projects must have public benefit. The company is responsible for developing the protocol and analytic plan, which is then implemented by an ICES analyst on their behalf. All commercially-funded research is listed on the ICES website. The company is given 1 year to publish or publically release the analytical results once completed. After that time, all results are made publicly available. This ensures that all relevant evidence is available to policymakers and other interested parties.

The NHMRC National Statement on Ethical Conduct in Human Research also offers guidance – the approving body must be satisfied that "the possibility of commercial exploitation of derivatives of the data [or tissue] will not deprive the participants of any financial benefits to which they would be entitled".

We recommend the data sharing principles identify commercial products that have no recognised public benefit, such as cigarettes/tobacco, and ensure the related commercial entities have no direct or indirect access to public sector data. Further, we recommend the social license for sharing data with commercial entities that profit from products that are potentially harmful or of questionable public benefit (e.g. alcohol, gambling) be ascertained as a matter of urgency.

7. Do you think the Data Sharing Principles acknowledge and treat risks appropriately? When could they fall short?

We believe the Data Sharing Principles acknowledge and treat risks appropriately. They could fall short if they are not assessed consistently and collectively, resulting in the prevention of public good research, and negative societal impacts.

They could also fall short if the protections applied to the data void their utility. How specific will the guidance be for "appropriate protections"? Will researchers have the opportunity to nominate protections that will maintain utility? We strongly encourage that the issue of data de-identification (specifically data modification and data reduction) vs. utility be clearly and fully addressed in the legislative reforms. If the Data Sharing Principles are applied holistically, data de-identification beyond the removal of direct identifiers may not be needed. The Best Practice Guide to Applying Data Sharing Principles (15 March 2019) articulates this issue well, *"In datasets where confidentiality may be a concern, most of the analytical outputs created by users will protect the data to some degree (for example, produce a table, regression, model, summary, etc.). In these cases it may only be necessary to remove the direct identifiers since confidentiality and privacy concerns are able to be controlled by the application of the Project, People, Settings and Output Principles."*

The 2017 *De-identification Decision-Making Framework* report by Professor Christine O'Keefe et al (https://publications.csiro.au/rpr/download?pid=csiro:EP173122&dsid=DS3) discusses this issue in detail, but is not referenced in the Legislative Reforms Discussion Paper. We would like to highlight some of the directly relevant points from the O'Keefe report:

1. Page iv: *"Data custodians should therefore be aware that the application of environmental controls, which go to the 'who', 'what', 'where', and 'how' of accessing data, may be more effective at reducing the risk of re-identification than modifying the data itself, and with a lower impact on the utility of the data.*"
2. Page 5: "*De-identification is a process to produce safe data but it only makes sense if what you are producing is safe useful data: You may wonder why we talk about the need to balance data utility with data safety in the de-identification process. It is easy after all to think about de-identification only in terms of producing safe data but if you do that you may well be taking a risk for no benefit. "*
3. Page 5: *"On the issue of data utility – there is little point in releasing data that does not represent whatever they are meant to represent. There are two possible outcomes that arise from low utility and neither are happy ones: - the data is of little or no use to its potential users and you will have wasted your time and resources on them, or - the data could lead to misleading conclusions which might have significant consequences if, for example, the data is used to influence policy or to make decisions."*

We recommend the issue of data de-identification vs. utility be explicitly addressed as it may help data stewards shift their focus from applying all data controls equally, to prioritising some over others, which in turn should help maintain an acceptable level of data utility. Otherwise, while the legislative reforms may improve data sharing, the shared data may not meet the standards for high quality research, and may lead to inaccurate evidence and potentially harm.

The discussion document notes that additional protections will be set for sensitive data in a binding **Sensitive Data Code**. The definition of 'sensitive data' includes 'sensitive information' as defined in the *Privacy Act 1988.* The definition of sensitive information in the Act:

(a) Information or an opinion about an individual's
- *(i) Racial or ethnic origin*
- *(ii) Political opinions*
- *(iii) Membership of a political association*
- *(iv) Religious beliefs or affiliations*
- *(v) Philosophical beliefs*
- *(vi) Membership of a professional or trade association*
- *(vii) Membership of a trade union*
- *(viii) Sexual orientation or practices*
- *(ix) Criminal record, or*

(b) Health information about an individual, or
(c) Genetic information about an individual that is not otherwise health information, or
(d) Biometric information that is to be used for the purpose of automated biometric verification or biometric identification, or
(e) Biometric templates

Given the breadth of the above inclusions, including all health information about an individual, we are keen to contribute to the consultation. Importantly, some of the above data types and population subgroups (e.g. ethnicity, incarceration records, health records) are currently being shared for public interest research, with appropriate safeguards. We are concerned that a sensitive data code may be a backwards step, as these data are critical to widely performed analyses in the public interest.

We note reference to the National Indigenous Australians Agency in the Discussion Paper. We look forward to learning more about the detail of this important work. RADiANT acknowledges the benefits for Aboriginal and Torres Strait Islander people, families and communities that could arise from the sharing of data about them. RADiANT also acknowledges the particular risks to Aboriginal and Torres Strait Islander people, families and communities that could arise from

(i) their over-representation in disadvantage data,
(ii) their distance from decision making about the collection and use of data about them,
(iii) the incorrect use of predictive algorithms, and
(iv) their social and economic position.

These risks could be mediated by Indigenous data governance mechanisms that help to achieve collective benefits and a cultural and social licence while reducing disadvantage and stigma.

8. Is the *Best Practice Guide to Applying Data Sharing Principles* useful? Are there areas where the guidance could be improved?

The Guide is useful. We believe the guidance about the roles of data custodians and ethics committees could be strengthened.

Importantly, the NHMRC National Statement on Ethical Conduct in Human Research defines "human research" broadly, including the "involvement of human beings through …. access to their information (in individually identifiable, re-identifiable or nonidentifiable form) as part of an existing published or unpublished source or database". This definition encompasses all research that will be undertaken using shared unit record government data.

The National Statement must be applied to any research that is funded by, or takes place under the auspices of, any of the bodies that have developed it (NHMRC, ARC and Universities Australia). However, the National Statement also "sets national standards for use by any individual, institution or organisation conducting human research. This includes human research undertaken by governments, industry, private individuals, organisations, or networks of organisations." It is essential that the Guide provide explicit advice to data custodians regarding how and when the National Statement is applicable. It will be highly problematic if different policies, processes and levels of scrutiny are in place for research that is conducted by researchers based in Universities and those based in government or industry.

Another key issue is whether and when human research using shared data that is conducted in accordance with the Data Sharing Principles can be classified as "low risk", which is defined in the National Statement as research "where the only foreseeable risk is one of discomfort". Arguably, proper application of the data sharing principles, in particular the "Setting" and "Output" Principles can mitigate the risk to human participants, which relates primarily to breach of confidentiality, to a "low risk" level.

We suggest that the ONDC and the NHMRC jointly consider these issues, which may best be clarified through amendments both to the Guide and the National Statement.

We note that generally speaking, ethics committees are not well equipped to deal with data privacy and data governance issues. Ethics committee members need to assure themselves that the applicants are aware of and will adhere to the Data Sharing Principles. The NHMRC and ONDC could ensure training is available and taken up by ethics committee members tasked with making decisions about research requiring data sharing.

9. Do the safeguards address key privacy risks?

We believe the safeguards broadly address the key privacy risks.

We understand that the safeguards include (i) legislative safeguards, (ii) the option for informed consent, (iii) offenses and penalties, and (iv) the risk management controls applied to each request for data sharing. These controls include:

- Restricting data access to projects or programs that meet the purpose test
- Restricting data access to accredited individuals
- Controlling the environment in which the data is shared
- Applying protections to the data
- Restricting data outputs

We recommend that there be greater information about the nature and extent of the privacy protections in a manner that is understood by the general public. This should include sufficiently detailed information about the types of environments in which the data is shared, and the controls in place. Similarly, the nature and impact of data protection steps, and the nature of data outputs. We recommend a centralised, standardised national approach to these controls for Commonwealth data at a minimum, directed and managed by the ONDC. The controls need to meet the public expectations and thus be generally predictable.

10. Are the core principles guiding the development of accreditation criteria comprehensive? How else could we improve and make them fit for the future?

The core principles guiding the accreditation criteria appear sufficient.

11. Are there adequate transparency and accountability mechanisms built into the framework, including Data Sharing Agreements, public registers and National Data Commissioner review and reporting requirements?

We agree Data Sharing Agreements should be published, and we are generally supportive of the proposed mandatory terms of the Agreements.

We are supportive of the notion of public registers of accredited data users and accredited data service providers.

Reporting to the National Data Commissioner by data custodians, accredited data service providers and accredited users is proposed. We recommend that the reporting of this information be streamlined within an integrated, shared digital application and governance system. Reporting of public benefits, in lay terms, is critical to maintaining a public license for data sharing.

12. Have we achieved the right balance between complaints, redress options and review rights?

Further detail is required. Particularly in relation to the timelines.

We recommend the legislation provide for merits review of data sharing decisions by Data Custodians (for example, decisions to share, the conditions of sharing [such as data protections], or decisions to deny access; p51). Stating that Data Custodians universally have "a greater understanding of the risks and benefits of sharing data" appears to be at odds with the findings of the Productivity Commission and other sections of the discussion document (e.g. 'inconsistent safeguards and standards' p1; acknowledgement that it will take time for the data system to mature and for Data Custodians and others to confidently use the new system p53). It is also at odds with achieving a data sharing ecosystem that genuinely reflects community expectations, rather than those of an individual.

13. Have we got our approach to enforcement and penalties right for when things go wrong? Will it deter non-compliance while encouraging greater data sharing?

Further detail is required.

14. What types of guidance and ongoing support from the National Data Commissioner will provide assurance and enable safe sharing of data?

We agree with compulsory training (and testing) for data users. Depending on the operationalisation of the legislation, compulsory training (and testing) for Data Custodians may also help them apply the Data Sharing Principles.