



## Attachment 5: Perceptions of Data Linkage: Examples

Perception	Clarification
1. WA's position as a leader in data linkage is declining.	<p>The Department of Health (DoH) supports healthy competition as a way of producing better results. However, it recognises there are currently no formally established or accepted criteria that allow meaningful and reasonable comparisons of 'data linkage' nationally or internationally.</p> <p>DoH also recognises that competition for research funding to support linked data projects has increased nationally, as a result of:</p> <ol style="list-style-type: none"><li>1. the intentional growth of linkage centres nationally;</li><li>2. the reduction in national funding for medical research (e.g. NHMRC).</li></ol> <p>It is important to avoid conflation of these separate issues.</p>
2. DoH Data Custodians and DLB obstruct requests for linked data.	<p>DoH supports the use and release of linked data for approved purposes. Linked data has been released for over 500 requests since 2010.</p> <p>The Data Linkage Branch (DLB) and Data Custodians are tasked with multiple responsibilities, including data governance. Rejection of linked data requests is rare, and to date has occurred for any of the following reasons:</p> <ol style="list-style-type: none"><li>1. legal advice from DoH Legal and Legislative Services and/or State Solicitor's Office advising data release was unlawful;</li><li>2. significant data quality issues that made it unfit for the proposed purpose, based on expert advice from Custodians (e.g. incompleteness, no agreed definitions); and/or</li><li>3. lack of DoH HREC approval (e.g. inadequate research merit).</li></ol> <p>In all situations, DoH strives to reach a reasonable compromise that allows data release.</p>
3. DLB has unlimited capacity.	<p>This is an unrealistic expectation for any business, unless matched with unlimited resources. DLB is not immune to WA and national fiscal austerity measures. Despite these challenges, DLB has built the broadest linkage system in Australia, with over 1000% growth since its inception (4 datasets in 1996 compared with 47 in 2015). Additionally, DLB is developing a new data linkage system that shows promising efficiency gains. This occurs alongside the continued delivery of high quality data linkage products and services, in an environment marked by increasingly complex projects.</p> <p>DLB uses intentional capacity and demand strategies to ensure it knows how capacity is being used and where improvements can be made (within existing constraints) to maximise efficiency. DLB, like other vital components of WA's data linkage infrastructure, would benefit from increased and sustained funding.</p>



	Perception	Clarification
4.	Linked data is unreasonably expensive.	<p>It costs to have a linkage system the size and complexity of the WADLS, and 'big data' more generally, e.g. The Secure Unified Research Environment (SURE) at the Sax Institute has attracted similar comments about perceived expensiveness.</p> <p>DLB's 2016/17 budget is ~\$2.2M, which has remained virtually unchanged for several years. Historically, DLB has recouped 5-10% of its operating costs through cost recovery. The median cost of linked data projects invoiced during 2010 – 2015 was only \$6,054, noting DLB experienced significant cost shifting during this time. Consequently, increased financial prudence was required to continue DLB's service excellence.</p> <p>To this end, DLB reviewed its charging policy in 2015; the first time in over ten years. Key features of the revised policy include:</p> <ul style="list-style-type: none"><li>• it remains <u>partial</u> cost recovery;</li><li>• it now applies to <u>all</u> linked data requests, including DoH.</li><li>• It includes charges for <u>all</u> key DLB products and services, such as: (i) the application, access, data coordination and delivery process; and (ii) streamlined data extraction from CARES.</li></ul> <p>DLB's revised charging policy ensures fairness, links charges to demand, and bridges the gap between other funding sources. Note, as a transition step, DLB has honoured all past cost recovery estimates provided under the previous policy.</p>
5.	Timeframes for delivery of linked data products are unreasonably long.	<p>Linking data and the provision of linked data products are complex activities. Timeframes are highly variable due to a variety of considerations. Resourcing and project complexity are significant challenges. DLB's total project complexity burden has increased significantly since 2010. In 2015, one third of all requests submitted to DLB were 'highly complex', which has quadrupled since 2010. See also Attachment 9.</p> <p>Frequently, stakeholders expect highly complex, cross-jurisdictional and proof of concept projects to take the same processing time as 'standard' linked data projects. This is not always achievable. Cross-jurisdictional data requests require significantly more time to set-up and approve. Proof of concept projects, by definition, break new ground, including the establishment of new processes and demonstration of feasibility. Therefore, it is reasonable these projects may require more effort (including time) from all stakeholders.</p>
6.	The application for linked data process is repetitive and time consuming.	<p>The process of applying for linked data comprises three phases: (i) draft application review; (ii) DoH HREC review (where relevant); and (iii) Custodian approval and data delivery. These phases have separate purposes that are underpinned by: (i) Custodian in-principle agreement; (ii) data linkage feasibility; and (iii) ethical approval.</p> <p>The draft application review process has been in place since 2008. It benefits all linked data stakeholders by providing a single pipeline that</p>



Perception	Clarification
	<p>facilitates information sharing. DLB notes that applications proceed faster through the application phase if Data Applicants engage with DLB and/or the Data Providers early in the process. The modular Draft Application Form is similar to the approach used by other linkage centres. It is reviewed annually by DLB and Custodians. Detailed information on the current draft application process can be found elsewhere.<sup>1</sup></p> <p>The value of the draft application review should not be underestimated. The process serves several functions that cannot be measured exclusively using time, notably:</p> <ul style="list-style-type: none"><li>• sharing knowledge of DoH data, linkage methods and processes;</li><li>• assisting Data Applicants to formulate a robust application;</li><li>• improving the likelihood of first-time ethical approval (DoH HREC reported that <u>no</u> data linkage project required resubmission in 2015).</li></ul> <p>It has previously been suggested to DLB that the draft application process should be dismissed. DLB strongly advises against this, since to remove it would add significant delays for Data Applicants overall and also lead to additional workload for DLB, Data Custodians and DoH HREC.</p> <p>There is also a frequent misperception that Custodians approve each project twice – once at draft review and again following DoH HREC approval. This is incorrect. The draft review process involves Custodian consultation to ensure the project is feasible and the data requested is fit for purpose. The Custodians are unable to give formal approval until DoH HREC approval is provided. By then, Data Custodians are already familiar with the project and so approval is typically fast.</p> <p>DoH HREC approval has been streamlined in the last five years and further changes may occur in light of the Health Services Act 2016 and related health reform work. DoH recognises that ethical review and approval is a key component of research using linked data. Under previous iterations of DLB's access policies, evidence of HREC approval was required (but not always the DoH HREC). By 2014, over 90% of linked data requests required DoH HREC approval. To reduce the burden of seeking multiple ethics approvals for 'multiple-centre' linked data projects, DLB formally changed its access policy in 2014 to allow single ethical review by DoH HREC.</p> <p>WA Health ethics processes have also been streamlined with introduction of the <a href="#">WA Health Research Governance and Single Ethical Review Standard Operating Procedures</a>. This development benefits clinical research within Health Services. Other linked data users are</p>

<sup>1</sup> <http://www.datalinkage-wa.org/access-and-application/application-process/custodian-engagement-draft-application>



Perception		Clarification
		expected to benefit when DoH HREC becomes a 'Lead WA Health HREC' under the aforementioned policy.
7.	Data linkage application forms are long and duplicate other forms.	<p>DLB maintains the modular DoH 'Application for Data Form' (the Form). It is used for both linked and unlinked requests to reduce confusion for Data Applicants. The Form was created before the inception of DoH HREC and prior to the introduction of the WA Health Research Governance Policy. It is similar to the form(s) used at other linkage centres and is reviewed annually by DLB and Data Custodians.</p> <p>As noted in the main submission, data linkage is a complex activity. Every project needs a blueprint to ensure DLB and Data Custodians have sufficient information to progress and design specialised linked data products that are fit for purpose. The information for this blueprint is often different to what DoH HREC requires, although some overlap exists. Progress in electronic application forms and interoperable systems may streamline future processes.</p>
8.	DLB is responsible for resolving all data linkage challenges and stakeholder constraints.	Data linkage is a complex field that depends on many stakeholders to achieve success. DLB, as the pioneer of data linkage within WA, has been the 'go-to' source for troubleshooting issues along the development journey. DLB (and DoH) has no authority to change, prescribe, override or hasten the processes of other data linkage stakeholders. e.g. DLB cannot bridge the gap when in-principle support from a Data Provider does not translate into actual data delivery. As listed in the recommendations (Section 4), DoH welcomes the development (and resourcing) of stakeholder roles and responsibilities, coupled with criteria for assessing the linkage 'readiness' of Data Providers.



	Perception	Clarification
9.	Linked data products and services are technically, logistically and cost trivial.	<p>Complexity is unavoidable in providing a multi-faceted data linkage service the size and scope of the WADLS. It requires significant investment, expertise and IT infrastructure, as evidenced in Attachment 11.</p> <p>DLB acknowledges a range of different linkage strategies worldwide. DLB uses methods of high quality probabilistic linkage that are appropriate to the WA setting. Namely, DLB uses identifying fields for linkage, noting no universal person identifier exists across multiple agencies. DLB's new data linkage system (DLS3, nearing completion) will bring considerable efficiency gains to the review of 'potential' links, which is currently a time intensive activity (and not unique to DLB).</p> <p>To complement the WADLS, DLB offers a range of other unparalleled data linkage capabilities to support highly complex projects. This includes 20+ data collections for multiple cohort and matched control groups, multi-generational genealogical links, and unique value-adds (e.g., geocoding). These services also require significant investment, expertise and IT infrastructure.</p> <p>DLB uses a variety of strategies to educate stakeholders (new and old) about the complexity of linking data. e.g. Data Linkage website, DLB's researcher training workshop, external courses such as The Analysis of Linked Data short course (UWA), and the draft application review process.</p>
10.	The WA Government Open Data Policy allows unrestricted access to all linked data.	<p>DoH supports the WA Government Open Data Policy. Consistent with this policy, DoH also recognises that some data is sensitive and potentially inappropriate for wider disclosure, particularly where there is a risk of individual re-identification by combining it with other datasets (existing and future). Anonymisation failures occur. Examples are not hard to find and include: (1) the Netflix data breach; and (2) re-identification of 87% all Americans in 2000 using only three pieces of information (post/zip code, date of birth and sex).</p> <p>Significant work remains to expertly assess the current state of privacy protection for de-identified data, especially in light of rapid advancements in re-identification technology, the speed of modern computers, and the amount of outside information that people volunteer on so-called 'anonymised' databases.</p>
11.	Linked data is inaccessible to other government agencies	<p>Linked data is available to any group or person that fulfils the requirements of the DLB Application and Charging Policy and relevant Custodian and ethical approvals.</p> <p>Data Custodians can request links within their own dataset following linkage, as formalised in a data exchange agreement with DLB. This free service is not used by all Government Departments that contribute data to the WADLS.</p> <p>Linked data can be requested from other Government Departments;</p>



Perception	Clarification
	noting these Departments might have their own specific application criteria for access. e.g., Access to linked data from the Departments of Corrective Services, Attorney General (courts) and WA Police is limited to the Developmental Pathways Project, and access to Department of Transport Drivers' Licence data is restricted to road safety related research.