Private & Confidential

Privacy Impact Assessment

Release of certain Opal Datasets in the context of the NSW Open Data policy

Transport for NSW

Draft 9 December 2016 – updated for Data 61 Privacy Preserving Opal Data Report

Contents

Executive Summary

Overview

This report sets out the findings of the privacy impact assessment (**PIA**) conducted by Holding Redlich of the proposal to release certain Opal Data sets under the NSW Open Data Policy (**ODP**) (the **Project**).

This report includes recommendations to mitigate the privacy risks that were identified during the PIA as applicable in the use, storage and disclosure of Personal Information relating to the Opal customers and in the context of ODP.

This PIA Report reviews the privacy impacts of disclosure and use of Opal Data for ODP generally and considers the particular issues raised in the context of ODP.

This PIA was conducted to ensure that the broader use of Opal Data meets the following objectives:

- ensure that TfNSW's privacy obligations are met by examining those processes against the relevant information protection principles (**IPPs**) and health privacy principles (**HPPs**);
- ensure no Personal Information is released;
- identify and assess any potential privacy risks in disclosure;
- identify privacy enhancing features and measures and incorporate these into data disclosure parameters including specifically de-identification and anonymization techniques;
- implement strategies to reduce and mitigate the likelihood of a privacy breach occurring in the form of subsequent re-identification; and
- ensure information is retained in accordance with the requirements of the *State Records Act 1998* (NSW) and TfNSW record retention and disposal authorities.

TfNSW has obligations under the *Privacy and Personal Information Protection Act 1988* (NSW) (the **PPIP Act**) and the *Health Records and Information Privacy Act 2002* (NSW) (the **HRIP Act**). Further detail about these obligations as they apply to the Project is set out in **Section 2**.

Evolution of the Project

It is noted that our initial briefing was in relation to the data set known as the Opal Journey Segment View Data and that was subsequently reviewed down to 13 elements that were provided to Data 61 for testing. As a result of the initial testing and subsequent testing of the de-identified data for utility, a further set of data elements was tested and the report updates as at 9 December 2016 reflect this final data set which is referenced in the Data 61 Privacy Preserving Opal Data Report published on 15 November 2016.

The evolution of the project and the reduction in the data set being considered for release is a direct consequence of the privacy mitigation strategies referenced in this report as implemented by data scientists at Data 61 and also the Open Opal Data industry reference group which included members of the University of Technology Sydney and the Data Analytics Centre as set out in the Open Opal Data Project Management Plan December 2016 version 1.

Identified Risks & Recommendations

In applying the relevant IPPs and HPPs, the PIA identified a range of risks related to the Project that can be treated with the controls recommended in this report. **Section 5** of this report analyses these privacy risks and makes recommendations that aim to maximise the Project's compliance with the IPPs and

HPPs, and to promote privacy best practice in the handling of Personal Information by minimising privacy intrusion and maximising privacy protection, including in particular, obligations to destroy or anonymise data. The ODP issues are more fully explored at **Section 4** of this Report.

The table below summarises the major privacy risks identified in the PIA and their corresponding recommendations. We note that this table does not assess or prioritise risks in accordance with the TfNSW Enterprise Risk Management System, as we are outside the business and are not qualified to make this analysis.

Key Issues

As set out in **Sections 2.3** and **2.4** of this Report, the risk analysis being taken in the ODP context is fundamentally different from that where a project is reviewed in the standard data life cycle context. In the ODP context there is one single opportunity for the government to mitigate risk before data is released to the public and at risk of re-identification. Determining the appropriate steps to undertake in order to release that data are matters which combine both legal analysis and technical analysis. There is Australian and overseas guidance on what might constitute best practice in this area and what sorts of treatments might be provided. At **Section 4.3** we provide our recommendation in relation to data testing being undertaken before any further steps to release are pursued. Included in this recommendation and attached as Appendix B to this PIA is a draft service specification for confidential service provider testing of the data to consider the legal issues and to make recommendations as to appropriate testing before the data or a subset of it is released. At **Sections 4.4** and **4.5** we set out a summary of the testing that has occurred and the conclusion that the privacy issues identified have been addressed and that based on the final Data 61 Privacy Preserving Opal Data Report published on 15 November 2016, the data set tested may be released subject to the limitations and risks identified in that report.

In addition to considering testing for full release, we have considered at **Section 4.2** that there are alternate methodologies to complete release into an open data framework which might still allow the ODP objectives to be satisfied but in more limited circumstances. These recommendations also draw their source from Australian and overseas guidelines. For example, if information is considered to be of high risk of re-identification then rather than not making it available it could be made available to a limited number of contractors under a free licence arrangement but with contractual protections built around it and sanctions for those contractors who breached those contractual obligations.

IPP/HPP	Privacy risks	Recommendations
IPP 2/HPP 3 (direct collection)	Collection is not relevant for this purpose other than in relation to open collection and transparency	The TfNSW and Opal privacy policies be updated to disclose the release of some data under ODP
IPP 3/HPP 4 (open collection)	ODP not currently disclosed as a possible use / disclosure	Consider revising both Opal Privacy Policy and TfNSW to clarify ODP use
IPP 5/ HPP 5 (secure)	High-risk if data sets not sufficiently de-identified	Apply industry best practice de-identification techniques before release of any data

IPP/HPP	Privacy risks	Recommendations
IPP 6/HPP 6 (transparent)	IPP6 not relevant	No change to BAU
IPP 8/HPP 8 (correct)	IPP8 not relevant	No change to BAU
IPP 10/ HPP 10 (limited)	IPP10 not relevant	No change to BAU
IPP 12 (safeguarded)	IPP12 not relevant	No change to BAU

This report is intended as an ongoing resource to assist TfNSW to monitor the Project's compliance with the IPPs and HPPs, identify and manage privacy risks and enhance stakeholder confidence in the Project. This report was prepared on the basis of the documents provided by TfNSW to date as well as the publicly available resources in relation to NSW ODP also referenced there, a list of which may be found at **Annexure A**. The report was prepared initially based on the data set known as the Opal Journey Segment View Data.

Where to from here? – Preliminary Conclusion

In our view, until such time as the service specification testing or a similar round of testing has been undertaken by a data analytics provider, no release should be made of any or all of the Opal Journey Segment View Data. Once technical experts have provided their recommendation as to the methodologies and treatments to be applied then an informed decision as to release can be made. At that point it may or may not be necessary to further update this PIA, subject always to the outcomes of the data testing.

Data Testing Update - 6 October Initial Summary Report

The balance of this report has considered the treatment necessary to fully release the set known as the Opal Journey Segment View Data. The initial work undertaken by data analytics provider Data 61 has tested a small subset of that for release, known as the "Minimum Viable Product Example" (**MVPE**) of the Opal Journey Segment View. The data segments that have been tested for release are separate tap-on tap-off data for the modalities of bus, train, ferry and light rail. The data sets tested are separated by mode and there is no capability to obtain a single journey view across modes. Similarly, no hashed CIN is contained in the MVPE. The 13 elements in the MVPE mean that some of the analysis relevant to elements of the Opal Journey Segment View are not relevant to the MVPE.

In relation to the MVPE and that data only, the conclusion of Data 61 in its testing as set out in paragraph 2.4 where the application of a differentially private algorithm gives a guarantee of inability to attribute the trip to a known individual, irrespective of any auxiliary information held, can be relied upon in order to release that MVPE data for the periods as referenced in the Initial Summary Report.

The Report notes at paragraph 5.13 it is not information about a single integrated journey and at section 6 recommends further testing before elements relating to any integrated journey could in fact be released.

On this basis it is our view that the limited tested data sets for the respective modalities of buses, ferries, trains and light rail being the MVPE as tested by Data 61 can be released on the terms set out in that report which is limited to 2 weeks of data and notes that if further data relating to the same data sets is to be released in the future then the application of the algorithm will need to be revisited. A copy of that report is attached as Annexure F.

Data 61 Privacy Preserving Opal Data Report and Addendum November 2016

Further to the testing undertaken in the 6 October Initial Summary Report, testing the utility of that data by the Open Opal Data Industry Reference Group determined that the data had very little utility. Accordingly further discussion was undertaken about what could be done to move to a position of an MVP with sufficient utility as to be worthwhile.

While we were not involved in this process we understand that it resulted in the further Open Opal Data Set which is described more fully in the document headed Open Opal Data Set Documentation December 2016 version 1.0. It involves six CSV files covering 14 days across the four public transport modes and includes tap-on and tap-off time, tap-on and tap-off location and tap-on and tap-off time and location.

There are protections applied to that data in accordance with the original PIA recommendations, in particular the removal of high-risk low-count queries as more fully set out in the Data 61 report. Given that one of the most significant risks identified in the original draft of this PIA was the risk of identification of individuals in the data set through use of geolocation data, the application by Data 61 of its expertise to determine a threshold of 18 entries to remove any low-count queries as a privacy risk appears to be an appropriate response to the identified risks.

Glossary and acronyms

BAU	Business as usual
Health information	Health information is defined in the HRIP Act and includes information about a person's physical or mental health or disability
HPPs	Health Privacy Principles, legal obligations under the *Health Records and Information Privacy Act 2002* (NSW) that apply when NSW public sector agencies collect, store, use and disclose health information
HRIP Act	*Health Records and Information Privacy Act 2002* (NSW)
IPPs	Information Protection Principles, legal obligations (under the PPIP Act) that apply when NSW public sector agencies collect, store, use and disclose Personal Information
Opal Card	A smartcard used in the electronic ticketing system for public transport in the greater Sydney region
Opal Data	Data collected from the Opal Ticketing System which uses a contactless smartcard (**Opal card**) for ticketing and/or payment. Some Opal cards function as an electronic purse from which fares can be deducted and which can be "topped up". The card contains a computer chip which stores value (a dollar amount), limited transaction history and a code that enables the correct fare to be charged. Opal Data also includes the elements specifically mentioned in sections 4.2 and 4.3 of the Opal Privacy Policy.
Opal Privacy Policy	The policy of that name published on the Opal website, cobranded TfNSW print version December 2015
Personal Information	Information (including information forming part of a database and whether or not recorded in a material form) about an individual whose identity is apparent or can reasonably be ascertained from the information (as defined in the PPIP Act)
PIA	Privacy Impact Assessment
PPIP Act	*Privacy and Personal Information Protection Act 1998* (NSW)
The Project	TfNSW's release of certain Opal Data under the Open Data Project
TfNSW	Transport for NSW

1. Privacy Impact Assessment

1.1 What is a privacy impact assessment?

A PIA is a comprehensive assessment of the privacy impacts and risks of a new project that handles Personal Information. A PIA identifies options for mitigating privacy risks, thereby promoting privacy compliance and best practice while meeting project objectives.

PIAs are typically undertaken as part of a sound risk management strategy. A failure to properly embed appropriate privacy protection measures in a project may result in compliance risks and harm later on- including harm to individuals, reputational damage to organisations, and/or legal action as a result of privacy non-compliance.

Best practice models for PIAs include not only an assessment of whether a project complies with legal privacy obligations, but also how a project responds to community expectations about privacy. Notwithstanding compliance with the legal framework that applies to a project, stakeholder expectations about a project can pose unacceptable risks if they are not appropriately managed.

A PIA demonstrates accountability and transparency. It provides a tool to enhance public trust and confidence in the way an organisation handles Personal Information about its customers.

A PIA should be regularly reviewed and updated to reflect project developments and inform project risk management.

1.2 Scope of privacy impact assessment

(a) **In scope**

This PIA is an assessment of the privacy impacts of the disclosure of certain Opal Data, being the data elements extracted from the set known as the Opal Journey Segment View Data Dictionary June 2016 version 0.6.

Systems and processes assessed by this PIA include the technological solutions and business processes for the management of Opal Data to ensure that prior to release under ODP all reasonable steps are taken to ensure no personal information is released.

(b) **Out of scope**

This PIA does not assess:

(i) TfNSW's electronic ticketing systems and processes for handling Personal Information generally;

(ii) Cubic Transport Systems (Australia) Pty Limited's (**Cubic**) systems and processes for handling Personal Information; or

(iii) third party systems and processes for handling Personal Information related to the Project; or

(iv) generic issues relevant to privacy as an indirect consequence of information security. That is where the technical dependencies are said to comply with NSW Government policies for ICT Security and cloud usage. We have not investigated these issues further.

1.3 Methodology

This PIA was conducted by Holding Redlich in May, June and July 2016. The methodology for this PIA is based on guidelines issued by the Office of the Australian Information Commissioner and the Office of the Queensland Information Commissioner.

The following steps were undertaken to prepare this PIA report:

(a) Desktop review of relevant documentation provided by TfNSW including the documents set out at **Annexure A** and various documents referenced therein.

(a) Interviews with staff from TfNSW business units, including:

(i) Dorothy Cheng;

(ii) Nathan Frick;

(iii) Micah Starkis;

(iv) Yvonne Lee;

(v) Alan Willmore;

(vi) Donna Hayward;

(vii) Melanie Kelly; and

(viii) Brett Johnstone.

(b) Analysis of privacy impacts and identification of privacy risks related to the ODP in applying the IPPs and HPPs;

(c) Developing recommendations to mitigate identified privacy risks; and

(d) Preparing a report based on feedback received.

Consultation with external stakeholders has not been undertaken for this PIA. However, given the comments by the Australian Law Reform Commission recently we have included a brief section of managing stakeholder expectations.

1.4 Assumptions and limitations

In addition to section 1.2(b) above, which sets out the out of scope description, there are a number of other assumptions and limitations that apply in this PIA. Essentially, this PIA looks at the risks that arise in releasing the Opal Data. We consider at **Section 5** the nature of the data and the specific elements.

TfNSW uses Opal Data for many purposes and there are many variants of the data that is retained depending on the type of Opal Card that is held by a customer. For example, cards for various concession holders might hold details about health information or other sensitive information that may allow them to be identified which would not be the case with unregistered cards. However, TfNSW also has the ability to identify the holder of an unregistered card through other mechanisms (see Section 5 of the Opal Privacy Policy December 2015 print version).

Taking a conservative approach, in this PIA we have assumed that Opal Data generally will be able to be matched to an individual and so on a default basis, we assume data is, or may be, Personal Information. This then becomes an exercise in ensuring that TfNSW has taken all reasonable steps to de-identify information and prevent re-identification of the Opal Datasets before their release under the ODP.

We were not asked to advise specifically on the Opal Privacy Policy other than to consider if its references to use extend to ODP disclosure. Currently it does not contemplate attributes of Opal Card holder information being published as part of ODP.

We also note by way of limitation that certain documentation provided to us is part of an iterative process and is in draft form when received by us. Accordingly, we have our review based on the documents set out in **Annexure A** and if at a later point in time those documents change in any significant manner, then this may influence and alter our view and our recommendations.

1.5 Sub-contractor assumptions

For the purpose of preparing this Privacy Impact Assessment we have referred at times to the third party Opal administrator, Cubic Transport Systems (Australia) Pty Limited (**Cubic**), who we understand to be a sub-contractor of TfNSW. In terms of undertaking our PIA, we have considered Cubic and TfNSW to be the one entity, and interchangeable as the collector and holder of Opal Data on the basis that the agreement between TfNSW and Cubic are such that:

(a) TfNSW does not release the handling of Personal Information from its effective control by having a binding contract with Cubic that requires it to handle the Personal Information only for the purposes for which it is provided; and

(b) TfNSW has ability to effectively control how the information is handled by Cubic.

In considering the role of Cubic in relation to the Open Data Project it is important to note that the Opal Data held by Cubic and the data released by Cubic to TfNSW as part of the Opal Journey Segment View is only a subset of the full suite of Opal Card information and the Opal Journey Segment View does not contain of itself any elements that would be identifying information (i.e. personal information) on their own.

Accordingly, the key issue we are considering is re-identification. It is important to note that the release of the Opal Journey Segment View dataset, or a sub-element of it, is an element that has been extracted from, and is not linked in any way, to the Opal database held by Cubic in its systems and is maintained by TfNSW in its systems separate and distinct from the Opal systems.

Accordingly, the issues of disclosure of Personal Information by Cubic do not need to be considered and the issue of contractual entitlements to data have not been considered on the basis that the elements that have been released by Cubic to TfNSW as part of the Opal Journey Segment View are not Personal Information, are not traceable to the Cubic system and represent solely a collection of data elements that are designed to assist TfNSW in its planning purposes. The information at this point is not Personal Information. The issue of re-identification risk means that the standard analysis of disclosure of Personal Information does not apply in this instance. Instead, a nuanced approach premised on mitigating the risks of re-identification is called for.

2. Privacy Requirements – Legislation, Policies and Stakeholder considerations

2.1 Privacy legislation

TfNSW must comply with privacy laws that apply to NSW public sector agencies. The *Privacy and Personal Information Act 1998* (**PPIP Act**) and *Health Records and Information Privacy Act 2002* (**HRIP Act**) regulate the way agencies handle personal and health information throughout the information 'life cycle', from collection through to storage, use and disclosure.

"Personal Information" is defined in the PPIP Act as:

> *"Information or an opinion (including information or an opinion forming part of a database and whether or not recorded in a material form) about an individual whose identity is apparent or can reasonably be ascertained from the information."*

Personal Information must be protected by agencies, including TfNSW, in accordance with the twelve Information Protection Principles (**IPPs**) in the PIPP Act.

Health information is defined in the HRIP Act and includes information about a person's physical or mental health or disability. Health information must also be protected by agencies, including TfNSW, in accordance with the fifteen Health Privacy Principles (**HPPs**) in the HRIP Act.

Together these IPPs and HPPs require TfNSW to ensure that:

(a) collection of Personal Information is open, reasonable, necessary, direct and lawful;

(b) Personal Information is stored in a secured manner;

(c) Personal Information is only used for the purpose for which it was collected; and

(d) any disclosure of Personal Information is consented to by the relevant individual.

TfNSW also has obligations to handle information in accordance with:

(a) *State Records Act 1998* requirements, including to ensure the safe custody and preservation of State records within TfNSW's control and not to dispose or destroy State records without proper approval processes;

(b) *Government Information (Public Access) Act 2009* requirements to provide access to information where there is no overriding public interest against disclosure; and

(c) Contractual arrangements with third parties that provide for the protection of confidential information.

2.2 Privacy guidelines

Internal privacy policies also apply to the way Personal Information is handled by TfNSW, including TfNSW's Privacy Management Plan version 1 issued in May 2014, and processes for managing privacy complaints and requests by individuals for access to their information.

The Opal Privacy Policy and Opal terms of use apply to the way TfNSW handles Personal Information when a customer applies for and uses an Opal Card.

As stated above at sections 1.2(a) and 1.5, the Opal Data proposed to be released under the ODP does not of itself contain any elements of Personal Information. Accordingly, the issue of IPP and HPP compliance arises in the context of mitigating the risks of potential re-identification. In our view, the existing Privacy policies for both TfNSW and Opal do not contemplate this type of

situation and will need to be amended to reflect the fact that information which may not, when held by the entity, be Personal Information, may by virtue of its release into a public environment by TfNSW be able to be re-identified. This would require the Privacy Policies to outline the steps that they have taken to minimise this risk.

2.3 Characterisation of the Opal Datasets – Do they contain Personal Information?

It is necessary to consider whether the datasets proposed to be released by TfNSW under the ODP constitute Personal Information or if they do not constitute Personal Information is there a risk that the information released could be re-identified so as to become Personal Information held by a third party?

The datasets proposed to be released are a subset of the dataset known as the Opal Journey Segment View. There are 117 attributes recorded in the data dictionary of the Opal Journey Segment View and none of them on their own constitute Personal Information as that term is defined in the PPIP Act. However, the concern is whether the information contained in those 117 attributes is sufficient in an open data context when combined with other information that is proprietary or publicly available, could be re-identified.

Accordingly, the issue is not the prevention of the release of Personal Information in the dataset as it currently exists, but the consideration as to whether the elements and attributes of the dataset will enable re-identification of an individual arising from the ODP release of the dataset.

This represents a fundamental shift in the types of protections and analysis that have been made of datasets in the past. A recent publication in the Berkeley Technology Law Journal in relation to open data states as follows:

> *"Emerging challenges in this area include the rising speed of data collection and processing (sometimes referred to as data velocity), heightened data integration, and increasing analytics sophistication. Capabilities for linking statistical data to auxiliary data sources are improving, and common techniques for limiting disclosures risks can greatly diminish the utility of the data. Agencies are pressured to release data faster, more cost effectively, and in a way that allows a greater range of analysis, including visualisations and data mining, and provides estimates for finer time scales and geographic areas."*

The conclusion arising from this is "thus e-Government and open data programs represent a fundamental shift in the way governments release data".

Addressing and responding to this fundamental shift is the concern of this PIA and the analysis of the attributes forming part of the Opal Journey Segment View. It also means this PIA considers the application of technology and learnings on technical ability to re-identify individuals from information as once data is released TfNSW cannot control its use. This is fundamentally different from a PIA that considers the collect, hold, use data lifecycle.

In this respect, regard may be had to overseas experience where we can look at approaches that have been taken in different jurisdictions to limit the amount of data that is released from a particular dataset and also to consider instances of data being released under an open data policy to find that the actual uses exceeded the intended uses and that re-identification was able to be undertaken despite the releasing agency's estimation that it would not.

Perhaps the most famous example is the New York City taxi information release where drivers and taxi details were able to be re-identified. The New York City Taxi and Limousine Commission then redacted a number of data fields for all subsequent data releases to prevent re-identification.

These are examples to be borne in mind in determining whether all of the 117 attributes or only a portion of them will be released.

2.4 Protecting the Datasets – De-identification

While as at the date of this report the NSW Privacy Commissioner has not provided any guidelines for anonymization of datasets and their use in research, two key documents which provide guidance and assistance in this area are the:

(a) OAIC *Privacy Business Resource 4: De-identification of data and information*; and

(b) UK Information Commissioner's Office document *Anonymisation: Managing data protection risk code of practice,* issued in 2012.

Both of these documents set out various categories of risk and make recommendations for dealing with those risks. It is noted at the outset that de-identifying data is never entirely risk free and there is always a risk that an individual's Personal Information could be re-identified to be held and used as Personal Information by a third party as a consequence of the data release.

Accordingly, the burden of the de-identification strategy for TfNSW under ODP is to maximise the single opportunity to mitigate the risks of all future re-identification.

2.5 Disclosing the Datasets

We have assumed that in collecting and holding Personal Information about individuals within the Opal Data and ETS database that collection and authorisation requirements of *PPIP Act* have been complied with. We further assume that the limitations on the number of elements or attributes is based on reliable assumptions as to de-identification of data, explored further below and considered in the service specification for testing of the proposed datasets prior to release.

2.6 Re-identification – issues and considerations

The OAIC Privacy Business Resource 4 provides useful guidelines in relation to de-identifying data. It helpfully defines the process as including two steps, being:

(a) removing personal identifiers; and

(b) removing or altering other information that may allow an individual to be identified, for example because of a rare characteristic of the individual or a combination of unique or remarkable characteristics that enable identification.

In this regard, it is noted that the risk of re-identification must be actively assessed and managed to mitigate the risk. In the case of the Opal Datasets, the existence of location data in relation to trips which may occur on a regular basis, create a potential risk of re-identification. However, the balance between removing identifiers and re-identification is whether further de-identification steps reduce the value and utility of the data set to achieve its objectives.

The OAIC Guideline states that one of the tests to be applied in considering the risk of re-identification is applying the "motivated intruder" test. This test is generally taken to be where a reasonably competent motivated person with no specialist skills would be able to identify the data or information. It assumes that the motivated intruder would have access to resources such as the internet and publicly available documents and data sources, which would enable them to

make further enquiries. One likely application of this to the Opal Datasets would be a potential stalker seeking to identify travel patterns and plans of an individual from the data.

This test is considered in detail in section 3 of the ICO Guideline, where it raises the question of jigsaw attack; that is, piecing together different bits of information to create a more complete picture of someone.

A jigsaw attack may be used to piece together various pieces of data and information to re-identify and create personally identifying information. If it is Personal Information, then sections 17 and 18 of the *PPIP Act* prohibit the use and disclosure other than for the specified purposes in the privacy notice.

Accordingly, there is a need to treat the datasets before release to reduce the risk.

There is a possibility that information may not be Personal Information even if it relates to the individual. This was considered in detail in the ongoing case of Ben Grubb and Telstra regarding metadata held by Telstra in relation to Mr Grubb's use of the Telstra voice text and internet services. The relevance of the Ben Grubb metadata case to the scenario of Opal Data is that it provides a good analogy as to where the distinction between personally identifying information and information that may be about an individual, but the individual is not the subject of the information, is crucial to making the distinction as to whether information is Personal Information or not.

At first instance, the Commonwealth Privacy Commissioner ruled that the metadata was in fact Personal Information and that Telstra was required to give that information to Mr Grubb (*Ben Grubb v Telstra Corporation Limited* [2015] AICmr 35). The Commissioner's decision canvassed at length the various categories of information held by Telstra in its various management information systems, how those systems co-ordinated and communicated with one another, and what data Telstra was able to extract that was relevant to Mr Grubb. As a background note, it is worthwhile to understand that one of the basic contentions made by Mr Grubb was that information was routinely provided by Telstra to law enforcement authorities on request and if law enforcement authorities could have Personal Information about Mr Grubb then his query was, why could he not have the same information?

The Privacy Commissioner undertook a detailed analysis and came to the view that information about Mr Grubb's use of the system, including various information in relation to calls, which went well beyond the information routinely provided through the billing system, was Personal Information and was to be released to him, except in circumstances where to do so would be to infringe the Personal Information of a third party.

Telstra appealed and the matter was heard by the Administrative Appeals Tribunal by Deputy President S.A. Forgie (*Telstra Corporation Limited v Privacy Commissioner* [2015] AATA 991). That decision overturned the decision of the Privacy Commissioner on the basis that a number of specific instances of metadata recorded information that while they concerned or related to the individual, they were not information about the individual. Deputy President Forgie gave, by way of example, the analysis of a traffic accident at paragraph 99:

> *"Putting the issue another way, how tenuous can the link be before information or opinion is not about an individual but about something else or, if still about an individual, not about a particular individual but another? If I were to imagine a road accident in which a car ran a red light and hit a pedestrian who was walking with a green light, the report of the accident itself naming the driver, the pedestrian and the circumstances of the accident could, as a whole, be said to be about the driver, the pedestrian, the circumstances of the*

accident, the witnesses, the state of the road surface and the weather and so on."

On this basis, while certain information will be about or concerning an individual, it will not necessarily be Personal Information. This is relevant as it might suggest that various data elements held in the Opal Data datasets, while about an individual, will not be Personal Information. The Commissioner appealed the AAT decision and so while this case cannot be taken as to reflect the settled state of the law at this time, it is indicative of the fact that some information may not be Personal Information as such. The appeal is due to be heard in August this year so more clarity on this point may be available after that date.

In any event, this class of information may then fall into the high-risk category for re-identification and anonymization as the fact that it concerns an individual may make it easier to re-identify in the event that a motivated intruder has access to other databases.

In our view, it is this risk which would cause us to treat information of this type in the same way as Personal Information from the perspective of *PPIP Act* protection and consequent use and release, and require the data be treated appropriately.

2.7 Data treatment

The logical consequence of the above analysis is that a robust approach to treating the Opal Data to ensure it is unlikely to be able to be re-identified before it is released is required. That is, industry best practice techniques for anonymisation and de-identification be applied prior to release. This is considered further in **Section 4**.

3. Stakeholder Consultation & Expectation Management

3.1 Stakeholder consultation is outside the scope of this PIA.

Accordingly, we have not advised of any risks associated with the stakeholder management expectations.

That said, we raise the issue in the context of community expectation, proposals for new laws to give rise to remedies for breach of privacy and the context of ongoing public and media scrutiny of the Opal card. One of these issues that has been raised in the media on an ongoing and relentless basis, is the ability of the Opal card to track an individual. This is a technical ability and the information is, of course, kept secure within TfNSW and Cubic.

However, to the extent that information is released to the public at large as part of the ODP, the issue may arise that there is scepticism and concern about the security arrangements in place. In the context of ODP this becomes a significant issue. In the recent ALRC report on Serious Invasions of Privacy in the Digital Era (Report 123, 2014), the report dedicated a chapter to considerations as to what is "a reasonable expectation of privacy".

While not coming to any clear conclusion, it was noted that different factors will vary on a case-by-case basis and will depend on community standards and expectations. It will also depend on the means used. The report stated that while individuals may not have an expectation of privacy from the use of a technology of which they are unaware, it was noted that many people *"may even reasonably expect privacy without even considering whether their privacy is remotely likely to be intruded upon at all"* (paragraph 6.42). Accordingly, to the extent that people's travel, whether it be for home or for work, is regarded as a private matter will depend on all of the circumstances.

Based on the Opal Privacy Policy, it is likely that an individual Opal card holder would consider that the information was held only by TfNSW and/or Cubic and would not be shared more widely. While this does not provide any clear answer, it raises the point that while media has mentioned at a number of points in time the likelihood of being subject to unwelcome surveillance by a stalker using location information, it is reasonable to expect that individuals would rely on TfNSW, the Opal privacy notices and the Opal privacy statements to suggest that this could not occur.

If Opal Datasets were shared more widely to allow this to occur, it is reasonable to expect that there would be some form of stakeholder backlash and potentially adverse media coverage. Accordingly, considerations as to whether some or all of the Opal datasets are suitable for ODP is a matter requiring further detailed consideration and is explored further in **Section 4** of this Report.

As a minimum, amendments to TfNSW and Opal privacy policies as set out at Section 2.2 would be required.

3.2 Open Opal Data Industry Reference Group

Notwithstanding the comments above, we note that subsequent to the initial testing of the MVPE TFNSW engaged the above referenced group to provide feedback throughout the project and guidance about potential use cases and particular issues arising in relation to the use.

It is apparent from reviewing the Data 61 report that there has been a level of informed discussion about the trade-off between potential risk of re-identification of individuals as a result of Opal Data and the utility of the data sets released. Given these are matters for the expertise of data scientists we make no comment other than to state that the fact that such engagement has occurred is, in our view, another step in mitigating any likely risks in release of the data.

4. Open data policy

4.1 Overview of NSW ODP

In September 2013, the government released Version 1 of its Open Data Policy. That Policy was reviewed and a simplified, updated policy was released in April 2016. However, the key objectives remain the same. They are to:

(a) simplify and facilitate the release of appropriate data by NSW government agencies;

(b) make explicit the NSW government's commitment to open data and open government;

(c) create a practical policy framework that enables high-value datasets to be released to the public;

(d) help agencies in understanding community and industry priorities for open data; and

(e) support the *Government Information (Public Access) Act 2009* (NSW) and promote simple and efficient compliance with the requirements set out in that Act.

The Open Data Policy is only one part of the NSW government ICT strategy and a number of steps have been taken to implement the Open Data principles and to ensure that there is transparency in government, where appropriate. However, the Open Data Policy also recognises at paragraph 6.2 that data should not be released, or not be released in full, when considerations including privacy and security preclude its release.

We note that the New South Wales Government Open Data Policy is a dynamic and changing one and accordingly we acknowledge that the key principles and objectives will remain the same but responses may vary over time and the way in which this PIA and Opal Data integrate into that regime may vary.

4.2 Specific Opal Data Risks

In the context of Opal Data and in particular, individual Opal Card user data, the potential risks to the privacy of individuals who can be re-identified by reference to their card and or journey activity, if indeed their habitual transport preferences can be monitored and predicted, can be significant. This means that the application of the de-identification processes to the Opal Datasets to prevent re-identification by a motivated intruder are a key element of any process of release of the Opal Data as part of the Open Data framework.

Ideally, TfNSW would like guidance around the release of Opal Data that classifies it into three headings – data that can be released; data that can never be released; and then a clarification of those items which fall into what might be called the "grey area".

However, for the reasons set out above in relation to de-identification, it is difficult to create such a clear set of guidelines. If we consider the Opal Datasets, then it is clear that identifying information which constitutes Personal Information can never be released. For registered card holders, this will include name, address, date of birth and card identification number. None of these are contemplated in the Opal Journey Segment View dataset. Information that then falls into the "grey area" would be hashed and otherwise secured card identification number (or CINs).

Whether additional information will also fall in the "grey area" depends on the ability to de-identify in the context of Open Data background. The ICO guideline states on page 21 that in such a framework it is necessary to consider whether a current release of data exposes it to re-identification in the future. In particular, it says as follows:

"A realistic assessment of the risk of re-identification occurring in the future should be made, meaning that organisations should not assume that data that is anonymous now will necessarily become re-identifiable in the future. However, organisations should carry out a periodic review of their policy on the release of data and of the techniques used to anonymise it, based on current and foreseeable future threats. There are certainly examples though of where a complacent approach to anonymization, and insufficiently rigorous risk analysis, had led to the substantial disclosure of personal data. This was a case where "anonymise" internet search results were released without proper consideration of the risk of individuals identifying each other from the search terms used."

Accordingly, in terms of using unique identifiers for cards and providing traceable individual card journey information, it is questionable as to whether this can be released or whether it could be subject to re-identification.

While there may be various databases which would allow re-identification, for example proprietary databases such as RMS DRIVES, these are not generally available to a motivated intruder and it is necessary to consider whether in the context of a motivated intruder, with access to a range of publicly available datasets and sources of information, information from Opal trip data as contained in the Opal Journey Segment View could be re-identified.

A conservative view would consider that allowing historic trip data to be analysed on an individual card basis would potentially provide enough points of reference for that card holder to be identified.

If trip data is available in some sort of aggregated form, then the likelihood of re-identification is reduced.

The ODP document considers that one of the benefits of Open Data Policy is that the provision of high-value datasets will allow new and possibly not contemplated uses, which will enhance both Government use of the data and public use of the data for the public benefit.

One way to mitigate is to only allow use in a closed and contractually regulated context, that is, contractual protections would need to be placed around the use of the data by third parties. However in the context of Open Data, it may be that a middle-ground is to provide a free licence to use certain data from the Opal Data datasets on contractual terms that deal with the potential risks to privacy.

The third option is to release a limited set of attributes that minimise privacy risks.

In many instances that we can envision, where transport data is used to predict demand and provide better services, the likely beneficiary of any such analysis would be TfNSW and/or the private operators concerned in the provision of the relevant services.

Where Opal Data datasets might be used for broader public research and analysis, then in our view, where individual card tracked trip data is released, it can only be released after it has been subject to rigorous testing for re-identification risk.

The mapping of such forms of release is a project that would need to be undertaken in conjunction with ICT security experts.

In our view, the privacy risks to Opal card users of re-identification at this point cannot be adequately assessed and on that basis, at this point, no information should be released as part of the ODP.

4.3 Next Steps – Recommendation

In order to move to a point where Opal Journey Segment View data can be released under ODP we have prepared a service specification for a suitably qualified service provider to rigorously test the dataset for risks of re-identification using a range of industry acknowledged best practice techniques and to recommend the appropriate treatment and or range of treatments to minimise risks.

Subject to TfNSW making an assessment of the risks after this formal testing process, it is our view that the privacy risks we have identified can be addressed by selecting and applying the methodologies recommended by the technical experts.

A copy of the draft service specification is attached to this PIA at **Annexure B**.

4.4 Results of Initial Data Testing

As set out in the Executive Summary, an initial testing has been undertaken of a limited data set MVPE and, based on the results of that initial testing, the data sets that were subject to the testing and the application of the differential privacy algorithm may be released for the period set out in the Initial Summary Report as part of the ODP.

This is on the basis of the opinions of the data scientists expressed in that Initial Summary Report attached as Annexure F. Before any further elements of the Opal Journey Segment View Data and in fact any single journey that combines elements of bus, train, ferry and light rail can be released as a single journey, further testing is required. Similarly, if a further set of MVPE is released for a different time period, it should be re-assessed for risk before release.

4.5 Results of Data 61 Privacy Preserving Opal Data Report

The Data 61 report dated 15 November 2016 and the addendum dated 17 November provide a final report in relation to the Open Opal Data as provided to Data 61 for the purpose of testing and concludes that the mitigation of privacy risks as set out in Section 6 of the report are sufficient to allow TFNSW to proceed to release the Open Opal Data Set as tested subject to employing the strategies that have been recommended by Data 61, including in particular differential privacy.

Section 8 of the Data 61 report sets out some of the concerns with privacy preservation that are specific to the nature of the Opal Data Set and consider the trade-off between the number of data elements and the utility. Section 8.2 of the report contains an important caveat being “the focus on getting data to release without assessment was not as envisaged in the scope of works” and notes that with more time allowed to do further testing then work to improve accuracy and errors may be improved. However, it is noted that this will not occur. The addendum addresses specifically the mitigation of privacy risks at section 4 and also makes comments in relation to functionality.

In particular, the application of the algorithm to the released data sets provides for differential privacy. We note the disclaimers applied by Data 61 if the data is varied. Provided the data is released in accordance with the Data 61 algorithm and the guidelines set out to ensure privacy including the exclusion of small groups below a threshold of 18. Based on the skill of the data scientists it is apparent that all relevant privacy protections have been considered and put in place for the benefit of Opal Card Users in relation to the data set.

5. Privacy Impact Analysis

This PIA assesses the Project's privacy impacts against the 12 IPPs in the PPIP Act and the 15 HPPs in the HRIP Act.

The IPPs and HPPs are legally binding obligations that apply to all public sector agencies in New South Wales when handling Personal Information. The IPPs and HPPs are summarised in plain language in italics in this section.[1]

As mentioned above, the scope of this PIA Report is fundamentally different from the traditional data life cycle analysis as the risk sought to be mitigated is the potential re-identification of data which, at the time of release, is unlikely to be Personal Information as such.

However, in order to demonstrate all of the issues have been considered, we set out below our approach to each IPP and HPP under this factual scenario.

5.1 Collection of Personal Information

The privacy requirements relevant to the collection of Personal Information are IPPs 1 to 4 and HPPs 1 to 4.

(a) Lawful collection (IPP 1 & HPP 1)

> *Only collect Personal Information for a lawful purpose, which is directly related to the agency's function or activities, and is reasonably necessary for that purpose.*

The Opal Privacy Policy specifies the purposes, this is BAU.

(b) Direct collection (IPP 2 & HPP 3)

> *Only collect Personal Information directly from the person concerned, unless they have authorised collection from someone else, or if the person is under the age of 16 and the information has been provided by a parent or guardian.*

Any Opal Data that is Personal Information is collected directly.

(c) Open collection (IPP 3 & HPP 4)

> *If collecting from an individual, inform the person you are collecting information from why you are collecting it, what you will do with it and who else might see it. Tell the person how they can view and correct their Personal Information, if the information is required by law or is voluntary, and any consequences that may apply if they decide not to provide their information.*

The Opal Privacy Policy specifies data analytics and research as a possible use and flags third party contractors as being used. It does not currently specify release of data as broadly as contemplated by the ODP hence revision is recommended. Similarly, the TfNSW Privacy Policy contemplates all relevant issues being covered off in the Opal Policy.

[1] The summaries of the IPPs are adapted from the Information and Privacy Commissioner's *Fact Sheet – information protection principles for agencies*.

Recommendation 1

As the release is by TfNSW, updating the TfNSW Privacy Policy to specifically address ODP is recommended.

Privacy Policy – an additional sub-paragraph should be added to paragraph 3.2 of the Opal Privacy Policy as follows:

> "Opal card use data as set out in clause 4.2.3 may be released to the public in accordance with and to meet the objectives of the NSW Government Open Data Policy https://www.finance.nsw.gov.au/ict/resources/nsw-government-open-data-policy.
>
> In addition, information which has been de-identified may be made available."

(d) **Relevant information (IPP 4 & HPP 2)**

Ensure that the Personal Information is relevant, accurate, complete, up-to-date and not excessive and that the collection does not unreasonably intrude into the personal affairs of the individual.

Information collected is necessary. No change to BAU

5.2 Storage

(a) **Secure (IPP 5 & HPP 5)**

Keep the information for no longer than necessary, dispose of it appropriately, and protect it from unauthorised access, use, modification or disclosure.

Information is kept secure in accordance with NSW Government policies. The treatment before data is released under ODP needs to ensure all reasonable steps are taken to de-identify. See our recommendation at **Section 4.3**.

5.3 Access and accuracy

The privacy requirements relevant to the collection of Personal Information are IPPs 6 to 8, and HPPs 6 to 8.

(a) **Transparent (IPP 6 & HPP 6)**

Explain to people what Personal Information is held about them, what it is used for, and any right to access the information.

No change to BAU.

(b) **Accessible (IPP 7 & HPP 7)**

Allow people to access their Personal Information without excessive delay or expense.

No change to BAU.

(c) **Correct (IPP 8 & HPP 8)**

Allow people to update, correct or amend their Personal Information where necessary.

No change to BAU.

5.4 Use of Personal Information

(a) **Accurate (IPP 9 & HPP 9)**

Make sure that Personal Information is relevant, accurate, up to date and complete before using it.

No change to BAU.

(b) **Limited (IPP 10 & HPP 10)**

Only use information for the purpose collected.

Recommendation 2

Any release of Opal Data under ODP initiatives be deferred until privacy concerns can be specifically addressed, having regard to each data field and the issues of potential re-identification.

5.5 Disclosure of Personal Information

(a) **Restricted (IPP 11 & HPP 11)**

Only disclose with a person's consent or if the person was notified about the disclosure at the time of collection.

Satisfied for the Projects.

(c) Safeguarded (IPP 12)

Only disclose sensitive information with the individuals consent or in order to deal with a serious or imminent threat to any other person's health or safety.

To the extent this occurs for certain Opal Concession card holders, it is covered in their Privacy Notice.

5.6 Other applicable protection principles

(a) **Identifiers (HPP 12)**

Identifiers can only be applied to personal health information if this is reasonably necessary to carry out the organisation's functions. Public health system identifiers may be used by private sector agencies, but only in defined circumstances and with strict controls.

Not applicable.

Annexure A – Documents provided by TfNSW & NSW Documents

	Description
1.	Briefing Note TD/04892 Releasing Opal Data
2.	Opal Journey Segment View – Data Dictionary June 206 Version 0.6
3.	NSW Government Open Data Policy April 2016
4.	Information Management Framework from finance.Nsw.gov.au/ict
5.	Open Data Action Plan 2016
6.	Open Opal Data Set Documentation December 2016 version 1.0
7.	Open Opal Data Project Management Plan December 2016 version 1.0

Annexure B – Service Specification for Confidential Service Provider Testing

1. Background – Open Data Policy & Opal Data

1.1 We have been asked to prepare a PIA which considers the risks in releasing certain Opal data (the Opal Data Set) under the NSW Government's Open Data policy (**Annexure D**). In accordance with both the *Privacy and Personal Information Protection Act 1988* (NSW) (the **PPIP Act**) and the *Health Records and Information Privacy Act 2002* (NSW) (the **HRIP Act**), personal information cannot be released. However, the issue in relation to the Opal Data Set relates to the steps necessary for anonymisation and de-identification in relation to the Opal Data Set and the potential risks of re-identification. We note that while the Opal Data Set does not contain any personal information as such, it does include a range of identifiers and elements that could be used to re-identify an individual or individuals. As the tests for determining this are primarily technically based, advising on the relevant steps to be taken to anonymise and de-identify is outside the scope of our legal expertise.

1.2 Attached to this document at **Annexure C** is the Opal Journey Segment View Data Dictionary, as already redacted by the business unit to remove elements that are considered as not of any utility to persons outside TfNSW. This document describes the data attributes that are available to be viewed by TfNSW on a daily basis (but not in real time) that may be available for release. It is noted that the focus for TfNSW in viewing the data attributes is to analyse journey segments for a range of purposes.

1.3 However, a number of the data attributes, although relating to journeys, may allow an individual to be identified. The purpose of this brief is to request that, having regard to the data attributes as defined in the redacted Opal Data Set, the Service Provider undertakes testing of a sample data set or sets with the attributes in the redacted Opal Data Set to provide certain recommendations as to the quantifiable level of risk in releasing the redacted Opal Data Set without removal of further data attributes. Alternatively, recommendations as to the key attributes that give rise to risk and should be removed.

1.4 Also attached at **Annexure E** is an extract from the internal TfNSW proposal for release of the Opal Data Set which identifies a number of possible techniques for de-identification and the benefits/limitations. It is requested that the service provider consider these benefits/limitations as identified by the business unit in the context of the testing and recommendations sought in this document.

1.5 Fundamentally, the question can be summarised as follows: even after removing the hashed CIN, what is the likelihood of identifying tracking a unique journey or journeys and then identifying an individual taking that journey?

1.6 Such testing should be undertaken in accordance with the Commonwealth Government Office of the Australian Information Commissioner's Privacy Business Resource 4: "De-identification of data and information" issued in April 2014 and certain elements of the UK Information Commissioner's Office Guide: "Anonymisation: managing data protection risk code of practice" issued in 2012, as set out below. It is noted that currently there is no specific guidance available from the NSW Privacy Commissioner.

1.7 The Australian Business Resource makes reference to a number of guidelines issued by the National Statistical Service confidentiality information series in relation to confidentialisation, managing the risk of disclosure in the release of data and how to deal with risks of re-identification. **Schedule 1** to this Scoping Request sets out the general guidelines referenced in the above publications relevant to this project.

2. Request – General issues

2.1 We request that the redacted Opal Data Set be tested in the context of the NSS guidelines and those set out at **Schedule 1**.

2.2 The testing to be undertaken includes testing alternative configurations of the data attributes, including whether there are certain key data elements that require suppression or removal in order to minimise risk, as indicated at 1.3 and 1.5 above. If the Service Provider recommends any treatment to the data whether the treatments set out in this specification or otherwise, please advise which treatments and how they might be applied.

2.3 There are 117 data attributes listed at **Annexure C** and some of those are clearly required to be removed or otherwise de-identified. Some of the attributes relate to vehicles, such as a bus or a ferry, and some relate to a fixed location, such as a transit stop or a train station. Some of the attributes refer to the name of a vehicle operator and some relate to the relevant individual, by virtue of being related to the Opal Card.

2.4 To the extent there are attributes that relate to an Opal Card, they are necessary to follow a journey that might be intermodal to benefit from understanding the different types of transport used.

2.5 However, identifying a single passenger whether through a hashed card identification number or similar, will potentially allow identification of the individual. This is the key data attribute for consideration and testing.

2.6 The other key types of attributes are passenger type code, card type code, and discount entitlement code and discount entitlement description which might lead to the ability to identify the individual cardholder by virtue of their attributes such as adult or child and any other discount entitlement.

2.7 Tracking a journey then through the various steps for “tapping on” and “tapping off” different vehicles and different routes could, as individuals are creatures of habit, again allow an individual to be located. This applies where there are multiple days of journey data released and a potential journey could be tracked through use of the weekly cap applied and whether discounts have been applied to continued use. In accordance with the Open Data Policy we assume that there will be ongoing releases over time. From the 117 attribute fields there are some general requirements and some specific requirements. One of the key issues with the journey segments is that location data can be used as a quasi-identifier and to the extent that it allows for re-identification, removal of those data attributes requires consideration. [Note, TfNSW to provide some guidance lists here as part of further details and briefing]. NOTE: This specification with this original comment was provided to Data 61 in the initial testing phase and while the specification has not been updated in a course of the PIA it is apparent from all report that there has been significant and robust discussion between the data scientists at the Open Opal Data Industry Reference Group and TFNSW, Entity of NSW and Data 61. On this basis we are satisfied that the mitigations proposed for the binning of times and the aggregation of “tap-on tap-off” locations is sufficient to mitigate the risks identified of tracking a single journey

3. Specific testing request

3.1 We ask that the testing consider the impact that the removal of each attribute has in terms of degrading the ability of the data set to be used for planning and other purposes. Accordingly, based on the guidelines referenced above and the specific general issues referred to below, our request is that confidential testing be undertaken with a view to providing:

(a) a minimum viable product of data fields that would be “low risk” in terms of re-identification; and

(b) an enhanced data set that would meet the requirements of addressing the testing issues set out below but would have some risk attaching but be acknowledged to be a more useful data set.

3.2 To the extent any recommendations are to be made in relation to further technical mitigation of the risk of re-identification in relation to the release of the Opal Data Set then this information should be provided as part of the deliverable to this Specification.

4. Next Steps

4.1 Further technical assistance and provision of sample Opal Data Set will be provided by TfNSW when a response with proposed testing methodologies and costing are agreed.

13 July 2016

Schedule 1 - Principles

General testing principles to be considered in testing design

1. Information concerning the likelihood of re-identification of individuals being attempted by members of the public and/or organisations.

    (a) How and why could non-personal information reasonably be linked to an individual?

2. The likelihood that such re-identification attempts would be successful.

    (a) Particular consideration to be given to two main methods of re-identification:

        (i) An 'intruder' uses its own personal knowledge and searches an anonymised dataset for a match; and

        (ii) An 'intruder' takes a record from an anonymised dataset, and seeks a match in other publicly available information.

    (b) In some situations data will necessarily be unique by reference to the hashed CIN identification.

    (c) Consideration of the other information that is available. This is inherently uncertain as we can never be certain what data is already available/what may be released in the future. Return/recall/removal of data is not a viable option once published/released to the public therefore it is essential to consider the inherent uncertainty (different members of the public will have different degrees of access to 'other information' needed to identify. Non-recorded personal information may pose significant risk but is hard to establish.

    (d) 'Motivated intruder' test: 'jigsaw attack' – person who starts without any prior knowledge but who wishes to identify the individual from whose personal data the anonymised data has been derived – would they be successful? Can this be undertaken? Assume the intruder is reasonably competent, has access to resources (internet, libraries etc), and would employ investigative techniques (e.g. making enquiries). Consider:

        (i) the characteristics of the data that facilitate data linkage from the data attributes;

        (ii) other 'linkable' information that is available;

        (iii) technical measures that might be commonly available and used to re-identify;

        (iv) the weight to be given to personal knowledge;

        (v) the results of a penetration test,

    (e) Consider identification 'in the round' - whether any member of the public/organisation could identify any individual from the data – on its own or in combination with other data → risk will vary according to local data environment.

    (f) Unique/uncommon characteristics – quasi-identifiers. This includes the hashed CIN and the combination of geolocation and time attributes.

3. The anonymisation techniques which are available to use are set out below .TfNSW seeks recommendation as to preferred or more effective methodologies.

    (a) Consider:

(i) Data masking (including partial or quarantining)

(ii) Pseudonymisation

(iii) Aggregation (low risk but can't identify individual)

(A) Cell suppression

(B) Inference control

(C) Perturbation

(D) Rounding

(E) Sampling

(F) Synthetic data

(G) Tabular reporting

(iv) Derive data items and banding (low risk and allows personal identification)

(v) Remove quasi-identifiers

(vi) Alter 'tolerable error'

(vii) Swapping

(viii) Synthetic data

4. The quality of the data after anonymisation has taken place, and whether this will meet the needs of the organisation using the anonymised information. For example, would the service provider recommend randomisation on a regular basis, eg weekly?

5. The risks to/potential effects on individuals.

(a) when an intruder has personal knowledge – the risk needs to be addressed whether they might learn something sensitive. The risk could be low when in order for re-identification to occur, the individual must already know a lot of information – will they learn anything new?

(i) Consider conducting a general assessment for some individuals and then making a global decision about the information and the likelihood that those people that could re-identify actually seeking out or coming across relevant data.

(ii) Information about groups of people – not personal data but still potential privacy risk.

Specifically on spatial data attributes

6. Location data as a quasi-identifier – there are a number of these in the Opal Data Set.

7. Spatial data will not necessarily be personal information, but the more precise it is the more possible it becomes likely to be personal data upon analysing it or combining with other information. Balance whether being published for a legitimate purpose with the protection of an individual's privacy. Can the data be useful if quasi-identifiers removed for privacy?

8. Distinction between 'statistical comfort zone' – eliminating all risk of identification, and other forms of information that pose a risk of an individual being identified (e.g. small numbers in small

geographical areas – increased risk, but not all small numbers should be automatically removed). Looking to consider minimum viable product data set.

Consider how published

9. Under the Open Data Policy it is intended that as much data as possible be made available, subject to PPIP Act and HRIP Act compliance. In this instance, the Opal Data Set is the extent of data under consideration.

10. It is noted in relation to the open disclosure of transport journey data in other jurisdictions it has been limited and there have been live issues of re-identification. One issue may be to provide a sample of travel on a particular day and not to release series of days so no pattern can be identified that may be linked to an individual. Any recommendations as to these issues would be kindly received.

11. Consider as an alternative Sampling – only making some parts of the database available.

Problems & risks for consideration:

12. Subsequent uses to which the collected and released data might be put are essentially unknown and uncontrolled.

13. Removing/encrypting/etc the unique card identifier would minimise risk of being able to track individuals but would greatly reduce the analytical potential of the data. Retaining the ability to track a journey but not an individual is a key issue.

14. Data that is available now and in the future – unpredictable – need risk assessment & periodic review.

15. Techniques available are set out below but query their relevance to this data set:

    (a) could 'degrade' or 'fade' the information – start off with details but before release to public could be changed to suburb;

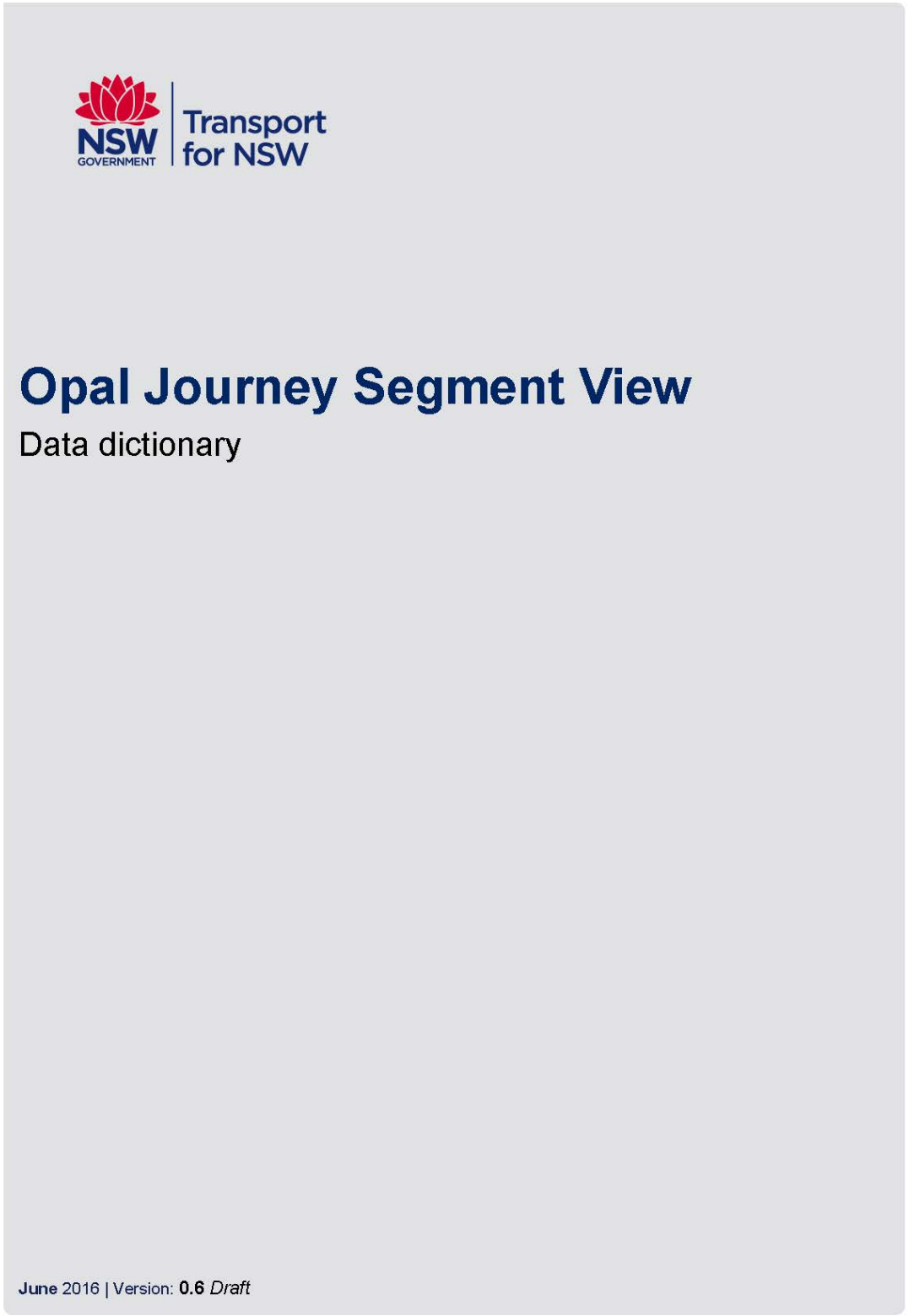
    (b) increase mapping area;

    (c) reduce frequency/timeliness;

    (d) formats – overview (e.g. heatmap);

    (e) avoid household level – could be personal data.

Annexure C – Opal Journey Segment View Data Dictionary



Contents

List of tables

Author:	Agius, Catherine – **Error! Unknown document property name.**
Date:	8 June 2016**Error! Unknown document property name.**
Version:	0.6
Reference:	TBA
Division:	Freight, Strategy and Planning
Review date:	N/A

1 Introduction and overview

1.1 Document purpose

This document describes the data attributes available through the Opal Journey Segment View. The Journey Segment View brings together data from several tables within the Oracle datamart housing Opal data. The purpose of the view is to provide a single set of meaningful journey segment data for each journey segment, simplifying the processes of working with and analysing journey segment data.

1.2 Document scope

This document focuses purely on the definitions of data attributes included in the Journey Segment View. The structure, data relationships and definitions of all data attributes in the Opal datamart can be found in the Opal data dictionary.

1.3 Audience

The intended audience of this document includes those who are responsible for developing data cubes, data visualisations and reports on Opal journey segment data for ongoing analysis, and those who wish to analyse journey segment data to answer specific journey-related questions.

1.4 Related documents

No	Document Ref	Document Title	Author
1		Opal data dictionary, v0.5, June 2016	Cathy Agius

2 Data dictionary

Table 1: BTS_DW.JS_V (Journey Segment View)

The attributes available through the journey segment view are defined in detail below. Unless specified otherwise, attributes with the following suffixes contain values taken directly from data extracts provided by Cubic:

- _ID
- _CD
- _DESC
- _NM
- _KEY
- _DID

Column	Attribute name	Attribute definition	Data type
CIN	Hashed Card Identification Number	A unique identification number for the Opal card, hashed/masked by Cubic for privacy purposes and different from the card number printed on the card	VARCHAR2 (50 BYTE)
CARD_FK	Card Foreign Key	Code which references the Card dimension	NUMBER
PSNGR_TYP_CD	Passenger Type Code	Passenger type (such as 'Adult', 'Child/Youth')	VARCHAR2 (50 BYTE)
CARD_TYP_CD	Card Type Code	Card type (such as 'Adult', 'Concession', 'Child/Youth', 'Senior/Pensioner')	VARCHAR2 (50 BYTE)
JRNY_ID	Journey ID	An identification number for the journey	NUMBER (5)
SGMNT_ID	Segment ID	An identification number for the journey segment	NUMBER (5)
JS_STRT_DT_FK	Journey Segment Start Date Foreign Key	A code which references the Date dimension, expressed in the format YYYYMMDD	NUMBER (8)
JS_STRT_TM	Journey Segment Start Time	The time the journey segment commenced, expressed in the format HH:MM:SS	VARCHAR2 (30 BYTE)
DISC_ENT_CD	Discount Entitlement Code	A code representing the discount entitlement type (such as '0', '1', '2'), linked to DISC_ENT_DESC	NUMBER (3)

Column	Attribute name	Attribute definition	Data type
DISC_ENT_DESC	Discount Entitlement Description	A description of the discount entitlement (such as 'Concession', 'Pensioner', 'No Discount')	VARCHAR2 (50 BYTE)
TS_TYP_CD	Transit Stop Type Code	A description of the type of transit stop (such as 'Train', 'Bus', 'Ferry', 'Light rail')	VARCHAR2 (20 BYTE)
TAG1_TS_TYP_CD	Tag 1 Transit Stop Type Code	A description of the type of transit stop associated with the journey segment's tap on (such as 'Train', 'Bus', 'Ferry', 'Light rail)	VARCHAR2 (20 BYTE)
TAG2_TS_TYP_CD	Tag 2 Transit Stop Type Code	A description of the type of transit stop associated with the journey segment's tap off (such as 'Train', 'Bus', 'Ferry', 'Light rail	VARCHAR2 (20 BYTE)
OPRTR_ID	Operator ID	A numeric code used to identify the operator (such as '0', '1', '2'), provides link to OPRTR_SHORT_NM and OPRTR_FULL_NM	NUMBER (10)
OPRTR_SHORT_NM	Operator Short Name	An abbreviation representing the operator (such as 'GNC')	VARCHAR2 (20 BYTE)
OPRTR_FULL_NM	Operator Full Name	The full name of the operator (such as 'Greens Northern Coaches')	VARCHAR2 (50 BYTE)
JS_TYP_CD	Journey Segment Type Code	A description of the journey segment record type (such as 'CompleteJourneySegment', 'PartialJourneySegment', 'LateReconstructedJourneySegment')	VARCHAR2 (40 BYTE)
JS_FARE_CENTS_AMT	Journey Segment Fare Cents Amount	The journey segment fare expressed in cents	NUMBER (10)
JS_DURN_SEC	Journey Segment Duration Seconds	The length of time between the journey segment start time and journey segment end time expressed in seconds	NUMBER (10)
IMTT_CD	Intermodal Transfer Type Code	A code representing the intermodal transfer type related to the journey segment (such as '0', '1', '2') for paid trips and links to the IMTT_CD description	NUMBER (5)
IMTT_DESC	Intermodal Transfer Type Description	A description of the intermodal transfer type (such as 'Ferry to rail', 'Bus to rail', 'No inter-modal transfer') for paid trips, based on IMTT_CD	VARCHAR2 (100 BYTE)
JS_TAG1_TYP_DESC	Journey Segment Tag 1 Type Description	A description of the tag type associated with the first tag for the journey segment (such as 'Purse Tag On')	VARCHAR2 (30 BYTE)
JS_TAG2_TYP_DESC	Journey Segment Tag 2 Type Description	A description of the tag type associated with the second tag for the journey segment (such as 'Purse Tag Off')	VARCHAR2 (30 BYTE)

Column	Attribute name	Attribute definition	Data type
TAG1_SEQ_NUM	Tag1 Sequence Number	The sequence number associated with the first tag in the journey segment record. Transactions on each Opal card are numbered sequentially from the first time it is used. Increments additional to TAG2_SEQ_NUM. Types of transactions include tap-on, tap-off, top-up, and auto top-up set-up.	NUMBER (5)
TAG2_SEQ_NUM	Tag2 Sequence Number	The sequence number associated with the second tag in the journey segment record. Transactions on each Opal card are numbered sequentially from the first time it is used. Increments additional to TAG1_SEQ_NUM. Types of transactions include tap-on, tap-off, top-up, and auto top-up set-up.	NUMBER (5)
TAG1_DT_FK	Tag1 Date Foreign Key	A code, relevant to the first tag for the journey segment, which references the Date dimension, expressed in the format YYYYMMDD	NUMBER (8)
TAG2_DT_FK	Tag2 Date Foreign Key	A code, relevant to the second tag for the journey segment, which references the Date dimension, expressed in the format YYYYMMDD	NUMBER (8)
TAG1_TM	Tag 1 Time	The time of the first tag for the journey segment, expressed in the format HH:MM:SS	VARCHAR2 (30 BYTE)
TAG2_TM	Tag 2 Time	The time of the second tag for the journey segment, expressed in the format HH:MM:SS	VARCHAR2 (30 BYTE)
TAG1_CARD_RDNG_VLDN_RSLT_CD	Tag 1 Card Reading Validation Result Code	A numeric code used to identify the card reading validation result for the second tag of the journey segment. Only '0' = 'Validation Passed' records included.	NUMBER (3)
TAG2_CARD_RDNG_VLDN_RSLT_CD	Tag 2 Card Reading Validation Result Code	A numeric code used to identify the card reading validation result for the second tag of the journey segment. Only '0' = 'Validation Passed' records included.	NUMBER (3)
BUS_TRIP_SK	Bus Trip Surrogate Key	An identifier for the bus trip, unique within the daily Bus Trip dimension table	NUMBER
BUS_OPRTR_FK	Bus Operator Foreign Key	Code which references the Bus Operator dimension and identifies the bus operator associated with the Bus Trip (such as '1', '17', '20')	NUMBER
BUS_OPRTR_SHORT_NM	Bus Operator Short Name	An abbreviation representing the bus operator (such as 'PBB')	VARCHAR2 (20 BYTE)
BUS_OPRTR_FULL_NM	Bus Operator Full Name	The full name of the bus operator (such as 'Punchbowl Bus')	VARCHAR2 (50 BYTE)
CUBIC_JRNY_DID	CUBIC Journey Data ID	The Journey Data ID referencing the bus trip to which the journey segment relates	NUMBER (10)
CUBIC_DUTY_DID	CUBIC Duty Data ID	The Duty Data ID referencing the bus duty (shift) to which the journey segment relates	NUMBER (10)

Column	Attribute name	Attribute definition	Data type
TRIP_ID	Trip ID	An identifier for the timetabled trip, as defined in TODIS	VARCHAR2 (12 BYTE)
ROUTE_ID	Route ID	A code identifying the timetabled bus route, as defined in TODIS	VARCHAR2 (12 BYTE)
ROUTE_VAR_ID	Route Variant ID	A code indicating the timetabled route variant, as defined in TODIS	VARCHAR2 (12 BYTE)
ROUTE_VAR_NM	Route Variant Name	A description of the timetabled route variant (such as 'Macquarie - Hornsby Via Ryde Rd'), as defined in TODIS	VARCHAR2 (100 BYTE)
ROUTE_VAR_TYP_CD	Route Variant Type Code	A code indicating the timetabled route variant type, as defined in TODIS	VARCHAR2 (10 BYTE)
ROUTE_VAR_TYP_NM	Route Variant Type Name	A description of the timetabled route variant type (such as 'Regular', 'School', 'Express', 'Limited Stops'), as defined in TODIS	VARCHAR2 (50 BYTE)
RUN_DIR_CD	Run Direction Code	A code representing the timetabled run direction (such as '1', '2', '3'), as defined in TODIS	NUMBER (10)
RUN_NUM	Run Number	A code indicating the timetabled run number (such as 'M002', 'L179'), as defined in TODIS	VARCHAR2 (15 BYTE)
SHIFT_ID	Shift ID	A code indicating the timetabled shift (such as 'MB328', 'LP758'), as defined in TODIS	VARCHAR2 (12 BYTE)
DRIVER_CONSOLE_ID	Driver Console ID	Identification number of the timetabled bus driver console, as defined in TODIS	VARCHAR2 (6 BYTE)
DRIVER_ID	Driver ID	A code representing the timetabled driver, as defined in TODIS	VARCHAR2 (6 BYTE)
DEPOT_ID	Depot ID	The unique identifier for the bus depot that the timetabled bus operates out of, as defined in TODIS	NUMBER (3)
BUS_ID	Bus ID	A code identifying the timetabled vehicle	VARCHAR2 (6 BYTE)
CONTRACT_ID	Contract ID	A numeric code identifying the timetabled contract (such as '2441', '2444', '2453), as defined in TODIS	VARCHAR2 (12 BYTE)
CONTRACT_CD	Contract Code	A code indicating the contract associated with the route variant (such as 'SMBSC009', 'SMBSC013', 'SMBSC007')	VARCHAR2 (50 BYTE)
TAG1_TS_FK	Tag1 Transit Stop Foreign Key	A code, relating to the first tag of the journey segment, which references the Transit Stop dimension table	NUMBER

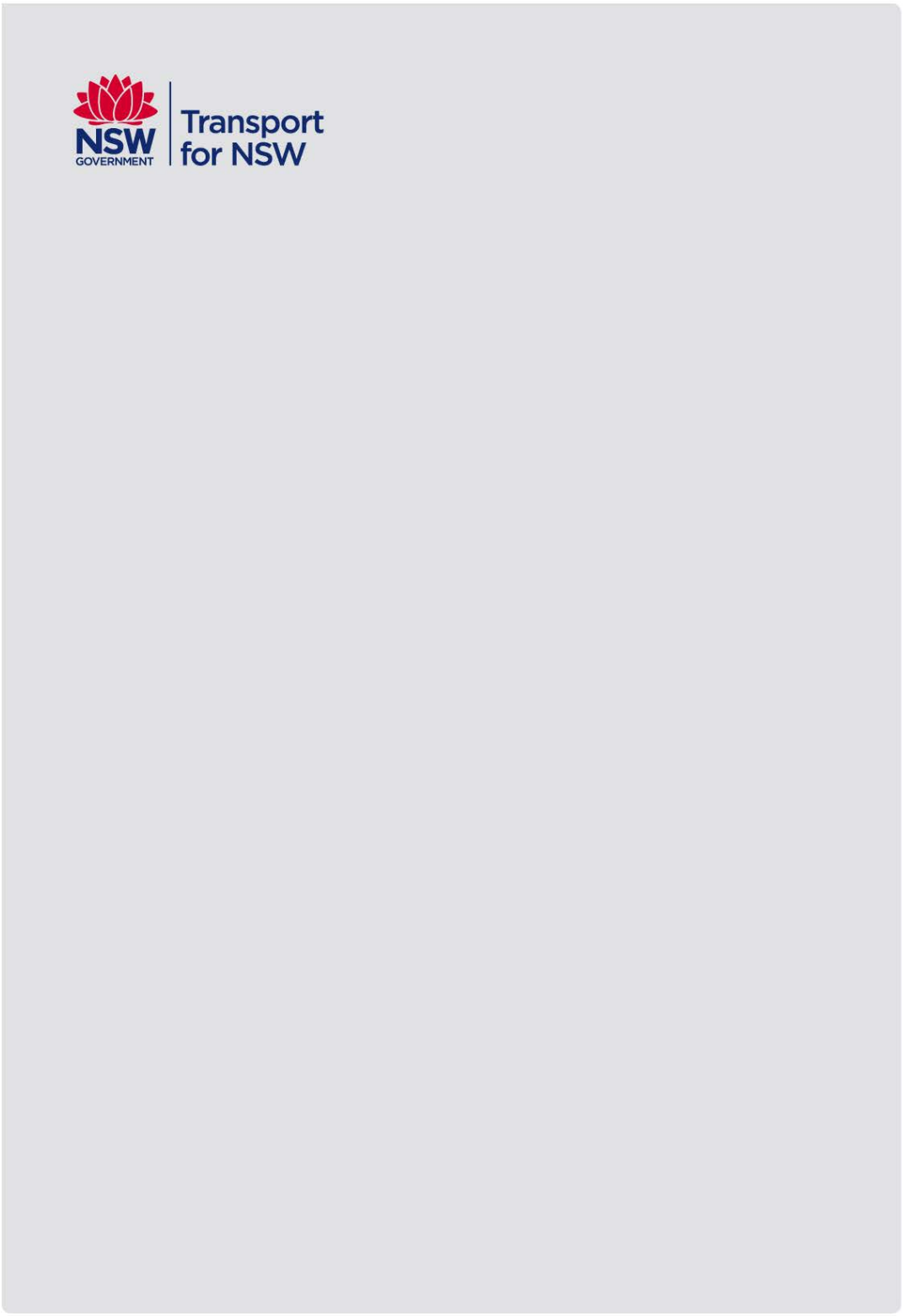
Column	Attribute name	Attribute definition	Data type
TAG1_TS_NUM	Tag 1 Transit Stop Number	A unique code used to identify the transit stop where the first tag of the journey segment occurred, as defined in Transit Stop Management (TSM)	NUMBER (7)
TAG1_TS_NM	Tag 1 Transit Stop Name	A description of the transit stop where the first tag of the journey segment occurred (such as 'Town Hall House, Druitt St, Stand L'). This description often indicates the Transit Stop location and is sourced from TSM	VARCHAR2 (255 BYTE)
TAG1_LOC_FK	Tag 1 Location Foreign Key	A code for the location of the first tag of the journey segment, referencing the Location dimension table	NUMBER
TAG1_LAT_VAL	Tag 1 Latitude Value	The latitude coordinate indicating the latitudinal position of the location relevant to the first tag of the journey. Sourced from TSM based on the Tag 1 Transit Stop Number	NUMBER (10,6)
TAG1_LONG_VAL	Tag 1 Longitude Value	The longitude coordinate indicating the longitudinal position of the location relevant to the first tag of the journey. Sourced from TSM based on the Tag 1 Transit Stop Number	NUMBER (10,6)
TAG2_TS_FK	Tag2 Transit Stop Foreign Key	A code, relating to the second tag of the journey segment, which references the Transit Stop dimension table	NUMBER
TAG2_TS_NUM	Tag 2 Transit Stop Number	A unique code used to identify the transit stop where the second tag of the journey segment occurred, as defined in TSM	NUMBER (7)
TAG2_TS_NM	Tag 2 Transit Stop Name	A description of the transit stop where the second tag of the journey segment occurred (such as 'Victoria Rd near Toelle St'). This description often indicates the Transit Stop location and is sourced from TSM	VARCHAR2 (255 BYTE)
TAG2_LOC_FK	Tag 2 Location Foreign Key	A code for the location of the second tag of the journey segment, referencing the Location dimension table	NUMBER
TAG2_LAT_VAL	Tag 2 Latitude Value	The latitude coordinate indicating the latitudinal position of the location relevant to the second tag of the journey. Sourced from TSM based on the Tag 2 Transit Stop Number	NUMBER (10,6)
TAG2_LONG_VAL	Tag 2 Longitude Value	The longitude coordinate indicating the longitudinal position of the location relevant to the second tag of the journey. Sourced from TSM based on the Tag 2 Transit Stop Number	NUMBER (10,6)
TAG1_CUBIC_HOST_DEVICE_KEY	Tag 1 Cubic Host Device Key	A code used to identify the host device used for the first tag of the journey. (Currently not applicable to Bus journey segment records)	NUMBER (10)

Column	Attribute name	Attribute definition	Data type
TAG1_DEVICE_LOC_ID	Tag 1 Device Location ID	A code used to identify the location of the device used for the first tag of the journey. (Currently not applicable to Bus journey segment records)	VARCHAR2 (10 BYTE)
TAG1_RTD_ID	Tag 1 Remote Ticketing Device ID	The unique ID used to identify the remote ticketing device used for the first tag of the journey. (Currently only applicable to Bus journey segment records)	NUMBER (10)
TAG1_DEVICE_TYP_DESC	Tag 1 Device Type Description	A description of the type of device used for the first tag of the journey (such as 'Gate Reader', 'Fixed Location Reader')	VARCHAR2 (30 BYTE)
TAG2_CUBIC_HOST_DEVICE_KEY	Tag 2 Cubic Host Device Key	A code used to identify the host device used for the second tag of the journey. (Currently not applicable to Bus journey segment records)	NUMBER (10)
TAG2_DEVICE_LOC_ID	Tag 2 Device Location ID	A code used to identify the location of the device used for the second tag of the journey. (Currently not applicable to Bus journey segment records)	VARCHAR2 (10 BYTE)
TAG2_RTD_ID	Tag 2 Remote Ticketing Device ID	The unique ID used to identify the remote ticketing device used for the second tag of the journey. (Currently only applicable to Bus journey segment records)	NUMBER (10)
TAG2_DEVICE_TYP_DESC	Tag 2 Device Type Description	A description of the type of device used for the second tag of the journey (such as 'Gate Reader', 'Fixed Location Reader')	VARCHAR2 (30 BYTE)
TAG1_DIST_BAND_CD	Tag 1 Distance Band Code	A numeric code used to identify the distance band for the first tag of the journey segment. Not currently in use	NUMBER (3)
TAG1_DIST_BAND_DESC	Tag 1 Distance Band Description	A textual description of the distance band. Not currently in use	VARCHAR2 (50 BYTE)
TAG2_DIST_BAND_CD	Tag 2 Distance Band Code	A numeric code used to identify the distance band applicable to the journey segment	NUMBER (3)
TAG2_DIST_BAND_DESC	Tag 2 Distance Band Description	A description of the distance band for the journey segment (such as 'Opal Train 5', 'Opal Bus 3', 'Opal Ferry 2', 'Opal Light Rail 1')	VARCHAR2 (50 BYTE)
TAG1_TAG_FLAG_SK	Tag 1 Tag Flag Surrogate Key	An identifier, unique within the Tag Flag dimension table, used to represent tag indicator data for the first tag of the journey segment. Each Tag Flag Surrogate Key value represents a unique combination of true/false values for a set of tag-related indicators	NUMBER
TAG1_AUTOLD_DISC_APPLD_IND	Tag 1 Autoload Discount Applied Indicator	Indication of whether an autoload discount was applied in relation to the first tag of the journey segment	VARCHAR2 (10 BYTE)

Column	Attribute name	Attribute definition	Data type
TAG1_DAILY_CAP_APPLD_IND	Tag 1 Daily Cap Applied Indicator	Indication of whether a daily cap discount was applied in relation to the first tag of the journey segment	VARCHAR2 (10 BYTE)
TAG1_FARE_DIST_BAND_APPLD_IND	Tag 1 Fare Distance Band Applied Indicator	Indication of whether a fare distance band was applied in relation to the first tag of the journey segment	VARCHAR2 (10 BYTE)
TAG1_FARE_FLAT_IND	Tag 1 Fare Flat Indicator	Indication of whether a flat fare was applied in relation to the first tag of the journey segment	VARCHAR2 (10 BYTE)
TAG1_FOU_DISC_APPLD_IND	Tag 1 Frequency of Use Discount Applied Indicator	Indication of whether a frequency of use discount was applied in relation to the first tag of the journey segment	VARCHAR2 (10 BYTE)
TAG1_OFF_PEAK_IND	Tag 1 Off Peak Indicator	Indication of whether the first tag of the journey segment meets business rules defining the journey segment as 'off-peak'. Not currently in use	VARCHAR2 (10 BYTE)
TAG1_TRNFR_DISC_IND	Tag 1 Transfer Discount Indicator	Indication of whether a transfer discount was applied in relation to the first tag of the journey segment	VARCHAR2 (10 BYTE)
TAG1_TRNFR_IND	Tag 1 Transfer Indicator	Indication of whether the first tag of the journey segment represents a transfer	VARCHAR2 (10 BYTE)
TAG1_WK_CAP_APPLD_IND	Tag 1 Weekly Cap Applied Indicator	Indication of whether the weekly fare cap has been reached during the journey to which the first tag of the journey segment relates	VARCHAR2 (10 BYTE)
TAG2_TAG_FLAG_SK	Tag 2 Tag Flag Surrogate Key	An identifier, unique within the Tag Flag dimension table, used to represent tag indicator data for the second tag of the journey segment. Each Tag Flag Surrogate Key value represents a unique combination of true/false values for a set of tag-related indicators	NUMBER
TAG2_AUTOLD_DISC_APPLD_IND	Tag 2 Autoload Discount Applied Indicator	Indication of whether an autoload discount was applied in relation to the second tag of the journey segment	VARCHAR2 (10 BYTE)
TAG2_DAILY_CAP_APPLD_IND	Tag 2 Daily Cap Applied Indicator	Indication of whether a daily cap discount was applied in relation to the second tag of the journey segment	VARCHAR2 (10 BYTE)
TAG2_FARE_DIST_BAND_APPLD_IND	Tag 2 Fare Distance Band Applied Indicator	Indication of whether a fare distance band was applied in relation to the second tag of the journey segment	VARCHAR2 (10 BYTE)
TAG2_FARE_FLAT_IND	Tag 2 Fare Flat Indicator	Indication of whether a flat fare was applied in relation to the second tag of the journey segment	VARCHAR2 (10 BYTE)

Column	Attribute name	Attribute definition	Data type
TAG2_FOU_DISC_APPLD_IND	Tag 2 Frequency of Use Discount Applied Indicator	Indication of whether a frequency of use discount was applied in relation to the second tag of the journey segment	VARCHAR2 (10 BYTE)
TAG2_OFF_PEAK_IND	Tag 2 Off Peak Indicator	Indication of whether the second tag of the journey segment meets business rules defining the journey segment as 'off-peak'. Only applicable for Train journey segments	VARCHAR2 (10 BYTE)
TAG2_TRNFR_DISC_IND	Tag 2 Transfer Discount Indicator	Indication of whether a transfer discount was applied in relation to the second tag of the journey segment	VARCHAR2 (10 BYTE)
TAG2_TRNFR_IND	Tag 2 Transfer Indicator	Indication of whether the second tag of the journey segment represents a transfer	VARCHAR2 (10 BYTE)
TAG2_WK_CAP_APPLD_IND	Tag 2 Weekly Cap Applied Indicator	Indication of whether the weekly fare cap has been reached during the journey to which the second tag of the journey segment relates	VARCHAR2 (10 BYTE)
DISC_CENTS_AMT	Discount Cents Amount	The discount amount, in cents, applied to the journey segment	NUMBER (10)
SAF_CENTS_AMT	Station Access Fee Cents Amount	The amount, in cents, of the station access fee applicable to the journey segment	NUMBER (10)
TAG1_TCKT_TYP_DESC	Tag 1 Ticket Type Description	For ticket-type Opal cards, a description of the ticket (such as 'Employee with SAF', 'Single Trip Rail without SAF', 'SSTS'). Where the Opal card used for the journey segment is a stored value ("purse-type") card, this field shows '-1'	VARCHAR2 (50 BYTE)
TAG1_GEO_VLDTY_TYP_DESC	Tag 1 Geographic Validity Type Description	For ticket-type Opal cards, a description indicating whether the ticket is geographically constrained (such as 'Ticket is not constrained geographically', 'Ride-based rail ticket, distance from the tag-on to tag-off')	VARCHAR2 (200 BYTE)
TAG1_RMNG_RIDES_CNT	Tag 1 Remaining Rides Count	The number of rides remaining on ride-based, ticket-type Opal cards (such as Opal single trip tickets) at the time of the first tag of the journey segment	NUMBER (3)
TAG1_DAYS_USED_CNT	Tag 1 Days Used Count	Not currently in use as this field relates to multi-day ticket-type Opal cards which are not currently available	NUMBER (3)
TAG1_DAILY_SAF_CAP_APPLD_IND	Tag 1 Daily Station Access Fee Cap Applied Indicator	Indication of whether a daily station access fee cap was applied in relation to the first tag of the journey segment. As there is currently no daily station access fee cap, this field shows 'false'	VARCHAR2 (10 BYTE)

Column	Attribute name	Attribute definition	Data type
TAG1_WK_SAF_CAP_APPLD_IND	Tag 1 Weekly Station Access Fee Cap Applied Indicator	Indication of whether a weekly station access fee cap was applied in relation to the first tag of the journey segment	VARCHAR2 (10 BYTE)
TAG1_DAYS_USED_TCKT_IND	Tag 1 Days Used Ticket Indicator	Not currently in use as this field relates to multi-day ticket-type Opal cards which are not currently available	VARCHAR2 (10 BYTE)
TAG2_TCKT_TYP_DESC	Tag 2 Geographic Validity Type Description	For ticket-type Opal cards, a description of the ticket. Where the Opal card used for the journey segment is a stored value ("purse-type") card, this field shows '-1'	VARCHAR2 (50 BYTE)
TAG2_GEO_VLDTY_TYP_DESC	Tag 2 Geographic Validity Type Description	For ticket-type Opal cards, a description indicating whether the ticket is geographically constrained (such as 'Ticket is not constrained geographically', 'Ride-based rail ticket, distance from the tag-on to tag-off')	VARCHAR2 (200 BYTE)
TAG2_RMNG_RIDES_CNT	Tag 1 Remaining Rides Count	The number of rides remaining on ride-based, ticket-type Opal cards (such as Opal single trip tickets) at the time of the second tag of the journey segment	NUMBER (3)
TAG2_DAYS_USED_CNT	Tag 2 Days Used Count	Not currently in use as this field relates to multi-day ticket-type Opal cards which are not currently available	NUMBER (3)
TAG2_DAILY_SAF_CAP_APPLD_IND	Tag 2 Daily Station Access Fee Cap Applied Indicator	Indication of whether a daily station access fee cap was applied in relation to the second tag of the journey segment. As there is currently no daily station access fee cap, this field shows 'false'	VARCHAR2 (10 BYTE)
TAG2_WK_SAF_CAP_APPLD_IND	Tag 2 Weekly Station Access Fee Cap Applied Indicator	Indication of whether a weekly station access fee cap was applied in relation to the second tag of the journey segment	VARCHAR2 (10 BYTE)
TAG2_DAYS_USED_TCKT_IND	Tag 2 Days Used Ticket Indicator	Not currently in use as this field relates to multi-day ticket-type Opal cards which are not currently available	VARCHAR2 (10 BYTE)



Annexure D – NSW Government Open Data Policy

NSW Government

Open Data Policy

Vision

Better, faster, more open data

NSW Government agencies:

Better	Release **better** data in accessible, consumable formats with metadata and quality statements
Faster	Release data **faster** using automated processes, standard data categories and trusted user models
More	Release **more** data and make it discoverable through central portals

Why open data?

Open data is a tremendous resource. It provides the intelligence for insight, invention and exploration that translate into better products and services that improve everyday life and encourage business growth.

Research shows that open data has a significant impact in four key areas:

- improving government
- empowering citizens
- creating opportunity
- solving problems.

The NSW Open Data Policy has contributed to improved outcomes for vulnerable populations protecting the environment, good government, and road safety.

The review of the policy in 2016 renews the NSW Government commitment to open data to support; transparency, innovation, and new knowledge and insights from combined data sources.

Introduction

The Open Data environment

This policy is part of the data environment underpinning Government initiatives which are leading to business growth and to improvements in everyday life. Open Government, Smart Cities, the Data Analytics Centre and the NSW Open Data Advocate are some of these initiatives.

Open Government

A smart government is transparent and accountable, and understands that solutions to policy challenges can come from outside government. It collaborates with community and industry partners, encourages participation in the policy formulation and decision making process, and encourages innovation from employees and external stakeholders.

Public service agencies are now more open and collaborative as a result of the reforms commenced under the Strategy. By 2014, the amount of information the NSW Government published online had more than doubled, more agencies conducted customer satisfaction surveys than in previous years, and most NSW Government agencies now had social media policies in place to better engage with customers. This shows a deeper involvement with the community and a commitment to ensuring that there are multiple contact points within government for customers.

Smart Cities

Smart cities are generally considered in terms of using technology to connect infrastructure such as transport, health, energy and water, generating data that can be used for better services and better planning. Open data makes this information available to everyone, from families to retailers, commuters to investors. Open data allows everyone to understand and interact with our cities, to contribute to their growth and development.

NSW Data Analytics Centre (DAC)

The DAC facilitates data sharing between agencies and manages whole-of-government analytics projects. It does that by leveraging internal and external partnerships so that the right capabilities, tools and technologies are applied.

The DAC works with data from a variety of sources, including open data, to derive insights that support strategic decision making, evidence based policy development and improved service delivery.

The DAC advises agencies on how data can be opened to the public in line with the Open Data Policy, including de-identification.

NSW Open Data Advocate

The Information Commissioner promotes public awareness and understanding of the object of the GIPA Act: to open government information to the public including authorising and encouraging the proactive public release of information by agencies.

The Information Commissioner has taken on the role of NSW Open Data Advocate. The Open Data Advocate will play a key role in ensuring that the new Open Data Policy is successfully implemented. Activating the role of Open Data Advocate will provide assistance to citizens making a request to access information, and provide advice to agencies to streamline processes for data release.

The Information Commissioner may investigate and report on the exercise by agencies of their functions under and compliance with the GIPA Act, and can investigate complaints made about an agency's conduct in carrying out its functions under the GIPA Act. This independent regulatory oversight role promotes accountability and public confidence in the release and management of data by government sector agencies.

The Information Commissioner oversees the transparency obligations of agencies in the GIPA Act including that:

- agencies must make information publicly available as open access information under Part 3 Division 1 of the GIPA Act
- disclosure logs where agencies must record information that has been requested formally and which the agency decides may be of interest to other members of the public (Part 3 Division 4).

The obligations of agencies are outlined in Part 3 Division 2 Agency Information Guides.

Progress to date

In 2009, the Government Information Public Access Act (GIPAA) established NSW Government's commitment to opening government information, proactively and responsively, to the public. Data NSW was launched as the first State level open data catalogue in the world. In 2013, the Open Data Policy was introduced to support simple and efficient compliance with GIPAA and encourage the release of high value datasets for use by industry and the community. The Policy aimed to assist agencies in embedding open data principles in their operations.

Data NSW now receives between 3,000 and 5,000 visits every month. The site contains approximately 350 catalogue entries which provide access to over 50,000 government datasets, which are published on agency portals. This approach allows open data to be searched in a central location, but also made available alongside data on the same subject – such as education, health or environment. Data NSW has also been federated with open data portals from other jurisdictions so that a search will provide links to datasets from South Australia, Queensland and the Commonwealth Government.

NSW Government has adopted an open access licensing framework to support the release and reuse of public information. We have also adopted the Australian Bureau of Statistics framework for standard data quality reporting.

The 2016 policy drives the release of more data, in better open formats, using automated processes for faster publication. The newly established Open Data Rolling Release Schedule indicates agency targets to improve the openness of their published data and to release new, high value datasets.

Case studies

Progress is demonstrated by the following three case studies.

Transport for NSW established the NSW transport data exchange (TDX) program. Data is available in a range of machine-readable, spatially enabled formats. TDX provides real time data on:

- timetables, routes and stops
- where your train, bus, ferry or light rail service is right now and estimated arrival times
- live traffic updates and incidents
- travel times
- scheduled road work.

Land and Property Information is now providing spatial data, such as the NSW Cadastre, imagery and administrative boundaries, as web services. The NSW Address Location Service web service allows the user to enter an address and pinpoint the location of that address. The web services allow users to access the most up to date information without having to store and maintain it in their own systems or applications. The web services can be used in multiple ways e.g. direct use, mashups, mobile apps and spatial systems.

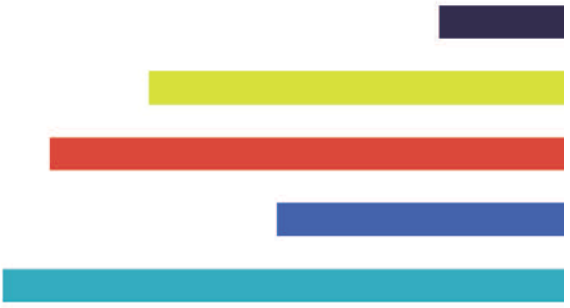
The NSW Education Datahub is a searchable central repository of publically available data with a range of video tutorials on using it. Department of Education is also a partner in the Code for Australia fellowship program. The Fellowship Program pairs civic-minded technologists, designers and community organisers with a government agency for a 12-month collaboration. For example, to develop a service that uses education and location data to make it easier for parents and carers to select a school for their children.

Information Management Framework

The NSW Government Information Management Framework (the IM Framework) establishes a common approach to managing government data and information.

The IM Framework is made up of policies, standards and guidance for managing data and information consistently across the sector. It enables agencies to appropriately protect, share and publish data and information in ways that support better services, and to derive the best value from investments in data, information and technology.

This policy forms part of the IM Framework.



NSW Government

Open Data Policy

Objectives

The objectives of this policy are to assist NSW Government agencies to:

- release data for use by the community, research, business and industry
- accelerate the use of data to derive new insights for better public services
- embed open data into business-as-usual
- use data to inform the design of policy, programs and procurement
- support the use of data by the NSW Data Analytics Centre for research and evidence-based decision making
- advance citizen engagement with government and the work of government
- support the *Government Information (Public Access) Act 2009 (NSW) (GIPA Act)* and promote simple and efficient compliance with the requirements set out in that Act.

What is open data?

This policy focuses on datasets as an aspect of information defined in the *Government Information (Public Access) Act 2009 (NSW) (GIPA Act).*

A dataset is an identifiable collection of government held information or data. Most commonly a dataset corresponds to the contents of a single database table, or a single statistical data matrix. The term can also be used to refer to the data in a collection of closely related tables. A dataset may comprise a smaller grouping (or subset) of data which, though limited by some constraint or feature type, is located physically within a larger dataset.

Data is open to the extent that its management, release and characteristics meet the principles of openness outlined in this policy. Open data should be both technically available and usable, and have licensing frameworks in place to facilitate its release and use.

Scope

This policy applies to all NSW Government agencies including cluster agencies, Departments, Statutory Bodies and Shared Service Providers. In accordance with *Premier's Memorandum M1999-19 Applicability of Memoranda and Circulars to State Owned Corporations*, the Policy is not mandatory for State Owned Corporations, although it is recommended for adoption. NSW Government funded research data must also be made open under this policy.

Target audiences for this policy include government agencies, non-government organisations (NGOs), academia, industry (including ICT developers) and members of the public who are interested in or have a specific use for government data.

Open Data Principles

Open data principles lead to more responsive and smarter government, and better service delivery. To meet the obligations of this policy, agencies must manage data as a strategic asset to be:

- Open by default, protected where required
- Prioritised, discoverable and usable
- Primary and timely
- Well managed, trusted and authoritative
- Free where appropriate
- Subject to public input.

Agency responsibilities

Open by default and protected where required

Agencies will start from a position of data openness, favouring the release of data, unless there is an overriding public interest against disclosure *(Government Information (Public Access) Act 2009 (NSW) (GIPA Act)*.

A Rolling Release Schedule for agencies' plans to publish open data is available on Data NSW. The Rolling Release Schedule targets improved formats for data and identifies new high value datasets for release.

Agencies must:

- Create and collect data in ways that support downstream processing and release
- Incorporate open data into project, program, service, system and policy design
- Make data available without reference to its potential end use by other parties
- Ensure they own all intellectual property rights in the dataset, or that they have a sufficiently broad licence to permit its release
- Comply with NSW Government policies for information security (e.g. NSW Government Digital Information Security Policy) and other relevant NSW Government information security and privacy policies, state and federal legislations and guidelines) and the management of privacy for the individual
- Apply information labels and security classifications which indicate whether data is protected, in line with the *NSW Government Information Classification, Labelling and Handling Guidelines*
- Ensure safeguards are in place to manage and assess the release of sensitive datasets
- Apply managed methods to support the release of otherwise sensitive data such as: redacting, consent, ethics approval, aggregating and third party brokerage (mediator or integrator).

Prioritised, discoverable and usable

High-value datasets will be prioritised for release, in line with demand from the public and industry, as a result of stakeholder consultation, or where the release of the datasets will contribute to better service delivery in NSW.

Data will be in a format that makes it easy to use, transform and reuse. Commonly accepted open data standards, are published by the World Wide Web Consortium (W3C) to enhance data usability.

When procuring or upgrading systems according to the *ICT Investment Policy and Guidelines*, agencies will consider the use of technologies that enable data to be consumed by other systems, without the need for manual intervention.

Metadata allows datasets to be found, understood, controlled and managed. It will be published with all datasets. It will define or describe content, quality, format or structure of a dataset, the system, location and context in which the dataset was produced, collected, processed or stored.

Agencies must:

- Release data which supports a range of outcomes, including; social, enhanced service delivery, agency core business, economic opportunity, generates efficiencies or reduced costs, supports evidence based research or policy
- Prioritise the release of datasets in response to public requests
- Link datasets to Data NSW for discoverability and availability
- Ensure datasets are technically open (machine readable)
- Apply creative commons licensing (default CC-BY) to facilitate the reuse of data, making it legally open
- Apply the correct data standards to manage their data. See the Information Management Framework
- Publish metadata with all datasets and the details of a contact person in their agency. Data NSW can assist with this and may be contacted through the contact form on the site
- Assess the quality of their datasets prior to release with the *Data Quality Standard*. The standard can be used to report whether datasets meet the above requirements.

NSW Government

Open Data Policy

Primary and timely

Data will be released as collected at the source, with a high level of granularity, and not in aggregate or modified forms unless required to safeguard confidential or personal data. Reuse will acknowledge the source and include the metadata.

Agencies must:

- Implement mechanisms that enable automated, ongoing or periodic regular release of data, without the need for manual intervention
- Publish live, real-time feeds where it enhances the utility of the dataset
- Include timestamps or other information for users to identify the currency of the data
- Make newer, replacement data available to users in a timely manner.

Well managed, trusted and authoritative

Users will be alerted to the quality and limitations of the data to ensure confidence that it is trustworthy and authoritative.

Data governance arrangements will be established and maintained in accordance with NSW Government's information management principles and the *Data and Information Custodianship Policy.*

Agencies must:

- Manage data in accordance with legislative and legal requirements, including those for data security and protection of personal information, intellectual property, business confidentiality and legal professional privilege
- Apply sound archival practices to preserve data.

Free where appropriate

Access to open data will be free by default. The widespread use of data for innovation will be encouraged to achieve the maximum value from the data for the people of NSW, and to enhance transparency of government.

Agencies must:

- Provide data free of charge
- Establish standard parameters for cost recovery (only for specialised data services).
- Publish any fees for data use.

Subject to public input

Agencies will engage with and are informed by the community, research sector and industry on the design of new datasets and data portals, and in deciding what data to publish and about publication practices.

Agencies must:

- Provide data in a way that is relevant to the transaction being undertaken or service being used
- Participate with activities across jurisdictions to share and integrate data at the national level.
- Publish community and industry feedback about the data.

Implementation

This Policy is supported by an action plan to assist agencies in achieving open data objectives.

This Policy is maintained by the Department of Finance, Services and Innovation.

Annexure E – Extract from Journey Segment View Release Proposal

Due Diligence Steps and Process – relevant issues

Opal Data and Privacy

Opal Journey Data, while not directly linked to a person, is linked to the use of a specific card - which generally correlates to a person or household. Every use of that card to complete a journey on public transport is recorded in the journey table.

The identity of the cardholder and any personal information about the cardholder (which is only known if the card is registered) is not included in the Journey dataset. Furthermore, the specific Opal Card Number that is printed on each Opal card is not included in the Journey dataset. Specific controls are in place to limit access to this customer data internally. This customer data would not be made open, as the Open Data Policy requires that personal information be protected on privacy grounds.

Nevertheless, each card is assigned a unique identifier (linked to the Card Number in a separate table) which distinguishes the use of that particular card from all others. A scenario could arise where, by direct observation or by a process of inference and elimination, the use of a particular card could be attributed to a specific individual. Once achieved, the use of that card could be tracked through the dataset, revealing location information for a specific individual.

Masking, encrypting or simply eliminating the unique card identifier would minimize the risk of the Journey dataset being used to easily track the use of a specific card, and therefore person - but would greatly diminish the analytical value of the dataset. Various data transformation techniques could be used to mitigate these privacy risks, but all entail some trade-off in the subsequent analytic potential of the data. Determining the most appropriate data anonymisation technique will be the primary objective of the Opal PIA.

Location Data as a quasi-identifier

Studies have demonstrated that personal travel behaviour itself is highly personal and unique, so that location data is a potent 'quasi-identifier' for a single person. That is, even completely removing any link between a journey and a card, may not prevent singling out a specific card utilising data mining and pattern-matching techniques.

The potential of being able to 'single out' a particular card is proportional to the total size of the dataset, the level of data granularity and its ‘linkability’ to other datasets. Expert advice may need to be commissioned to assess the level of risk.

Potential Data Anonymisation Techniques for Opal Journey Data

De-identification Technique	Impact Benefits / Limitations
Release whole dataset but **Remove Card Identifier**	Minimises ability to single out use of a specific card Prevents important analysis: Cannot identify multi-modal transfers Cannot identify variation individual travel behaviour across a week (eg: who is 'gaming' 8 trip limit? Do we need to track this to an individual)
Release whole dataset but **Replace Card ID with temporary random unique identifier**	Eliminates ability to track travel behaviour beyond duration of temporary identifier Allows multimodal trip analysis but only for duration of temporary identifier
Random sample	Reduces risk of tracking as a specific card has only a small probability of appearing in any given sample (Transport for London uses this methodology, releasing Journey Data for 5% of all card used on any given day) However, with significantly fewer cards in use, a statistically significant sample size would likely be much larger rendering this technique far less effective. (For a random sample to big enough to be representative of the whole, it may be as high as 20% of the whole)
Release whole dataset but **reduce granularity of chronological data**	Reduce the risk of associating a specific card with an individual person by converting actual time of tap on / tap off to time intervals (eg: time only represented in 5, 10 or 15 minute blocks). (Larger the block the greater reduction in risk, but less analytic utility.(?) Reduced granularity may impact detailed analysis in some cases
Release whole dataset but **Randomise Card ID (?)**	Randomise Card IDs tapping on at a specific location within a certain time block (15 mins - 1 day) reduces risk of associating a specific card with an individual via direct observation. Card cannot be tracked across time. Variation individual travel behaviour across a week cannot be analysed. (The larger the time block the greater the risk reduction, allows analysis within the duration of the time block.)

Annexure F - Initial Summary Report

Privacy Preserving Opal Data Release: Initial Summary Report

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

1 Introduction

Transport of New South Wales (TfNSW) plans to release part of its Opal tap on/tap off data publicly, maintaining the privacy of individuals while simultaneously providing adequate utility (ability to use) to third parties. The target dataset consists of two weeks of Opal tap on/tap off data from four modes of transport: buses, ferries, trains and light rail. We shall refer to this collectively as the Opal dataset. For the proposed release of the Opal dataset, we have chosen to use algorithms that have a formal proof of privacy associated with their output. This formal proof of privacy is known as differential privacy [3]. Based on our expertise and experience [1, 2], we strongly recommend this approach over ad hoc sanitisation of datasets. Datasets released via a differentially private algorithm ensures that all *known* classes of privacy attacks (such as, re-identification) or any discovery of new privacy attacks in the *future* have little chance of success. Ad hoc techniques may be susceptible to currently known attacks, attacks developed in the future, or to newly available background knowledge which the analyst can use to harm the privacy of users in the Opal dataset.

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Privacy Preserving Opal Data Report

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

1 Introduction

Maintaining the privacy of individuals is paramount in any data release and the challenge with Opal data is releasing privacy preserving data that still provides adequate utility (ability to use) to third parties. The target dataset discussed with Transport for NSW consists of two separate weeks of Opal tap on/tap off data from four modes of transport: buses, ferries, trains and light rail. We shall refer to this collectively as the Opal dataset. For the proposed release of the Opal dataset, we have applied an algorithm that has a formal proof of privacy associated with its output. This formal proof of privacy is known as differential privacy [3]. Based on our expertise and experience [1, 2], we strongly recommend this approach over ad hoc sanitisation of datasets. Datasets released via a differentially private algorithm ensure that all *known* classes of privacy attacks (such as, re-identification) have little chance of success. In particular and very importantly the mathematical properties of the algorithm and resulting dataset provide the same protection for new privacy attacks that may be developed in the *future*, ensuring they too have little chance of success.

It is important to note that ad hoc techniques alone such as aggregation or anonymisation may be susceptible to currently known attacks, attacks developed in the future, or to newly available background knowledge which the analyst can use to harm the privacy of users in the Opal dataset. Furthermore, no mathematical guarantees underpin such techniques.

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Re

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Privacy Preserving Opal Data Report Addendum

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

1 Introduction

This document is an addendum to the final report titled "Privacy Preserving Opal Data Release". That report provides an overview and background to the project, differential privacy, and the difficulties of applying differential privacy to the Opal data. This document is focused on the application of a second differentially private algorithm to the Opal data. It has been agreed that this application would apply this second algorithm to a reduced subsection of the Opal data to provide 6 data sets per day per mode. 14 days have been treated across the 4 different modes, being train, bus, light rail and ferry. The 6 data sets are:

- tap on time (single attribute data set)
- tap on location (single attribute data set)
- tap off time (single attribute data set)
- tap off location (single attribute data set)
- tap on time and location (double attribute data set)
- tap off time and location (double attribute data set)

This report presents those details specific to the second dataset. In particular it will give an overview of the algorithm, output data definition and an overview of utility (including information on detailed utility information provided with the dataset).

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

Redacted - Out of scope

6 Conclusions

This data set is likely to have a great general appeal when compared to the first data set. The data sets themselves are easier to use and the accuracy is significantly improved. To this end, the data sets ought to provide the functionality Transport for NSW are looking for. However it has come at a cost both in terms of the added complication of needing to work out which data set to use for a particular query, and in a reduced set of queries that can be asked. In particular the relationship between tap on and tap off has been removed. So this data set does not reveal any direct information of origin and destination of trips.

This algorithm has potential to be taken further to improve utility but requires Transport for NSW to understand the trade-offs in various approaches, particularly the potential impact on group privacy.

Redacted - Out of scope

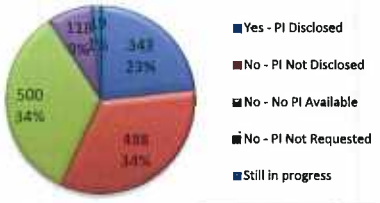
Privacy Requests Dashboard - Summary

As at 31 January 2017

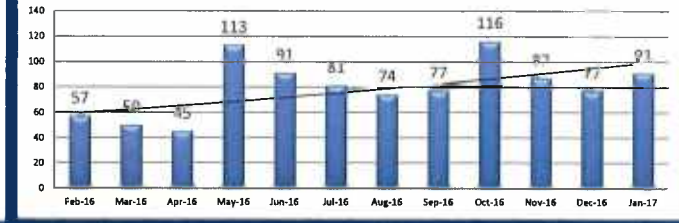
Summary of NSW Police, AFP and Dep Immigration Privacy Requests

To be actioned = 19 | Actioned = 1459 | Total requests received 1478

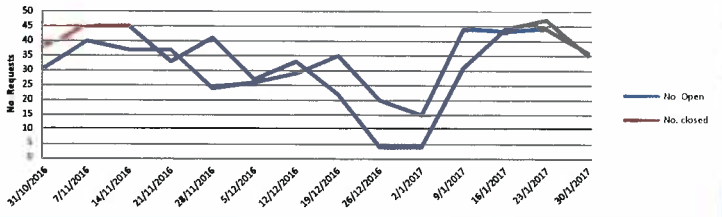
PI Disclosed - All Requests



Total No. Requests Received/Mth - 12 mth view



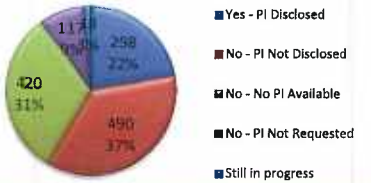
Trend - Average time to resolve all LEA requests by week - 3 mth view



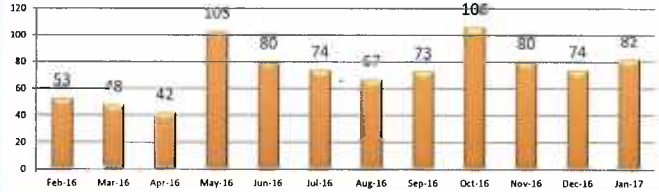
NSW Police

To be actioned = 18 | Actioned = 1325 | Total requests received 1343

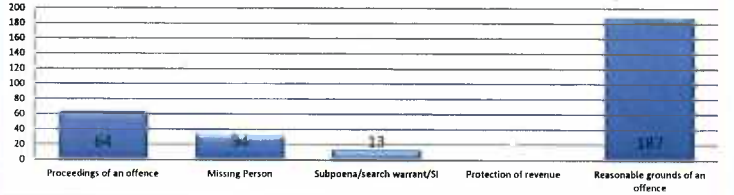
Personal Info Disclosed



No. Requests Received/Mth - 12 mth view



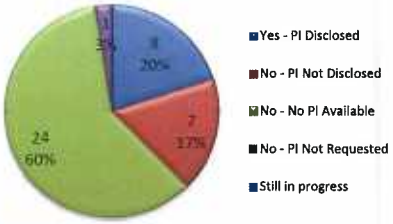
Business Rules - Yes Personal Information Disclosed



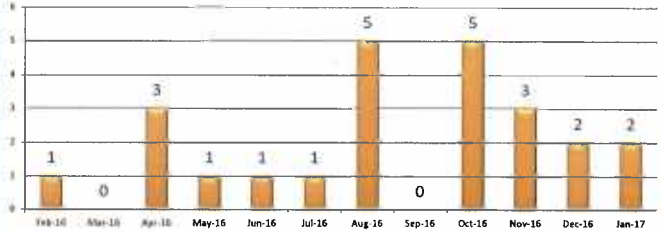
Australian Federal Police

To be actioned = 0 | Actioned = 40 | Total requests received 40

Personal Info Disclosed



No. Requests Received/Mth - 12 mth view

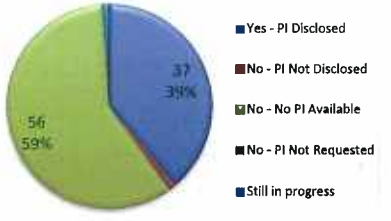


Business Rules - Yes Personal Information Disclosed

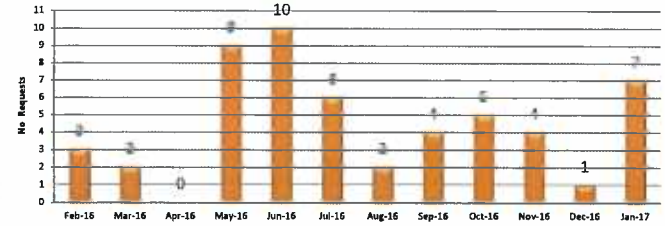
Department of Immigration

To be actioned = 1 | Actioned = 94 | Total requests received 95

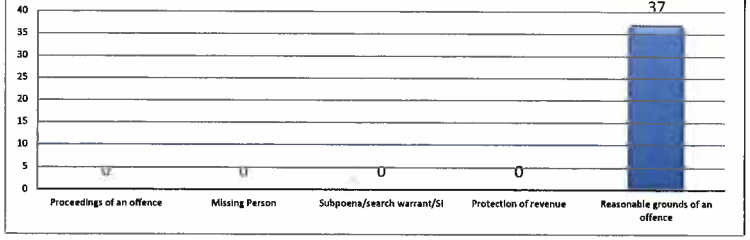
Personal Info Disclosed



No. Requests Received/Mth - 12mth view



Business Rules - Yes Personal Information Disclosed

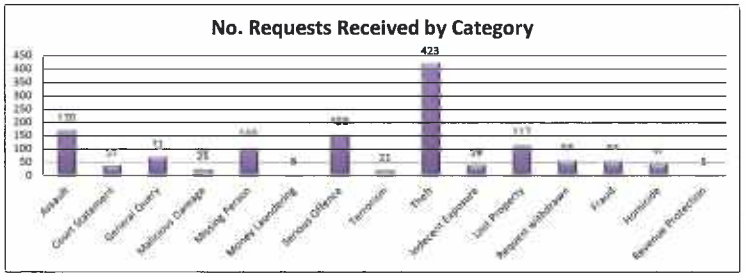


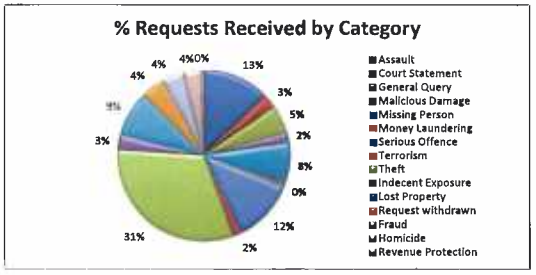
NSW Police Requests - Dashboard

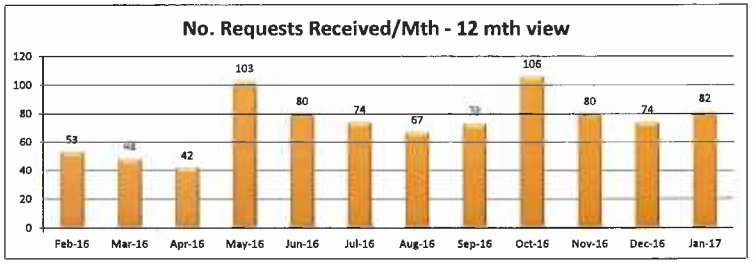
30 April 2014 - 31 January 2017

No. Requests Currently Open:	18	No. Requests Currently Complete:	1325	Total Requests Received to Date:	1343

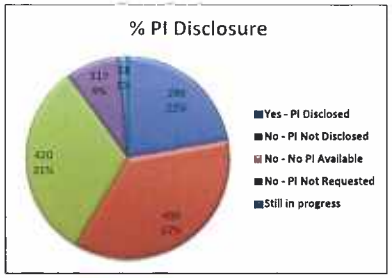
No. Requests Information

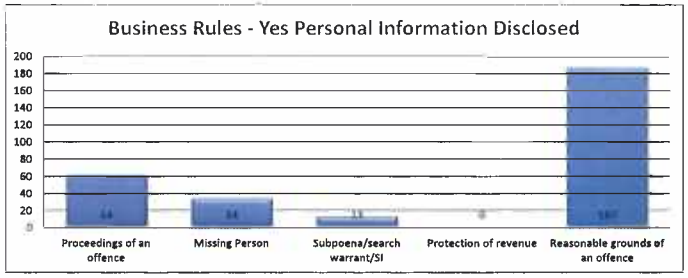


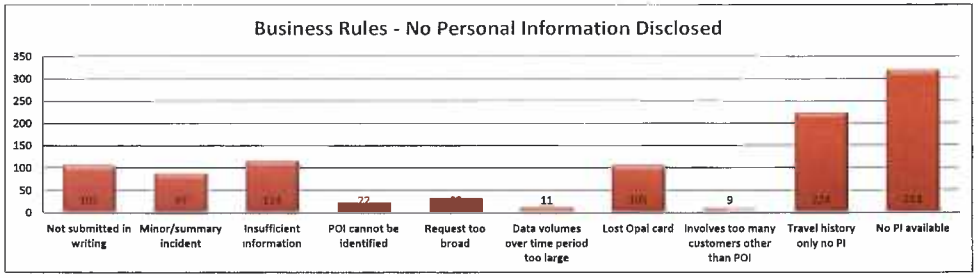




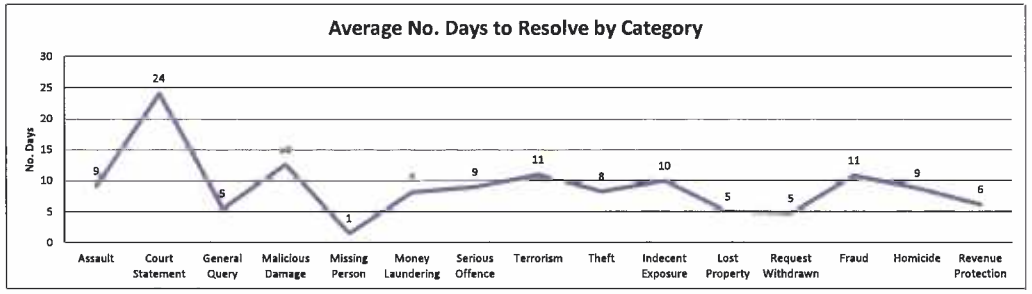
Personal Information Disclosed/Not Disclosed

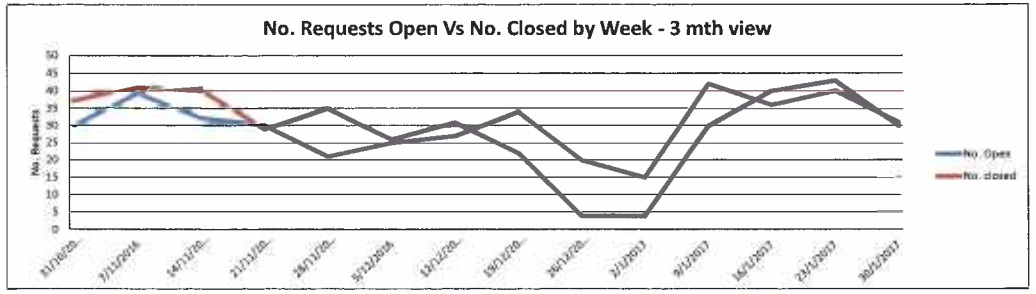






Trends



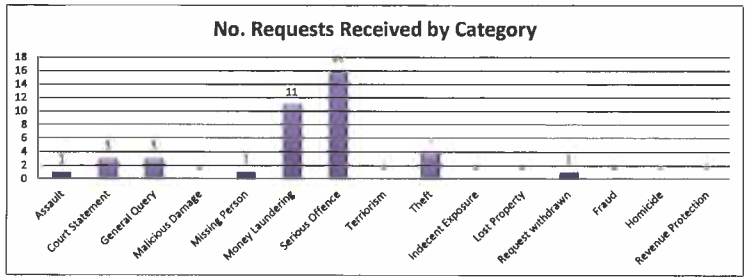


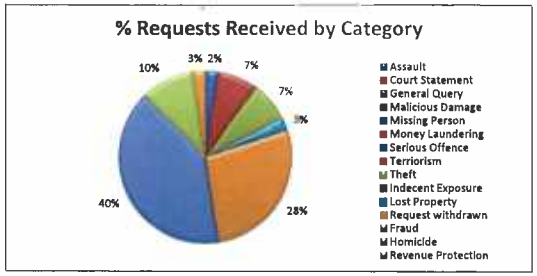
Australian Federal Police Requests - Dashboard

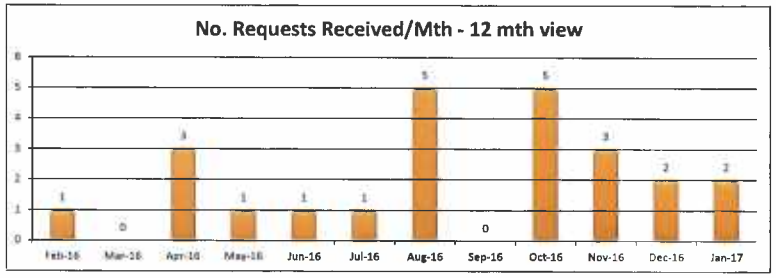
13 November 2014 - 31 January 2017

No. Requests Currently Open:	0	No. Requests Currently Complete:	40	Total Requests Received to Date	40

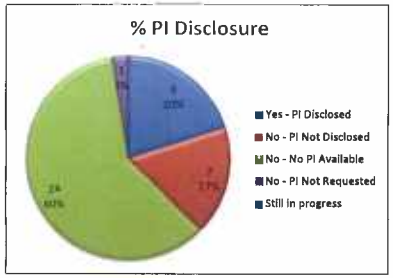
No. Requests Information

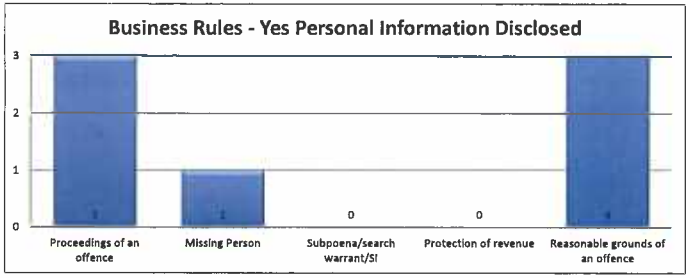


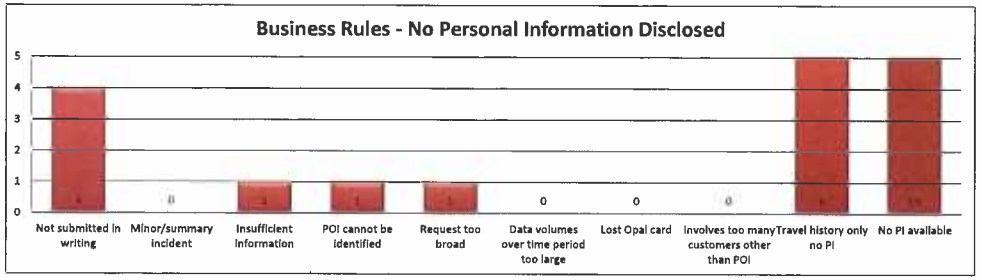




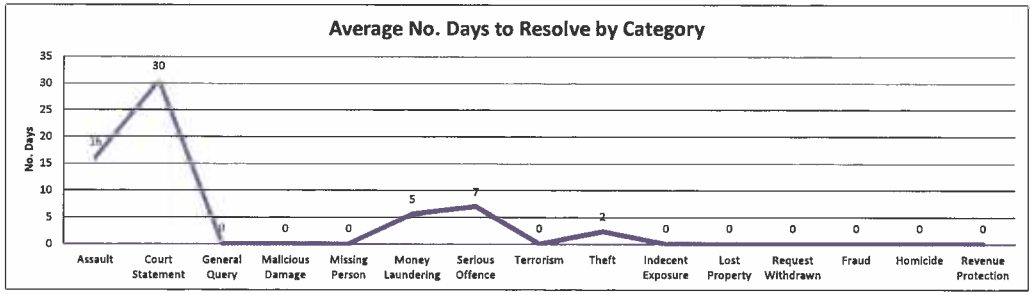
Personal Information Disclosed/Not Disclosed

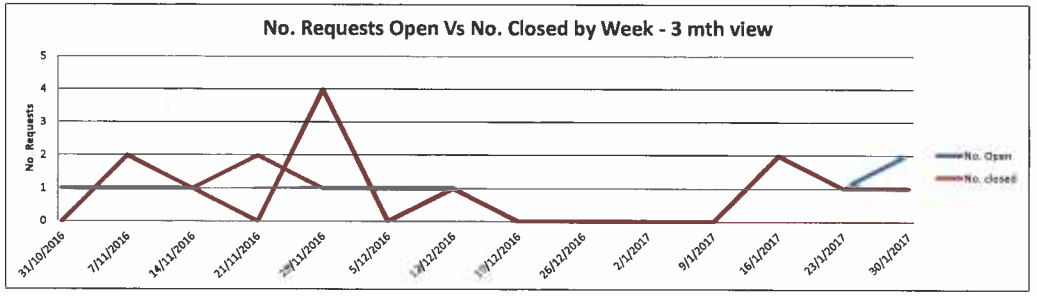






Trends



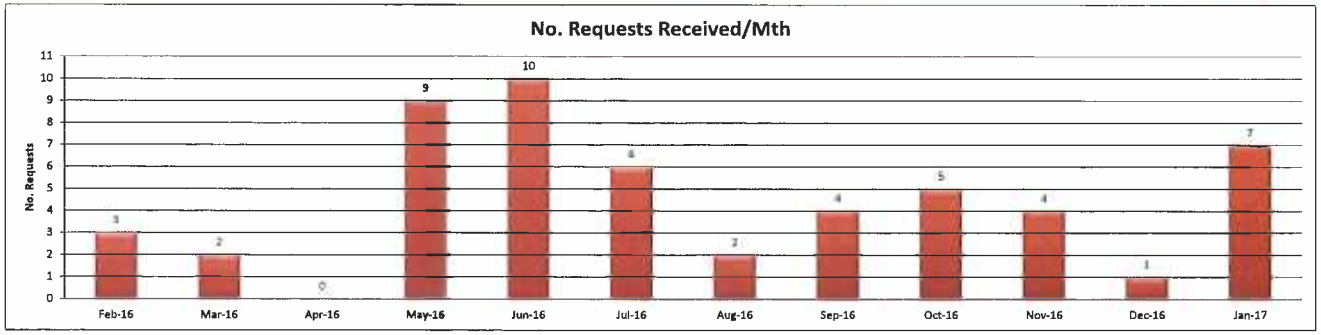


Department of Immigration Requests - Dashboard

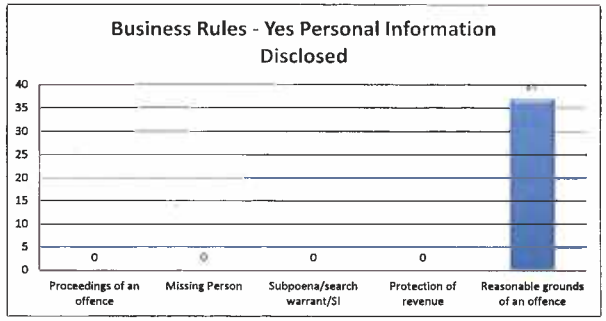
21 October 2014 - 31 January 2017

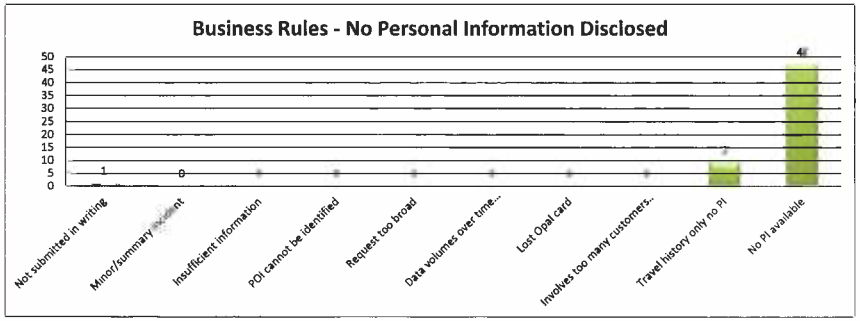
No. Requests Currently Open:	1	No. Requests Currently Complete:	94	Total Requests Received to Date:	95

No. Requests Information



Personal Information Disclosed/Not Disclosed





Trends