

WA Data Linkage Branch

Response to the Issues Paper for New Australian Government Data Sharing and Release Legislation

Submitted 1 August 2018

WA DATA LINKAGE BRANCH SUBMISSION ON THE ISSUES PAPER FOR NEW AUSTRALIAN GOVERNMENT DATA SHARING AND RELEASE LEGISLATION

This submission from Western Australia's Department of Health (WADoH) Data Linkage Branch (WADLB) highlights a number of considerations that will improve the completeness and value of this public consultation process, particularly as it relates to data linkage. WADLB has over 20 years' experience in this highly specialised field, which it shares here in the interest of building national data linkage infrastructure that complements and strengthens (not duplicates) the significant work that jurisdictional data linkage agencies have already invested in this space. The opinions expressed here are those of the WADLB and do not necessarily represent the views of the WADoH.

The WADLB comments that follow are underpinned by these broad views:

- Secondary analysis of routinely collected electronic government administrative datasets is a powerful and cost-effective means of fulfilling a number of evidence-based information needs (policy, planning, purchasing, evaluation, quality assurance, clinical decision support, innovation and research).
- Government administrative datasets become even more useful when data linkage is used to join together records within or between different data sources that are thought to belong to the same entity (i.e., person, place or event).
- Big Data comes with big challenges, particularly insofar as efficiency, cost, quality, scope, scalability, risks (especially privacy) and benefits. While fast and simple solutions are desirable, they are not always possible or produce the best long term outcomes.
- Current access to Commonwealth administrative data sources for data linkage purposes is labyrinthine and time intensive for various reasons. Some of these reasons stem from legislative impediments (absence, inconsistency and/or density), both at a State and Commonwealth level.
- The development of *National Data Sharing and Release (DS&R) Legislation* and appointment of a National Data Commissioner are both important developments that have the potential to grow Australia's data linkage infrastructure for approved purposes while protecting the privacy of individuals.

1: Overview: WA Data Linkage System and Data Linkage Branch

WADLB builds and operates a data linkage infrastructure (the WA Data Linkage System, WADLS) that is capable of securely linking data collections from a wide range of sources to enable a variety of approved activities including policy, planning and research. From its modest beginnings in 1995, the WADLS is now one of the most comprehensive, high quality and enduring linkage systems worldwide. It currently comprises a linkage infrastructure spanning approximately 50 data collections, representing 115 million linked records dating back to 1945. Of these, 103 million records are linked to at least one other record in the WADLS. Additionally, approximately 400 'ad-hoc' datasets have been linked against the WADLS to service project-specific data requests. Health linkages form over half of the WADLS. This focus is largely attributable to WADoH's established track record of collecting high quality electronic administrative data, driven by regulatory requirements and WADoH's need to support a wide range of approved medical and health-care functions (e.g., patient safety

and quality, disease surveillance, population health management, clinical decision support, cost effectiveness, service purchasing, planning and performance management, research and innovation).

Over a decade ago, WADLB also pioneered the linkage of Commonwealth datasets involving the Medicare Benefits Schedule (MBS), Pharmaceutical Benefits Scheme (PBS) and residential aged care. Although successful on a proof-of-concept basis, and despite WADLB's interest and capability of progressing these data linkages on an enduring basis, the Commonwealth withdrew this function from WA. The Commonwealth subsequently introduced the concept of Accredited Integrating Authorities (AIA) to undertake 'high risk data integration projects' involving Commonwealth data. WADLB contends that improved understanding of the basis of this AIA approach is warranted before it is adopted or built upon as part of the new *Australian DS&R Legislation*.

While WADLB generally supports the principle of an accreditation process to credential data linkage agencies as being trustworthy, technically competent and safe, WADLB also advocates for the new *Australian DS&R Legislation* to:

- ensure any AIA requirements are: 1) desirable (adds value); 2) feasible (is the best option available); and 3) viable (sustainable and fits with the way stakeholders want to use linked data);
- include more representative and expert involvement (i.e., jurisdictional data linkage agencies) in establishing AIA requirements and the assessment of AIA applications than exists currently;
- assess if (and on what grounds) Commonwealth data is truly more 'high risk' than the state-based data collections that WADLB (and possibly other jurisdictions) link without AIA status (e.g. WA HIV Register, WA Notifiable and Infectious Disease Register, WA Mental Health Information System, WA Police, Courts and Corrective Services);
- mitigate any conflicts of interest (real or perceived) in undertaking the aforementioned functions.

2: Models of data linkage: centralised -vs- distributed

Traditionally, models of data linkage have involved a single data linkage agency/node, otherwise known as a centralised linkage model. This is how most Australian jurisdictional linkage agencies operate, including WADLB until recently. This centralised approach also currently underpins the newly established *National Integrated Health Service Information Analysis Asset (NIHSI AA)*, led by the Australian Institute of Health and Welfare (AIHW). However, there is increasing interest in distributed linkage models that build integrated data linkage infrastructure by using a common 'spine of data' that is shared by a number of linkage agencies. Currently, the Population Health Research Network is finalising a report developed in consultation with the Commonwealth and jurisdictional data linkage agencies that supports a distributed linkage model as the preferred long-term basis for establishing a *National Master Linkage Key*.

By way of example, WA is currently pioneering a distributed linkage model as part of a large, state-based project called *Target 120*, which aims to reduce juvenile crime and improve community safety. It involves sharing, linking and analysing data from a range of WA-based data collections to inform efficient and quality decision making. Both WADLB and another data linkage agency are undertaking the data linkage work by establishing a spine via a small number of common WA administrative datasets (chosen to provide the best population coverage for the cohort). Importantly, some of these linkages will be established using fully identifiable person information shared by the Data Provider(s) with the linkage agencies, whereas privacy preserving record linkage (PPRL) will be used in other circumstances¹. Subsequently, the content/service data will be merged via spine record identifiers and loaded into a secure repository (*the Social Investment Data Resource*) for access and analysis.

These data linkage developments highlight a number of important considerations for the new *Australian DS&R Legislation*:

- while Commonwealth data linkage agencies have ‘strong experience in data curation, collation, linkage, de-identification, sharing and release’, so too do jurisdictional data linkage agencies such as WADLB;
- there is no one-size fits all approach to data linkage, nor should there be;
- legislation should cater for the secure bi-directional flow of data to jurisdictional data linkage agencies and not just from jurisdictions to Commonwealth agencies (such as AIHW and the Australian Bureau of Statistics);
- special care is needed to avoid legislative enablers for some data linkage agencies becoming legislative barriers for others;
- the ‘gold standard’ of data linkage (insofar as linkage quality and stability) requires the exchange, storage and use of person identifiable information between data linkage agencies (e.g., names, addresses, dates of birth);
- where the exchange of person identifiable information between data linkage agencies is problematic due to legal or other restrictions, privacy preserving record linkage techniques should be considered;
- work involving Commonwealth and jurisdictional stakeholders is needed to develop an agreed and interoperable *Information Classification Framework* that consistently describes the sensitivity and risk associated with State and National datasets;
- an explicit list of personal identifiers, perhaps modelled on the US *Health Insurance and Portability and Accountability Act (1996)*, will lead to improved understanding of the difference between identified and de-identified data;
- potential consideration of a *National Anonymisation Framework*, perhaps modelled on the United Kingdom approach;
- criminal penalties for those who purposely re-identify individuals or link to other datasets (regardless of success) without prior approval.

3: Governance of data sharing arrangements

¹ PPRL uses masked encoded quasi-identifiers to preserve privacy when the ‘gold standard’ of sharing person identifiable information is problematic.

In WADLB's experience, governance occurs at two levels:

- with the Data Provider(s); and
- with the Applicant seeking access to linked data.

Data sharing with Data Providers

WADLB coordinates approximately 50 Data Sharing Agreements (DSAs) with Data Providers. The majority of these relate to WA data collections, but they also include Commonwealth data such as the *Australia Early Development Census*, plus the Sax Institute's Secure Unified Research Environment. Our experience suggests that while standardised templates and use of the *Five-safes Framework* may provide a useful starting point, flexibility (and sometimes additional complexity) is needed to achieve DSAs that satisfy all parties, noting this can change over time (e.g., due to a change of personnel or legislative amendment). Some of WADLB's DSAs are non-legally binding Memoranda of Understanding based on goodwill, whereas others are legally binding agreements; this decision is informed by legal opinion from at least one of the involved parties. Additionally, WADLB's default position is not to make these DSAs publicly available; in some cases this is due to legal privilege. Importantly, DSAs can take many months to finalise while each party seeks professional advice before signing and agreeing to data sharing.

WADLB's DSAs typically include:

- the relevant legislation that applies to the DSA;
- the approved purpose and scope of the DSA;
- the nominated parties and contact details;
- obligations of the nominated parties, particularly regarding data security, transmission, storage, retention, use and disclosure of data;
- the data items being shared (for both linkage and extraction of linked content/service data);
- DSA commencement and termination dates;
- grounds for modifying or terminating the DSA;
- the application requirements for others seeking access to the linked data;
- exchange of linkage keys and other WADLB value adds (e.g., geocoding) with the Data Provider;
- indemnity arrangements.

Data sharing with Applicants

WA has a formal application process for linked data that comprises two main stages:

1. Data Custodian Engagement incorporating a Draft Application for Data

This includes expert advice from Data Custodians pertaining to the data collection(s) (including data item selection, quality and availability), technical feasibility of the data linkage, any special project requirements (e.g., cohort or control selection criteria), plus data governance considerations (e.g., legislative impediments, privacy impact assessment, potentially contentious issues such as overseas data release).

2. Formal Application for Data

This includes review and approval by the WADOH specialised Human Research Ethics Committee (HREC)², followed by formal Data Custodian and Research Governance approval. i.e., Scientific and ethical approval by HREC, individual declarations of confidentiality by the Applicant and associated staff, assessment and management of site risk (including resourcing), plus a declaration from the institution(s) hosting the project. The documents collected as part of the *Formal Application for Data* process act as the DSA with the Applicant.

These WA experiences are of relevance to the new *Australian DS&R Legislation* because:

- data linkage and data release more generally operates within a legislatively and policy dense environment, especially when it occurs without informed consent;
- there is no simplified one-size fits all approach to data governance, not least due to the differing sensitivity of data collections and applicable legislation;
- jurisdictions such as WA already operate with robust and mature data sharing arrangements, and so it encourages a complementary national approach.

4: Linked data repositories

Traditionally, the linked data requests received by WADLB have involved bespoke extracts for specific and time-limited projects. Over time, the size and complexity of these WA projects have expanded to the point that bespoke extracts can be time consuming to create (and update), costly to maintain, and sometimes share significant similarity/overlap with other projects. This has generated increased interest in the establishment and maintenance of enduring linked WA data repositories³ that can potentially service a number of projects, reduce data delivery timeframes and support easier and faster analysis and reporting. Currently, a number of WA stakeholders intend to develop enduring data repositories for various cohorts of interest (e.g., cardiovascular disease, Aboriginal children, and juvenile offenders). To date, WADLB has progressed these via the existing data sharing governance arrangement described previously (see Section 3). It is reasonable to expect that future data repositories will expand to request linked Commonwealth datasets such as the MBS and PBS.

While linked data repositories may be appealing (particularly from a user perspective), they are accompanied with a number of challenges that are likely to be relevant to the new *Australian DS&R Legislation*. These issues are at times counter to traditional WA considerations of linked data governance and include:

- the social licence regarding the acceptability of this approach has not been established;
- best practices for data management of repositories has not yet been established insofar as roles and responsibilities (e.g., stewardship), regulation (e.g., administration, data sensitivity classification, identifiability, data ownership,

² The WADoH HREC is a specialised health data HREC that comprises technical expertise and ongoing training and development of members in the area of linked and unlinked health data research.

³ The term 'data repository' is used here in the general sense, rather than in any technical sense that distinguishes it from 'data lakes', 'data warehouses' and 'data marts'.

- approved uses, approved users, security and retention periods) and compliance monitoring;
- there are potentially bigger security considerations when linked data is in a repository that can be accessed and shared by many;
 - the size and scope of some repositories will be inconsistent with WA's principle of releasing the minimum amount of data required for a specific purpose;
 - different linked data repositories are likely to contain some overlapping content, which creates multiple sources of truth that will potentially differ to the data source;
 - enduring repositories challenge the standard practice of data disclosure for temporary use over a defined period (followed by data destruction);
 - repositories require the release of data prior to review and approval of all possible uses, which may encourage data mining without any public benefit and possibly raise privacy concerns;
 - there is a risk that stakeholder demand for linked data repositories will outpace the development of best practices in this space (refer to dot point 2).

WADLB encourages multi-sector involvement in establishing a best practice for creating, maintaining and working with data repositories (linked or unlinked) that is compatible with the *Five-safes Framework*. This should also acknowledge that current national developments such as the *NH&MRC AA* are not the only approach to achieving this end.

5: Charging

It costs to have a linkage system the size and complexity of the WADLS, and 'big data' more generally. Workforce, equipment, software, and secure facilities, etc are not cost-trivial considerations. To this end, DLB accepts that charging for technical expertise is currently a necessary requirement in the absence of adequate recurrent funding. The new *Australian DS&R Issues Paper* requires clarification insofar as jurisdictional data linkage agencies who become 'Accredited Data Authorities'. Specifically, will they be treated as 'Commonwealth entities' (i.e., bound to comply with the *Australian Government Charging Framework*)? If confirmed, this charging arrangement could prove difficult if it differs materially to jurisdictional charging instructions.

6. Responsibilities of the Data Sharing Commissioner

A number of questions posed in the new *Australian DS&R Issues Paper* seem to overlap with the *Privacy Act (1988)*, which regulates the handling of personal information about individuals using established and ongoing principles. For example:

- the *Privacy Act (1988)* allows applications to the Commissioner⁴ to assess whether a particular use of data represents an infringement on privacy rights;
- the Commissioner can consider National Security and Law Enforcement issues, taking into account special situations that might exist in an emergency or disaster;
- the Commissioner can impose sanctions for improper data sharing;

⁴ Australian Information Commissioner or Privacy Commissioner.

- Sections 95 and 95A of the *Privacy Act (1988)* allows the National Health and Medical Research Council to work with the Commissioner to publish guidelines for research and privacy principles for the use of data in medical research.

Additionally, WADLB highlights these considerations for the new *Australian DS&R Legislation*:

- the security and protection of personal information are of prime importance to data use of any size and complexity;
- while powerful (when understood and implemented correctly), de-identification is not necessarily the best and only solution to protect privacy;
- clarity is needed about what privacy and/or DS&R legislation prevails when linked data products include Commonwealth, jurisdictional and possibly other data sources (e.g., survey data collected by a Researcher);
- data linkage solutions are often highly bespoke for a particular use, and so a principles-based approach to privacy and DS&R may be more beneficial than a prescribed set of rigid rules;
- care is required to avoid duplication of function between the National Data Sharing Commissioner, the Office of the Australian Information Commissioner and jurisdictional equivalents (where these exist), otherwise it risks adding more (not less) red tape to data sharing and release;
- multi-sector collaboration and decision-making is paramount to the growth, regulation and sustainability of Australian data linkage innovation.

7: Conclusion

WADLB is generally supportive of the new *Australian DS&R Legislation*, however, significant jurisdictional investment, interests and processes may be weakened if the issues raised in this submission are not considered and resolved through multi-sector collaboration and decision-making. Additionally, special care is needed to avoid the functions of the new National Data Commissioner from overlapping with the information management and privacy regulation functions of the Office of the Australian Information Commissioner.