Using Big Data from TOTOR ETS to optimise public transport operations

Facilitating a privacy-protecting empirically-driven continuous-optimisation approach to sustainable public transport operations using Big Data recorded by Tap On Tap Off electronic ticketing systems

Mathew Hounsell
+61 2 9514 4961
mathew.hounsell@uts.edu.au
www.linkedin.com/in/mathew-hounsell

Institute for Sustainable Futures
University of Technology Sydney

Level 10, Building 10,
235 Jones Street Ultimo, NSW, Australia, 2007

PO Box 123, Ultimo, NSW, Australia, 2007

55,000 words.

Use your ETS records to understand your operations and make your customers happy

We can only understand what we can measure; we can only act on what we understand.

Certificate of original authorship

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Mathew Hounsell declare that this thesis, is submitted in fulfilment of the requirements for the award of Masters of Sustainable Futures (95583), in the Institute for Sustainable Futures at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature: Signature removed prior to publication.

Date: 15th September 2020

Acknowledgements

With thanks to:

Michelle Zeibots and Paul van den Bos for their supervision.

Dr John Stone and Matt Faber for reviewing the submitted thesis

Chris Riedy and Suzanne Cronan for their administration

Jason Prior for his advice; Damien Giurco for his support

Claudine Moutou for her help and support

Ben Simons and Darren Lee (UTS Data Arena) for their support and help

The UTS Library and the Australasian Legal Information Institute (AustLII)

Transport for NSW

Transport Performance and Analytics

Transport Research Hub

Open Data Team

Transdev

The Capstone Students

Joel Smalley, Igor Mileusnic, Nathan Murdoch, Opinderjit Samra, Veronica Thorpe

My Group Accountability and Support (GAS) group

Andrea Turner, Anh Nong, Katelyn Bywaters

With special thanks to

Gavin Gatenby and Michelle Zeibots for giving me the push to pursue transport

My mate Delilah Slack Smith

My parents Ken Hounsell & Diana Hounsell

and especially to Bharath Ragothaman

Abstracts

Facilitating a privacy-protecting empirically-driven continuous-optimisation approach to sustainable public transport operations using Big Data recorded by Tap-On Tap-Off electronic ticketing systems

Medium (200 words)

This thesis contributes to development of the trans-disciplinary field of Transport Analytics, which aims to facilitate a sustainable customer-centric approach to transport service delivery through continuously measuring and optimising transport operations and through better targeting of customer preferences and needs.

Sydney's Opal Electronic Ticketing System (ETS) records Tap-On and Tap-Off (TOTOR) pairs and represents a census of passenger responses to the services promised and delivered. Unfortunately, these datasets contain private biographical travel histories of individual customers. This thesis assists practitioners in actualising new analytical opportunities, by describing the methodological barrier that is the intrinsic human right to privacy, and then demonstrating a method to overcome that barrier.

The methods proposed transforms biographical-datasets into privacy-safe activity-datasets through deidentification, disassociation, aggregation and elimination. The method proposes three stages of transformation to create three levels of activity datasets with increasing privacy that can be distributed and shared respectively with service providers, collaborators, and then the general public.

This thesis provides case studies demonstrating how the high-resolution activity datasets enable novel analytic techniques to assist service partners in analysing and interpreting passenger response to transport operations. These case studies leverage inherent aspects of the data that were not available in previous data forms.

Long Abstract (400 words)

This thesis contributes to development of the trans-disciplinary field of Transport Analytics that aims to better target customer preferences and needs while optimising public transport operations. It demonstrates how use of empirical data acquired from Electronic Ticketing Systems (ETS) such as the Opal card system in Sydney, can provide more accurate and unexpected insights into demand patterns for public transport services. Systems that record every Tap-On and Tap-Off (TOTOR) pair, such as Opal, effectively provide a census of passenger responses to the services delivered, potentially increasing certainty and consensus on key aspects of operations such as required capacity, appropriate frequency and interchanging.

Previously, the lack of detailed empirical data led public transport service providers to rely on top-down system-level models of macro-behaviour. The availability of high resolution TOTOR datasets provides an opportunity to develop bottom-up human-level models of micro-behaviour that can then be used to construct more accurate macroscopic system models. As will be shown, this difference in approach can lead to significantly different conclusions about patronage and appropriate service levels.

The thesis approaches this new data and analytical opportunities in two ways. Firstly, by acknowledging and addressing concerns about privacy through development of a method to construct privacy-safe datasets; and secondly through new analytical methods to take advantage of the TOTOR data.

The travel histories of individual customers contained within TOTOR datasets provide detailed biographical information; and so, to protect the privacy of individuals, access to these datasets has been highly restricted. In response, this thesis describes the methodological barrier created by the need for privacy protection, proposing a method to overcome this. The thesis provides several example case studies undertaken using analytics developed for activity datasets to improve public transport operations. These leverage inherent aspects of the data that were not available in previous data forms.

The methods proposed leverages the ability to transform biographical-datasets into privacy-safe activity-datasets through deidentification, disassociation, aggregation and elimination. The method proposes three stages of transformation to create three levels of activity datasets with increasing privacy that can be distributed and shared

appropriately between service providers and coordination agencies, to collaborators (such as researchers), and then to the general public.

At all times, the research has been carried out within a customer-centric (customer service) approach to public transport service delivery. In the case studies, improved analytics has been shown to assist service partners in analysing and interpreting passenger behaviour in transport operations.

Structure of this thesis

This is a conventional thesis. There are no papers / publications included. There are no contributions from other authors.

Figure 1: NSW Opal smart-card ticket device on a table of NSW paper tickets

Opal is from NSW, Australia; Octopus is from Hong Kong; Oyster is from London, UK; MyKi (pronounced as My Key) is from Melbourne, Australia); and Gautrain Gold is from Johannesburg, South Africa. Picture from (Giles 2013b)

Figure 2: The Sydney Harbour Bridge – Built for the future

Table of Contents

Table of Tables

Table of Figures

Abbreviations

API	Application Programming Interface
APP	Australian Privacy Principles
ASGS	Australian Statistical Geography Standard
BCP	Best Current Practice
CBD	Central Business District
EIS	Environmental Impact Statement
ETS	Electronic Ticketing System
FY	Financial Year
GCLR	Gold Coast Light Rail
GTFS	General Transit Feed Specification
GHG	Greenhouse Gases[1]
GMA	Greater Metropolitan Area
HTS	Household Travel Survey
IPP	Information Protection Principles
IWLR	Inner West Light Rail
LCP	Least Cost Planning (framework)
LGA	Local Government Area
MAAS	Mobility as a Service
PKT	Passenger Kilometres Travelled
PKTpc	Passenger Kilometres Travelled per capita
PTPM	Public Transport Project Model (New South Wales)[2]
PVD	Passenger Volume Distribution
RFC	Request for Comment
RPI	Responsive Passenger Information (system)
SA1 SA2, SA3, SA4	Statistical Area level 1 etc[3]
SATS	Sydney Area Transportation Study[4]
STM	Strategic Travel Model (New South Wales)
TAP	Transport Access Program (New South Wales)
TFN	Tax File Number (Commonwealth of Australia)
TOR	Tap-On Recording
TOTO	Tap-On Tap-Off
TOTOR	Tap-On Tap-Off Recording
TWT	Tuesday, Wednesday, Thursday[5]
UML	Universal Modelling Language
VKT	Vehicle Kilometres Travelled
VKTpc	Vehicle Kilometres Travelled per capita

[1] GHG are climate changing pollutants

[2] The PTPM is a recent addition to the STM in NSW modelling.

[3] Statistical Areas are from the ABS Australian Statistical Geography Standard

[4] (NSW 1974a, 1974b, 1974c, 1974d)

[5] Tuesday, Wednesday, and Thursday are used to calculate the average for normal workdays

Part I

Thesis

CHAPTER 1 – INTRODUCTION

The advent of Big Data has changed many aspects of our day to day lives and our relationship with technology. Its impact on practitioner perceptions of public transport has been significant and the resulting changes will be profound. This is because Big Data sets arising from the new generation of Electronic Ticketing Systems (ETS) like the Tap-On, Tap-Off recorder (TOTOR) system in Sydney (known as Opal), have the ability to enable high-resolution measures of passenger travel patterns, thereby providing insights into customer travel behaviour and their responses to public transport services. These measures have the power to answer questions that were previously unanswered or contested.

Before Big Data sets like Opal, public transport service providers had comparatively little empirical knowledge of how many customers were using public transport, when they were using it, where they were going, or what their origin and destination pairs might be. Big Data arising from TOTOR systems have now changed that, and it is as if for the first time, customers are able to tell us what they are actually doing with the services delivered to them. In the process, many long-standing assumptions that transport professionals have made about customer travel preferences are beginning to shift as results from empirical data prove many long-held assumptions as incorrect.

This thesis is the first stage in an investigation of what is now a rapidly changing field. It focusses on three things:

- An overview of what Big Data can tell us about public transport customer preferences and how they differ to previous views and beliefs arising from transport models
- How to protect the privacy of individual citizens whose travel patterns and preferences are revealed in Big Data, and why privacy protection must be assured before accessing data and the new insights it potentially provides
- A preliminary outline of the research questions that arise as a result of this first stage of research, which will form the basis of a later doctoral research program.

Results from the early stages of my research into the application of Big Data to public transport service analysis have provided significant insights. I have been able to demonstrate the ability to measure both the performance of transit services and customer responses to those services in ways that reduce guesswork and ambiguity. The level of detail in these datasets has also created new challenges. The ability to monitor the movements of private individuals raises the spectre of systemic or incidental privacy infringement, and there is a clear need for systemic and thorough privacy protections.

This thesis describes how to overcome the methodological hurdle of privacy protection that impairs the use of Big Data in facilitating the move from decisions that are model-driven to empirically-driven operations. The rest of this section outlines the structure of the thesis:

Chapter 1 – Introduction – begins with a discussion on the advantages of Big Data and what it can tell us about the revealed preferences of public transport customers, and how those preferences differ to previous views and beliefs arising from stated preference surveys and transport models.

Next, the research is described in the context of the questions that arose from this research programme. These questions are described as the three stages of organisational transformation (from model-centric product-centric to empirically-driven customer-centric). These questions will form the basis of a later doctoral research programme.

Then the chapter discusses leveraging Big Data to measure demand, service delivery, and customer response, as well as how Big Data allows the expansion of measurement beyond the peak hours to 24 hours a day 365 days a year at no extra cost.

This chapter finishes by describing theoretical frameworks for considering the knowledge gaps within transport operators, including how operators may miss important industry knowledge like the Service Quality Loop and the *Seven desires for useful transit, and how transit serves them*.

Chapter 2 – Privacy – describes why privacy protection must be built into the analytical processes *before* accessing Big Data. The chapter begins by outlining the ethical behaviour requirements of a professional engineer. The chapter then describes the ISO 31000 Risk Management Guidelines, and how a continuous cycle of risk management can be used to ensure privacy is protected.

The chapter then describes the reasons for why privacy must be protected, starting with the direct hazards to safety of the individual in the context of NSW crime statistics, then the systemic hazards of social norms and social control, and how these are significant risks for minority groups within society. Finally the chapter outlines how privacy is an essential need for liberalism and democracy and where the use of Big Data could be perceived or become a threat of a unceasing surveillance.

Chapter 3 – Privacy Protection – describes how to protect the privacy of individual persons whose travel patterns and sensitive information can be revealed in Big Data. The chapter describes Information Protection Principles, which are the Australian institutional and legal framework for privacy protection. This is discussed within the context of the pivotal legal case of Waters v Transport for NSW [2018] NSWCATAD 40 and its significance for transport operators.

Within this context the concepts of identifiers, biographies, activity datasets and privacy budgets are described. The chapter outlines why the Need-to-Know risk management framework is a key approach for information security and privacy protection. Then the chapter describes how derived datasets use the four methods of Deidentification, Disassociation, Aggregation, and/or Elimination to remove biographic information, in order to ensure the composition of datasets with the right level of detail for each users' Need-to-Know.

Chapter 4 – Method – describes the outcome of the research, which was to identify a method to prepare derived datasets to allow measurement of transport operations and the subsequent customer response. Careful preparation of the data ensures biographical information is protected in-line with the Need-To-Know risk management framework.

The chapter also describes the use of contextual datasets (for example holidays and weather conditions) and organisational datasets (for example stop locations and service routes) needed for rigorous analysis. Then the chapter outlines the data preparation needed to create privacy-protecting primary and secondary derived activity datasets, followed by privacy-safe tertiary datasets, which can be used to examine public transport operations and customer responses.

Chapter 5 – Case Studies – demonstrates the new insights that Big Data can provide through several case studies undertaken as part of this research programme. These case studies use primary-, secondary-, or tertiary- datasets which can be produced using the method proposed in Chapter 4. The first set consider examples of where measured passenger volumes of public transport use have exceeded model forecasts. The next set show that higher vehicle frequency (shorter vehicle headways) correlate with higher passenger volumes. Evidence for diverse travel patterns is then demonstrated using stop-pairs from TOTOR datasets. Finally, the chapter describes the ability to analyse actual customer responses using passenger volumes across screenlines computed from primary datasets.

The chapter concludes with an overview of how such empirical data might improve decision-making on public transport service levels by bringing it more into alignment with what customers value because it is based on knowledge of actual customer responses rather than assumptions.

1.1 Key advantages of Big Data

In 2015, Transport for New South Wales (TfNSW) relied on the five-yearly Commonwealth of Australia (CoA) census carried out by the Australian Bureau of Statistics (ABS), a Household Travel Survey (HTS) with a 3,000-person sample carried out by TfNSW predecessor organisations, and data from ticket sales to drive the department's four-stage Sydney Strategic Travel Model (STM) to estimate passenger behaviour and patronage levels (TfNSW 2001a, 2001b, 2018c).

In 2016, the first Opal datasets became available, just as my research began into measurement methods to assist transport operators in optimising their transport service offerings. My aim was to determine how transport bureaucracies could change from an approach based on small-scale surveys of passenger travel behaviour driving limited models, towards an empirical approach based on continuous large-scale measurements from ETS such as Sydney's Opal system.

During my investigations I found that the approach of any transport organisation will lie somewhere on a spectrum between model driven and empirically driven information sources, see Figure 3 below. An organisation described as empirically driven would seek to make its decisions based on measurements of the real world. By contrast, a model driven organisation would base its decisions on outputs from a theoretical model. Of course, most organisations will use a mix of models and empiricism, and that mix will vary from project to project and team to team depending on data availability and practitioner skill sets, which is why the approach is conceptualised as a spectrum. The approach of an organisation, when assessed as a whole, may move back-and-forth between the two poles over the lifetime of the organisation.

Figure 3: The spectrum of organisational approaches to decision making

Empiricism and modelling are two complementary analysis techniques. Modelling and the assumptions embedded in its structure enable estimations of travel behaviour, such as the passenger changes from changes in service levels. Modelling will always be an integral part of transport planning and operations. However, the accuracy and consequent usefulness of modelling is significantly improved when it is informed by empirical knowledge which leverages Big Data. The industry should seize the opportunity to improve the performance of modelling because that is an opportunity to improve customer service. During the course of this research, institutional resistance to changes of this kind was experienced and confronted. During the research, ways to overcome potential resistance to the changes needed for improvements were developed.

Through necessity, models are a simplified simulation of reality based on assumptions. Consequently, models do not include corner-cases and unusual behaviours. Models can never completely reflect the nuanced picture of reality that can be derived from empirical measurements. In addition, a model relying on small datasets will have limited accuracy because it only samples a small number of behaviours.

The NSW HTS is designed to undertake a deep analysis of individual travel behaviour across all modes and includes questions on trip purpose. However, the sample size is less than 1 in 1,000 residents in the New South Wales (NSW) Greater Metropolitan Area (GMA) of Newcastle, Wollongong, and Sydney. As such, the HTS can only ever be indicative, as it will never be able to detect the full range of behaviour of over two-million daily public transport users (the normal case before COVID-19).

The five-yearly Australian census has a large population sample size and can be considered comprehensive, however, it too has a small sample size because it applies only to trips for one purpose on one day every five years and only asks only one question: how did you travel to work today?

The lack of resolution in the temporal and purpose dimensions limits the utility of the census as a data source for analysis of transport usage. For example, the August census is used to assess cycling in Victoria. However, as Melbourne researchers are keen to tell you, August is the coldest and wettest month for the city. If funding and planning models assume mode share for cycling remains constant all year round, any models relying on the census will underestimate cycling during more clement weather.

Results from this research will show that transport models often provide results that are significantly different to empirical results. Consequently, provider expectations can differ from user experience. This difference reduces the ability of model-driven service providers to make decisions that are in the best interests of users and undermines their ability to improve and enhance more sustainable public transport modes.

Many providers have deployed smartcard based ETS that record some details on trips made with public transport. There has been substantial research into using richer datasets from ETS records to inform transport models. For example, NSW Transport Performance and Analytics (TPA) developed the Public Transport Project Model (PTPM) from the STM using inputs from the ETS to analyse the Sydney Metro (TfNSW 2018c, p. 112). There is an opportunity to use analysis of ETS records to derive more empirical measurements as shown in Chapter 5 - Case Studies.

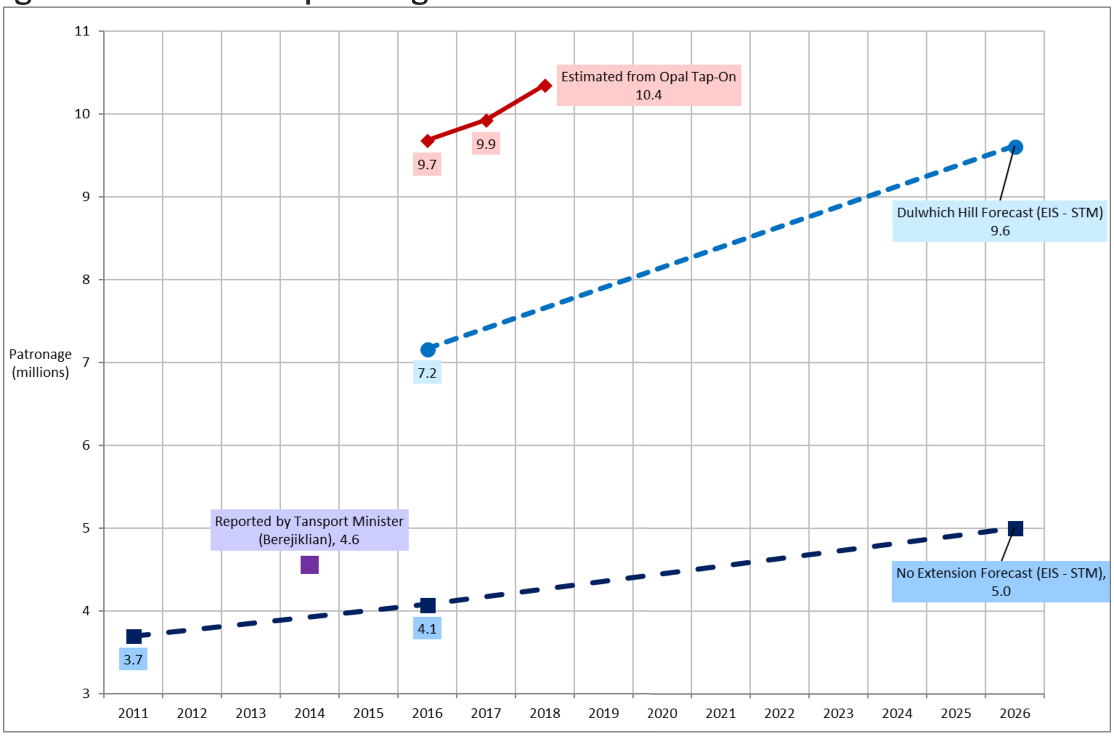
Most ETS deployed in the world are Tap-On only smartcard systems, meaning the user's entry point to the system is known but the exit point (Tap-Off) is unknown (Hounsell 2017c). Tap-On-Tap-Off Recording (TOTOR) ETS, like Opal in NSW and Queensland Go-Card, are less common and provide rare insights into customer travel behaviour. Since Tap-On Recording (TOR) systems are common, many research papers have examined methods to estimate probable Tap-Off locations (for example (Munizaga & Palma 2012)). However, TOTOR systems record the time and location for Tap-On and Tap-Off, thus these datasets are intrinsically richer than their TOR counterparts.

Some ETS are truly multi-modal and some have monopoly coverage over large conurbations, such as Opal's monopoly over the NSW GMA. A recording ETS with a monopoly over a large conurbation can be considered to compile a census on public transport usage within its coverage area. As such, monopoly ETS generate census datasets that have the power to support analysis of actual passenger behaviours from multiple viewpoints that can facilitate empirically-driven decision-making.

***Please Note**: This thesis discusses transport projects and models from NSW. This thesis considers the forecasts of certain models in retrospect. The discussion of these particular projects is designed to illuminate models in general. This thesis is not a criticism of any person or any model in particular, nor any particular engineering decision.*

Results from the STM provided the basis for decisions within the NSW Department of Transport (NSW DoT) relating to operational strategies and improvements for the Inner West Light Rail (IWLR) such as service frequency and capacity (GHD 2010). However, as Figure 4 below shows, the modelled results significantly underestimated the number of passengers on the IWLR after the opening of the extension.

Figure 4: Annual IWLR patronage observed and forecast for different scenarios.



Sources: (Hounsell 2018, p. 5 - revised with FY 2018-19), (GHD 2010)and (TfNSW 2019a, 2019c)

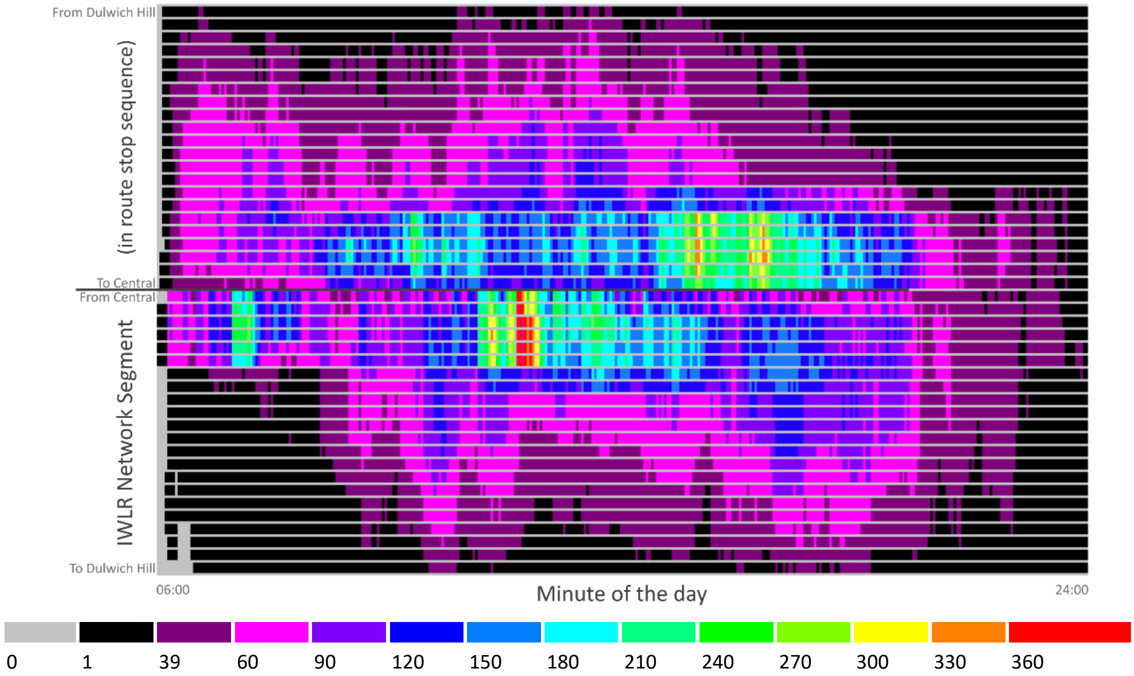
The assumptions coded into the STM during its design and implementation prevented the model from accurately estimating the patronage on the changed public transport network. The performance of a model should be regularly assessed against empirical measurements, with assessments examining the outputs for systemic biases. This is discussed in more detail in Chapter 5 – Case Studies.

Broadly, there were three observed responses to these statements of fact regarding the models results: disinterest, interest, and anger. Humans can respond to 'perceived' challenges of their deeply held beliefs with emotion (Wood & Porter 2019). The above statements are not intended as a challenge to the orthodoxy, they are merely an identification of an opportunity to improve transport modelling in NSW.

1.1.1 Action research programme

In May 2016, a demonstration TOTOR analysis of the IWLR was prepared for delivery during the launch of the UTS Transport Research Centre (TRC). The analysis was presented to the assembled service partners from TfNSW, the Ministry, and various public transport operators including Transdev, the operator of the service. The presentation included a visualisation of spatial-temporal Passenger Volume Distribution (PVD) for the IWLR, see Figure 5 below and see Section 5.7 for more explanation.

Figure 5: Saturday load on the IWLR 06:00 to 24:00 showing event crowds



A colour change represents a 30 person increase.

These results showed empirical evidence of crowding spikes (indicated in red and orange) during the weekends at The Star. This analysis shed new light on long running differences in perception over the existence of crowding between TfNSW and the operator (Transdev). Previously, TfNSW had relied on estimates from the STM and not empirical measurements to determine patronage and crowding levels. As shown in Figure 4 and Figure 5, empirical measurements demonstrated actual patronage far exceeded the estimated patronage from the STM, and customers would have been experiencing ‘crush loading’ as patronage reached and exceeded vehicle standing capacities (TfNSW 2001a, 2001b, 2018c).

My research triggered a programme to optimise the IWLR: ‘The research programme was begun using the principles of least cost planning, our objective was to determine ways to improve service quality with the least investment in capital and operating expenditure. Ideally the programme would find ways to improve services and also reduce costs. Initial investigations determined that there may be many ways to improve the operations of the IWLR.’ (Hounsell 2018a, p. 1)

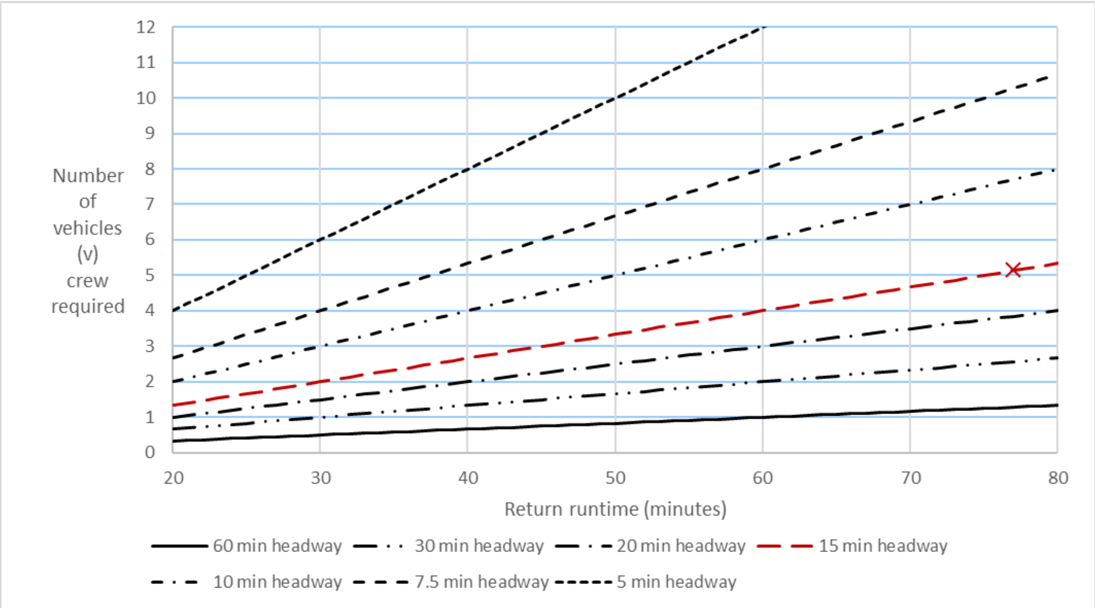
Next, the impact of return-runtime on the number of vehicles required to operate the IWLR at different headways was modelled. The return runtime is the time taken for a tram to run from Dulwich Hill to Central, return, and be ready to run the next service.

Return runtime is the sum of dwell times at stops, travel times between stops (including waits at traffic lights) and turnaround time (Vuchic 2005). In the process:

> 'It was determined that saving four minutes from the end-to-end runtime on the Inner West Light Rail could reduce the size of the deployed fleet. Those saving could then be used to improve the farebox-recovery-ratio or reinvested as lower vehicle-headways. The key approaches identified were operating speeds, dwell times, intersection delays and targeted service delivery.' (Hounsell 2018a, p. 1)

The initial results provided a focus for the UTS TRC Network Planning & Optimisation research stream — empirical measurement of patronage and service delivery to achieve service optimisation through controlling and reducing service runtimes.

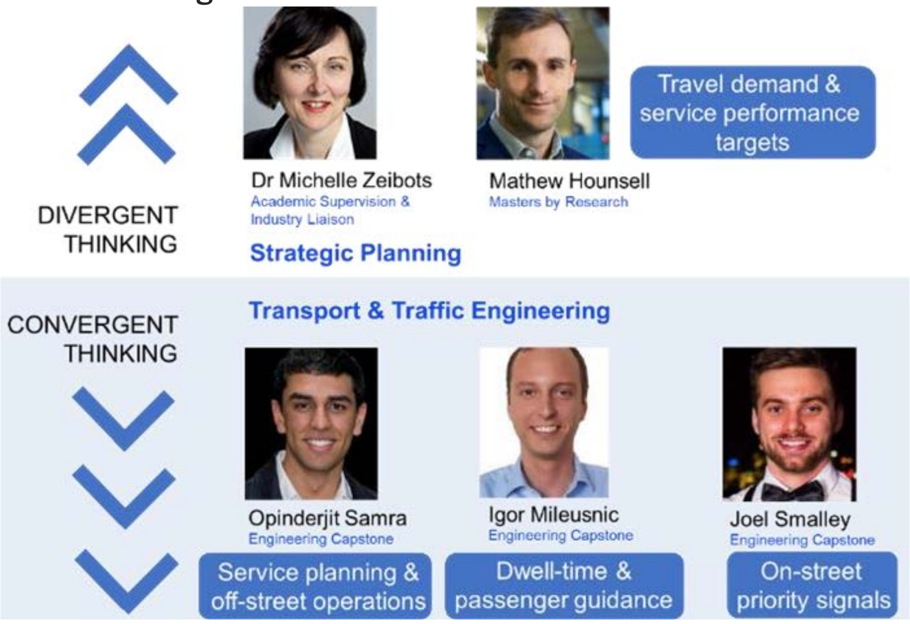
Figure 6: Number of vehicles per runtime per headway – two dimensional



Source: (Hounsell (2018), p 11, Figure 10)

The research programme was not just theoretical. After the results of the initial research, a cluster of engineering capstone projects were undertaken to investigate specific actions for reducing runtime on the IWLR, as shown in Figure 7 below. In collaboration with the students, I created transport analytics methods using the Opal dataset to assist in their investigations. Samra (2017) and Smalley (2017) in particular used a variation on the PVD methods developed above. I presented our findings with the team to TfNSW, outlining several approaches to improve runtimes on the IWLR. This programme of identifying mechanisms for runtime reductions was continued with several additional students in 2018–2019.

Figure 7: Inner West Light Rail research team



Source: (M. Zeibots (2018) via Hounsell 2018a, p. 2)

While reviewing this thesis one examiner commented that 'The proposition that faster trams require less vehicles to operate a set [L]evel [O]f [S]ervice is self-evident and would not be contested by any operator'. Similar comments were made while reviewing (Hounsell 2018a, p. 1). However, the on-the-ground experience from this action research programme suggests this assumption is incorrect in so far as many practitioners, especially those working in road traffic signals divisions, are not aware of that proposition. TfNSW is a large department that draws staff with expertise from multiple disciplines, such as project management and contract law. Key personnel in TfNSW were not familiar with the fundamentals of tram operations. Those in management and support roles benefited from UTS distilling the proposition into a simple optimisation target – saving four minutes return-runtime enables saving one tram and operating crew.

> 'The industry partners were pleased with the results of our research programme and found the workshop to communicate the results extremely useful. They were also heartened that we took extra care to explain to the lay-persons that reducing end-to-end-runtimes could reduce the number of deployed vehicles and lead to cost savings. They were also impressed by the common theme demonstrating that eliminating runtime variability caused by signalised intersections etc would also ensure those savings could be realised. The department was also quite happy to see that the research focused on delivering low-cost high impact changes. ' (Hounsell 2018a, p. 1)

The findings on runtime savings were developed into a significant component of subject 48370 Road & Transportation Engineering that is delivered to engineering students at UTS through the School of Civil and Environmental Engineering.

The empirical findings had a positive impact on transport in NSW through improving relations between the IWLR operator and TfNSW. The empirically-driven approach reduced disagreement between service partners and induced greater operational collaboration. The service partners responded to empirical knowledge of passenger crowding by collaborating to allocate additional trams to inter-peak services.

1.1.2 Opportunities for continuing the research

During this research programme it became clear the scope could be broadened into the philosophical and theoretical space, examining and expanding the theories that assist the delivery of customer-centric transport operations. The need, or gap, in our knowledge which this fills is anticipated by Levitt:

> ‘The railroads did not stop growing because the need for passenger and freight transportation declined. That grew. The railroads are in trouble today not because the need was filled by others (cars, trucks, airplanes, even telephones), but because it was not filled by the railroads themselves. They let others take customers away from them because they assumed themselves to be in the railroad business rather than in the transportation business. The reason they defined their industry wrong was because they were railroad-oriented instead of transportation-oriented; they were product-oriented instead of customer-oriented.’ (Levitt 1984)

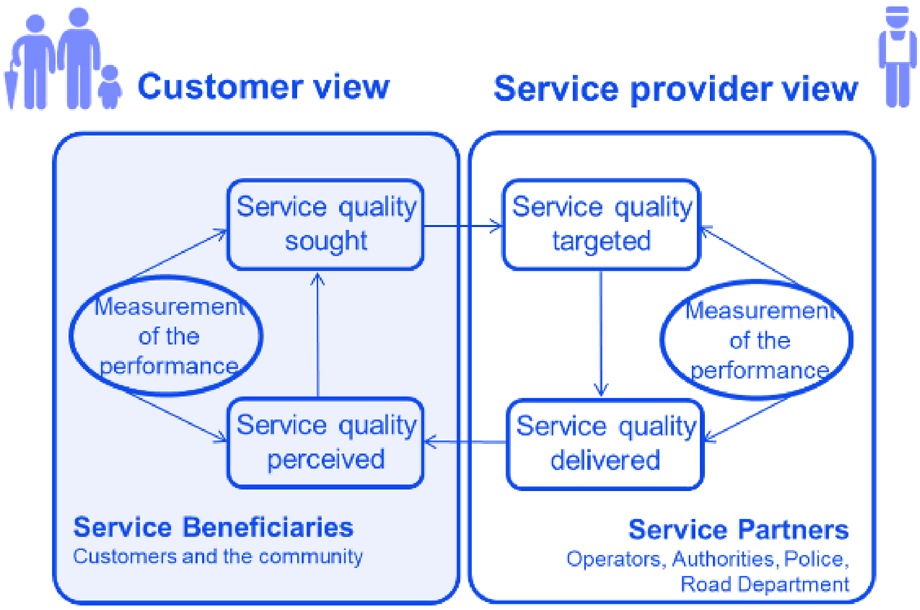
Traditionally, passenger railway focused on delivering a timetable, they were a scheduled-service that operates at a fixed-speed[6]. Since passengers do not have control over the vehicle they require a schedule to plan and coordinate their usage of the trains, buses and other scheduled services in order to successfully access mass transit (Zeibots 2007). This focus on delivering a timetable has been fundamental to the success of passenger services for well over a century. In this way, it might be said that railways were focused on delivering their product — the timetable — and as such, could be considered *product-centric*.

Recently there has been a realisation that Levitt was correct and transport operators needed to focus more on their customer’s needs – to become *customer-centric*.

[6] Public transport is operated by corporations (public/private) with scheduled services, and those achieve fixed speeds. Private transport which is operated by individuals to a random schedule and achieves variable speeds.

The Comité Européen de Normalisation (European Committee for Standardization) (CEN) defined the Service Quality Loop as a standard for assessing public transport operations (CEN 2002; Hounsell 2017c, 2018d). The Service Quality Loop aims to be a customer-centric conceptual framework based around the control-system concept of continuous measurements directing a feedback loop as shown in Figure 8. The Service Quality Loop is useful in explaining the importance of continuous optimisation and customer-centric decision-making to the creation of successful public transport services.

Figure 8: The ideal Service Quality Loop (EN 13816) is Customer Centric

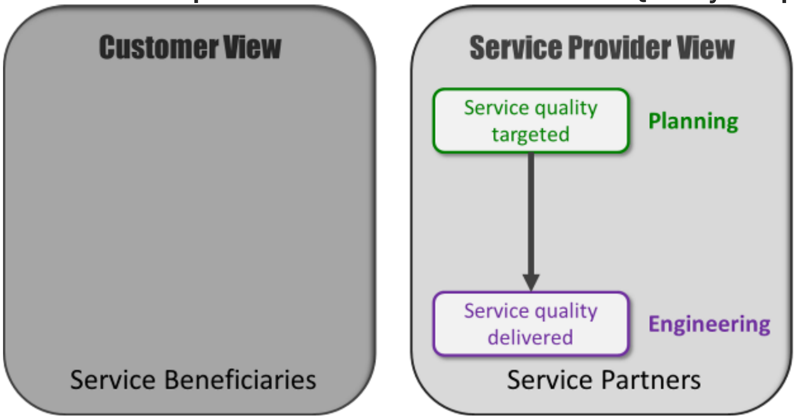


To explain the terminology, consider this hypothetical example, the government instructs the bureaucracy to engage a private-sector railway operator to provide a railway service on a newly constructed railway. The government, bureaucracy and railway operator are all service partners collaborating to provide services. In this example, when specifying the new operation, the service partners agree a fixed price, with the operator targeting a service frequency of one train every ten minutes. To ensure value for money, the bureaucracy will measure the number of minutes between trains to ensure the service delivered meets the services targeted in the contract. This example agreement is considered product-centric because there is minimal consideration of the customer's perspective when developing the services targets —Figure 9 below.

In theory, service partners should first engage the service beneficiaries to determine the services they seek then develop targets to serve those desires. The beneficiaries will develop perceptions based on both the targets and their experience of services delivered.

To be customer centric, the partners must regularly engage the beneficiaries to determine if they perceive the services delivered as meeting their expectations.

Figure 9: Product-Centric operations shown on the Service Quality Loop



Work with TfNSW indicates the Service Quality Loop is a useful conceptual framework. However, there are gaps in the literature regarding the implementation of customer-centric, continuous optimisation approaches like that advocated by the Service Quality Loop in public transport operations.

1.2 The 3 stages of transforming an organisation

1.2.1 Customer at the centre

Generally, it is easier for an organisation to be product-centric than to achieve customer-centric operations. Being customer-centric requires constant work, engaging with customers to assess their actual needs so that the organisation can target services to be used in satisfying those needs (Bartels 1988; Constantinides 2006).

Recent research indicates that to thrive in today's marketplace, transport operators need to understand that customers only seek transportation to fulfil intrinsic needs. The recent literature, including that on Mobility-as-a-Service, demonstrates that customers are more demanding, less patient and will not seek transport if the internet can satisfy their intrinsic needs (Crayen 2018; Li & Voege 2017; Rosenfeld, Thomas & Hausen 2019).

In 2011, the NSW government changed, and Gladys Berejiklian became the Minister for Transport. A significant policy initiative was reorganising the State's roads, railway, and maritime sectors by consolidating the departments, agencies and operators into a single cluster overseen by a single coordination authority with a focus on customer service. The intention of the initiative is laid out in an exert from this media release:

> ‘A new era in transport planning and customer service starts today with the formal establishment of the new integrated transport agency - Transport for NSW. … NSW Minister for Transport Gladys Berejiklian and Minister for Roads and Ports Duncan Gay said the occasion was a turning point for transport delivery in NSW.
>
> “Many months of hard work have gone into this process and now, for the first time, NSW has an integrated transport authority,” Ms Berejiklian[7] said. … “All planning and policy functions have been consolidated into the one authority, signalling the end of silos and the removal of duplication. The customer will be at the centre of everything we do.”’ (TfNSW 2011, p. 1)[8]

This policy influenced the planning process including the NSW Long Term Transport Master Plan (LTTMP) in 2012:

> ‘The LTTMP ‘sets the framework for the NSW Government to deliver an integrated, modern transport system that puts the customer first. … It sets the strategy and the direction required to deliver the customer-focused, integrated transport system that NSW needs.’ (TfNSW 2012, p. 9)

The Future Transport Strategy was a 2018 update of the 2012 LTTMP, it states:

> ‘Just like the 2012 [LTTMP], Future Transport 2056 places the customer at the centre of everything we do. It outlines a vision, strategic directions and customer outcomes, with infrastructure and services plans underpinning the delivery of these directions across the state.’ (TfNSW 2018b, p. 4)

In TfNSW (2018a), the agency describes its culture and values as follows:

> ‘Customer Focus – We place the customer at the centre of everything we do.’

Placing the ‘customer at the centre’ is a reference to the marketing and business concept of customer-centricity (Gummesson 2008; Shah et al. 2006). This change in focus raises several important questions:

- What does being customer-centric mean in practice?
- What impact will customer-centricity have on the behaviours of groups such as public transport passengers, transport services operators and their partners in government oversight?
- What benefits will customer-centricity deliver to passengers, businesses, government and society?
- Is customer-centricity a better approach than product-centricity, and why is one better than the other?

[7] Gladys Berejiklian: Liberal, MLA Willoughby

[8] In 2019, the now Premier Berejiklian, reissued her directive to merge the departments because the RMS had re-established itself as an independent road-centric entity (Saulwick 2019).

There is a large body of work addressing the concept of customer-centricity (Grönroos 2006; Gummesson 2008; Rihova et al. 2015; Vargo, Maglio & Akaka 2008). Some works specifically considers the perspective of passengers, businesses, and the society that are deriving value from the delivery of transport services.

Through becoming customer-centric, the transport operations may increase the overall benefit they deliver to society when compared to their costs of operations. Through optimisation of transport operations, it should be possible to minimise waste and reduce externalities, such as pollution, further improving the ratio of benefits to costs.

The customers of a transport service are all entities that *actualise value co-creation* from the direct and indirect use of the transport services. A transport operator that does not share the same values as its customers will develop and deliver services that the customer considers to be undesirable or of a low quality. By achieving values-alignment with their customers, a transport operator can deliver an efficient implementation of highly desirable (high-quality) public transport services (Jittrapirom et al. 2017; Levitt 1984; Lusch & Vargo 2014). As such, a transport operator can attract more passengers to sustainable public transport modes with a customer-centric approach.

There is a large body of work in business analysis – especially Lean Services and Six Sigma (Rooney & Rooney 2005; Smith 2014; Suárez-Barraza, Smith & Dahlgaard-Park 2012) – and sustainable systems – Least Cost Planning (Fercoq, Lamouri & Carbone 2016)– that clearly identify how an efficient and sustainable business is more customer-centric because it strives to deliver only the services or outcomes that users and customers need in ever more cost effective ways and that produce less pollution.

In 2016, service partners provided services for 1.2 million train trips and 0.9 million bus trips on an average workday in NSW. Organisational transformation does not occur overnight, and the NSW transport cluster continues to strive for customer-centricity. This research can assist in that transformation by demonstrating a method to overcome the methodological barriers of privacy protection preventing the use of a travel history dataset to in empirical transport analytics.

1.2.2 Product-Centric Model- to Customer-Centric Empirically-Driven

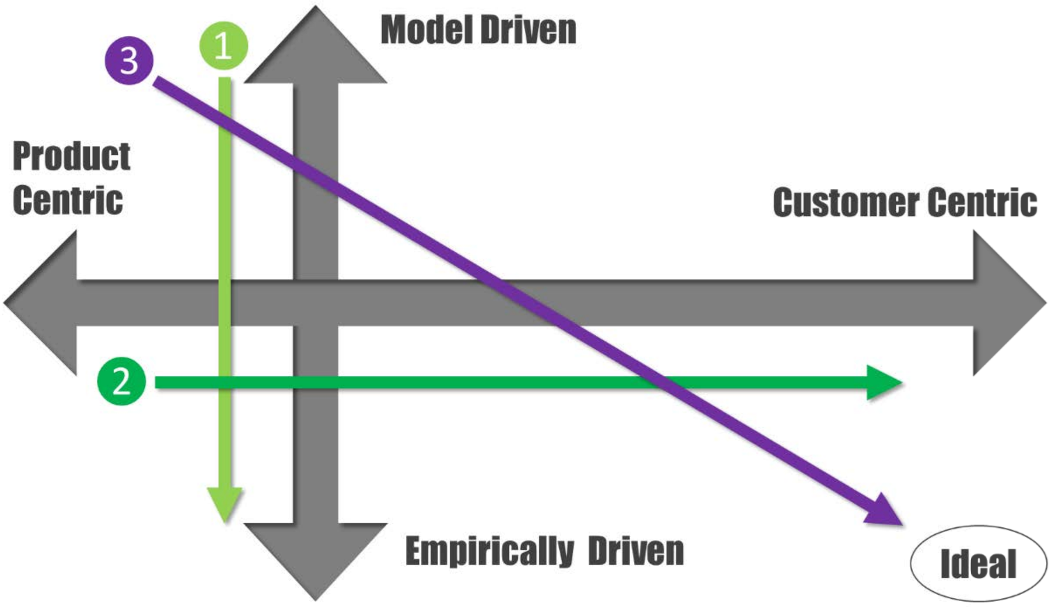
Put simply, a product-centric business develops products and then uses marketing to create customers whereas a customer-centric business uses marketing to analyse and develop services that customers will need (Gummesson 2007, 2008; Shah et al. 2006).

There are three questions that need to be addressed to achieve the professional and organisational transformation towards more sustainable customer-centric outcomes in the transport sector (for both roads and transit):

❶ How to move from model driven to empirically-driven analysis methods?
❷ How to move from product-centric to customer-centric mindsets?
❸ How to combine these and overcome institutional inertia to shift from the model driven product-centric transport operations of today to the empirically-driven customer-centric transport operations required for a sustainable future?

This research examines and documents a method to facilitate the shift from model to empirical analysis driven methods of investigation – ①. There are opportunities for further research into customer-centric approaches in transport – ②, as well as into the combination of these two transitions towards the ideal – ③. This third transition could result in the emergence of a significantly different approach to service analysis and delivery. The relationship between these is summarised in Figure 10 below.

Figure 10: The three stages of organisational transformation



There are additional sub-questions that arise from further analysis of this model:

❶ Model driven to empirically-driven:

i) How is the transport network actually being delivered?

ii) How are passengers responding to the delivered transport network?

❷ Product-centric to customer-centric:

i) Who are the customers?

ii) What do customers need, want & value?

iii) How would customer-centric public transport behave?

❸ Model driven product-centric to empirically-driven customer-centric:

i) How do we trigger change in transport operators and their partners?

This thesis examines how to overcome the methodological hurdle of privacy protection that impairs the use of TOTOR datasets in facilitating the move from model-driven to empirically-driven operations – ①. Analyse of why transition ② or ③ are desirable, how they might be achieved, and the background research needed to frame new service driven approaches to transport, are topics more suited to doctoral research.

1.2.3 Problem statement for this thesis

In 2011, the Minister for Transport declared the transport cluster would become customer-centric. The department outlined the need for a frequent inter-connected transport network in the LTTMP and Sydney's Rail Future (TfNSW 2012, 2013).

Starting in 2013, NSW began rolling out a smartcard based ETS called Opal. The Opal system is a TOTOR ETS which generates high resolution datasets that can be used to identify problems and target improvements. For example, by detecting common issues in Sydney like overcrowding on off-peak trams.

In 2016, this research into advanced analytics techniques began with a demonstration analysis delivered to the transportation industry and bureaucracy. That analysis empirically demonstrated inter-peak congestion on the IWLR. At the time, the existence of inter-peak congestion was hotly disputed by TfNSW because the models showed insufficient patronage to cause congestion. However, the models did not consider the 2,000 seat Lyric Theatre located next to one of the light rail stops, and thus did not include daytime event crowds in their congestion estimates. The models were also estimating the patronage level for 2016 at just 74.2% of the actual annual patronage.

There are two questions to address in the transition from a model-driven organisation to an empirically driven organisation – i.e. ① – i) how the transport network is actually delivered and ii) how passengers are responding to the delivered transport network.

In order to approach all decision making using the customers perspective – i.e. ② – the organisation must answer the central questions i) who are the customers? ii) what do customers need, want and value? and iii) what is customer-centric public transport?

While working with TfNSW it became clear that there is a systemic issue that needed resolution before the department and their partners could continue their transition to a more empirically driven approach. The travel histories contained within TOTOR datasets are clearly biographical information and using those datasets directly risks violating the privacy of the individuals described. As such, access to Opal datasets had been restricted. Therefore, privacy protection has become a methodological barrier that prevents a more empirical approach.

The identification of this methodological barrier - Chapter 2 and Chapter 3 - and a method to overcome - Chapter 4 - it are a significant contribution to the transport industry. Privacy is increasingly becoming an important issue for the public. In 2018, privacy became a chronic problem for TfNSW after they lost a legal case regarding privacy protection (NSWCATAD 2018).

The method proposed leverages the ability to transform biographical-datasets into activity datasets, through deidentification, disassociation, aggregation and elimination. The method proposes three stages of transformation to create four levels of datasets with increasing privacy that can be distributed as appropriate within the service-partners, to collaborators (such as researchers), and then to the general public. In the case studies (Chapter 5), improved analytics has been proven to assist service-partners in analysing and interpreting passenger behaviour in transport operations.

The TfNSW approach to releasing data as required under GIPA and Open Data policies has to date been inconsistent. Until 2014, the Department published monthly patronage results for every railway station in the network, such as Lithgow, Katoomba and Blaxland. However, that ceased in 2014 during the roll-out of the Opal system. TfNSW now only publishes the monthly patronage totals for “railway lines”, such as the Blue Mountains Line (which combines all three listed stations and many others). In contrast, TfNSW publishes monthly patronage for tram stops and ferry wharfs. However, other jurisdictions like Queensland are happy to regularly releases higher-resolution datasets on passenger responses to the transport network including monthly mobility summaries.

TfNSW genuinely wishes to assist service partners with open data but they are struggling with how to use the Opal datasets while protecting privacy and reputation. As such, this thesis tries to square the circle and provide a way forward that allows the department to protect privacy and embrace an empirically driven future.

1.3 Leveraging Big Data to close the knowledge gap

Computation for computations sake is a fun exercise for mathematicians and computer scientists, but without a purpose the results are little more than the dance of electrons.

At the start of the 21st century, humans are amidst the 4th Industrial Revolution due to the increasing deployment of Information Systems (Schwab 2017). This transformation has led to the development of the new fields of data science and analytics. Effective analytics requires engaging domain expertise to define the industry's problems, otherwise the complexities of real-world systems, such as public transport networks, will be missed, resulting in an unproductive naïve analysis.

As part of this industrial revolution, service partners have replaced paper tickets with smartcard based ETS, like Opal. Technological development has allowed the recording of data in previously unthinkable volumes, such as every Tap-On and Tap-Off pair in NSW. Since the fare compliance rate is over 90%, the Opal dataset represents a census of the passengers' response to the transport services targeted and delivered.

TOTOR datasets provide a valuable resource for assessing public transport network performance and consequent passenger behaviour, especially for off-peak and minor services. The size and accuracy of these datasets enable more comprehensive and accurate measurements than previous methods. Subsequent research will develop more complete and accurate models of passenger behaviour and network performance.

Previously, service-partners relied upon top-down system-level models of macro-behaviour – such as mode-split per purpose for inter-regional mobility patterns. The availability of high resolution datasets provides an opportunity for service-partners to develop bottom-up human-level models of micro-behaviour, which are then used to construct more accurate system-level models of macro-behaviour.

Operators have the opportunity to expand their knowledge using TOTOR datasets:

1) Testing the operators and industry's assumptions to detect any ideas that are actually incorrect (Misknowns; see section 1.4)
2) Assessing benefits of adopting industry best-practices (Blind-Spots)
3) Detecting new truths about customers and operations (Unknowns).

This section uses examples to show how TOTOR datasets are a significant improvement on previous datasets due to their improved resolution and coverage. As well, these examples demonstrate insights gained into Sydney's transport network and customer response that were made possible through analysis of Opal datasets

1.3.1 Measuring demand

This section examines the urgent need to improve our understanding of the transport network delivered in NSW and passengers' response to that network.

On an average workday in NSW during 2016, there was over 1.2 million embarkations on the heavy rail and 0.9 million embarkations on the buses (Hounsell 2017a). As shown in Figure 11 below, during July 2018 to June 2019 Australian Financial Year (FY) the Opal ETS, recorded 733 million embarkations (Tap-Ons), which was a 9.83% increase over FY 2016-17 (667 million).

With over two-million records per day, these datasets are beyond manual analysis and comprehension. The fact that Big Data regularly exceeds the limits of individual understanding, is forcing computer scientists (and others) to develop new methods for analysis and communication (Tao, Rohde & Corcoran 2014; Zhong et al. 2015).

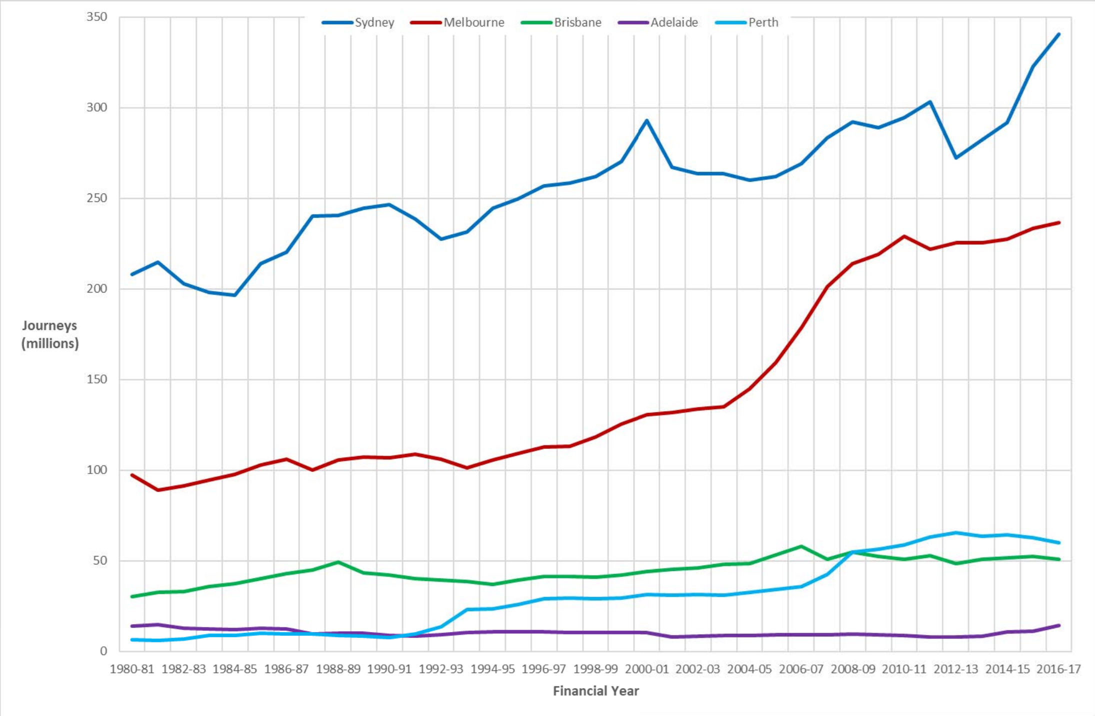
Figure 11: Public transport embarkations in NSW GMA per Financial Year

Source: • (TfNSW 2019b)

Before operators analyse passenger choices, they must accurately measure the number of trips. Incorrect measurements of trips result in incorrect conclusions that inevitably lead to incorrect, or even counterproductive, actions by service partners – see Chapter 5.

Public transit patronage on heavy rail in Australian metropolitan regions has increased steadily since 1980-81. In Sydney, the Airport Rail Link opened in May 2000 and the Epping to Chatswood Rail Link in February 2009. In Perth, the Armadale Line opened in May 1989, the Joondalup Line opened in 1993, and the Mandurah Line in December 2007. After the opening of these rail lines there were observed increases in patronage in Sydney and Perth, possibly from released demand or induced demand (McNally & Rindt 2007; Rose et al. 2013). Melbourne's patronage increased substantially from 2002 to 2008, however the increase cannot be directly attributed to a specific new line.

Figure 12: Heavy rail patronage in Australia per FY since FY 1980-81



Sources: (BITRE 2018, Table T 5.5a)

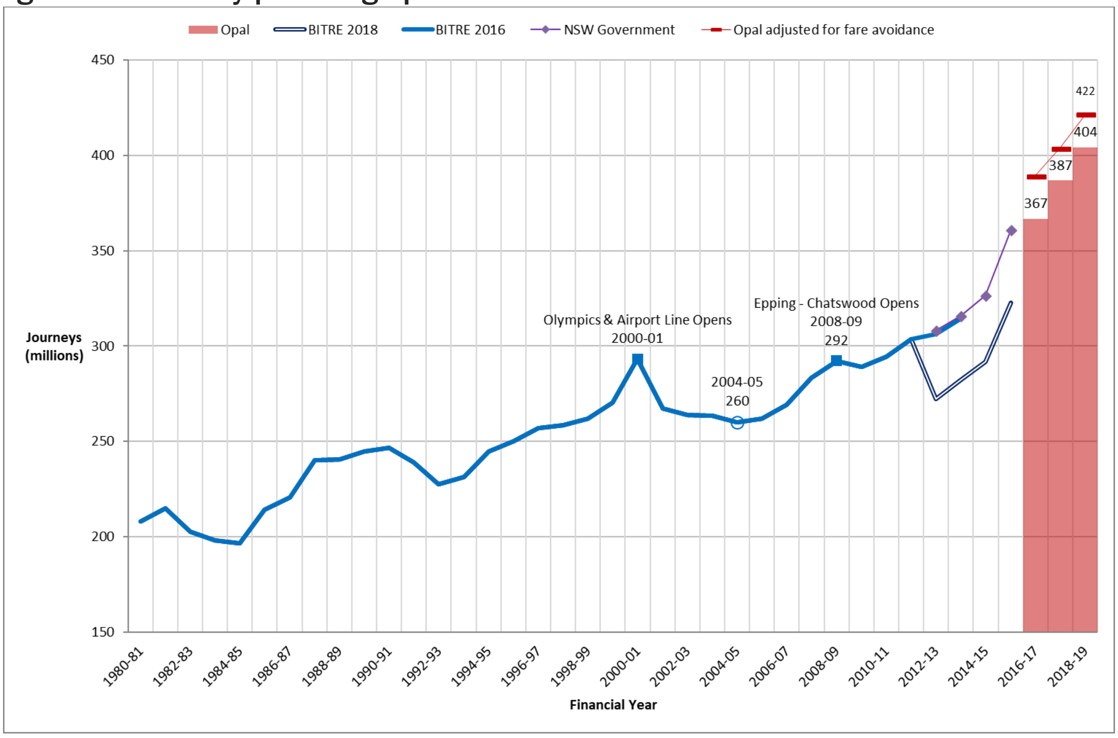
Figure 12 above was based on the 2018 Infrastructure Statistics from the CoA Bureau of Infrastructure, Transport and Regional Economics (BITRE). According to BITRE (2018), there was a significant reduction in heavy rail patronage in Sydney in FY 2012-13 and FY 2013-14. However those results are different to the data provided in BITRE (2016). In addition, the BITRE numbers differ from those in the TfNSW Annual Report for 2013/14 (TfNSW 2014b). As well, the 2014-15 and 2015-16 BITRE numbers are different to those reported in the NSW Budget 2017-18 (NSW-Treasury 2017).

As shown in Figure 13 below, the different numbers lead to substantially different interpretations of public transport patronage changes in NSW from July 1980 to June 2019. When examining the BITRE's 2018 statistics it appears there was a 10.2% fall in the patronage. However, considering the NSW Government's and Opal's numbers, there was a steady and substantial increase in the number of passengers.

As this example demonstrates, transport operators and partners need consistent and accurate measurements of patronage to support successful transport policy formulation and service delivery to the community.

Note: In Figure 13, the "Opal adjusted for fare avoidance" numbers are estimated using the Opal totals divided by the upper limit of fare compliance (≈95%) (TfNSW 2019a).

Figure 13: Railway patronage per FY in NSW GMA since FY 1980-81



Sources: (TfNSW 2019b), (BITRE 2016, Table T 5.5a), (BITRE 2018, Table T 5.5a), (TfNSW 2014b), (NSW-Treasury 2017)

1.3.2 The urgent need in NSW

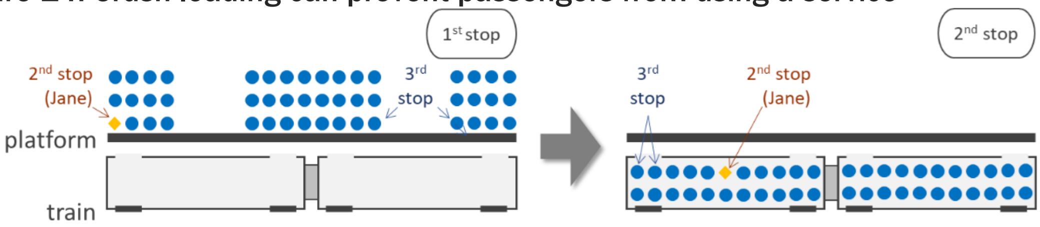
Transport partners and operators need high resolution data because aggregation over a month provides a very limited picture of the situation on a transport service. As an example, in Sydney there has been significant increases in patronage on the NSW urban rail network. The Western Line and Illawarra Line are two key lines into the Central Business District (CBD). Both are now at the functional limits of the system (TfNSW 2016c), but overcrowding is not evident from annual patronage figures.

The full-load of a vehicle is the level of passenger loading where the limited space within the vehicle delays passengers boarding and alighting and thus causes the dwell-times to increase, causing reliability and on-time running to be compromised (assuming an efficiently scheduled service).

> 'A load factor of 100% means there is a seat for each passenger. A load factor of 135% is the benchmark beyond which passengers experience crowding and dwell times can impact on-time running.' (TfNSW 2016c)

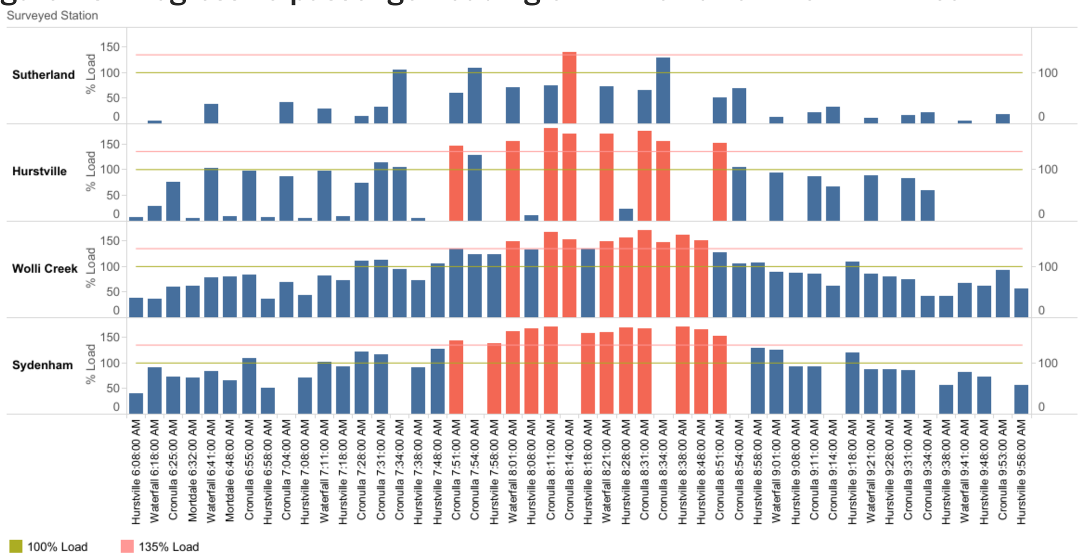
The crush-load of the vehicle is the level of passenger loading where a person is unable to move through the carriage to alight – the absolute maximum-load of a vehicle. At crush loads, the random distribution of passengers causes a high probability of a person being trapped in the vehicle and unable to alight at their stop because the majority of passengers in the vehicle unintentionally block the exits waiting for a later stop.

Figure 14: Crush loading can prevent passengers from using a service



The estimates of vehicle passenger loadings were compiled from manual inspections of the trains and are graphed in Figure 15 and Figure 16 below(TfNSW 2016c). On many lines to the Sydney CBD the services regularly, and for extended periods, exceed the operational full-load for Sydney Trains fleet.

Figure 15: Progressive passenger loading on T4 Illawarra Line in AM Peak



Note: the legend in the bottom right refers to the lines overlaid on each station plot; not the columns.

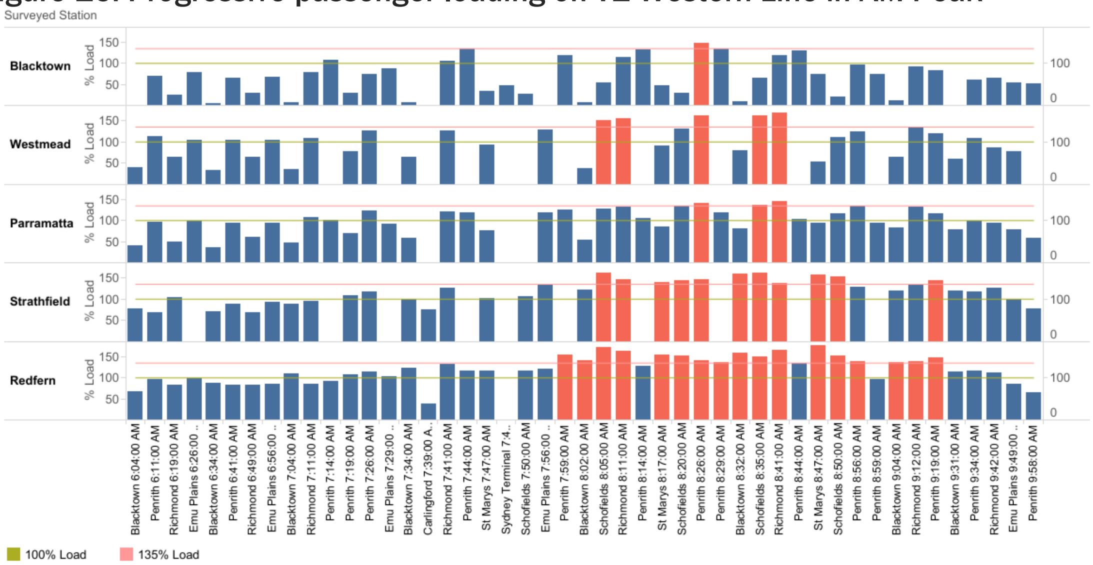
Note: the horizontal axis lists the origin station and arrival time at Central Station at the edge of the CBD.

Source: (TfNSW 2016c)

Increased passenger demand is causing the system to suffer disruptions to services and reduces on-time running performance. There is an urgent need to address the crowding as it is impeding the operator from delivering the services they promised to customers.

In the last few years, service partners have turned to weight sensors, door counters, and the Opal ETS to compute the probable loads on individual vehicles. This loading data is generated even in off-peak periods and on minor services and at all locations. As such, this Big Data allows operators to identify crowding on all services, at all locations, and at all times without incurring additional survey costs. The ability to diagnose problems, wherever they occur, is a powerful new tool for service partners.

Figure 16: Progressive passenger loading on T1 Western Line in AM Peak



Note: the legend in the bottom right refers to the lines overlaid on each station plot; not the columns.

Note: the horizontal axis lists the origin station and arrival time at Central Station at the edge of the CBD.

Source: (TfNSW 2016c)

1.3.3 Resolution in historical versus TOTOR datasets

Resolution is the scale of a change that can be detected from a dataset by a given measurement. The resolution of a dataset and its measurement is a key factor in its utility as a tool for transport analytics by operators.

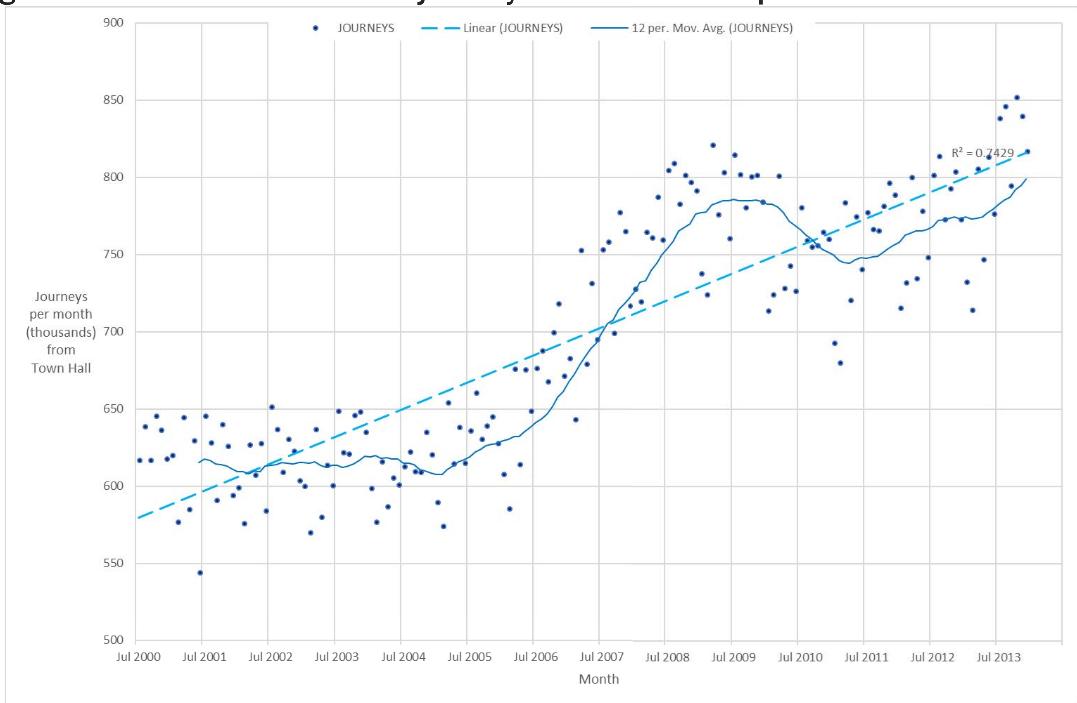
Consider this fictitious scenario, the standard business hours in the city are 09:00 to 17:00 and the peak of passenger travel (near full-loads) is between 08:00 and 09:00. As such, the transport operator uses the most vehicles to provide the highest route capacity during that hour. Now say the city's business leaders got together and decided to change the standard business hours to between 08:00 and 16:00. What measurements would have the resolution to detect the resulting crowding in passenger loads on the vehicles from 07:00 to 08:00? If the total number of trips does not change, the annual patronage figures would not show the change as it lacks the resolution to detect behaviour change.

1.3.3.1 Monthly patronage

Before 2015, NSW used a Magnetic Stripe Ticket (MST) to recover operating costs. Only a select number of stations, such as Town Hall, had ticket gates to record the number of passenger entries. Smaller stations did not have gates and there was no dataset recording actual patronage levels at small stations like Petersham. In reality, Town Hall's MST patronage dataset was also incomplete and used estimates for late evening and night operations because the fire and life safety policies required the ticket gates to be open when the barriers were unstaffed.

The Department of Transport (TfNSW) produced and release counts for the number of embarkations at stations like Town Hall for each month from July 2000 until December 2013. This monthly embarkations dataset for Town Hall is plotted in Figure 17 below.

Figure 17: Estimated number of journeys from Town Hall per Month



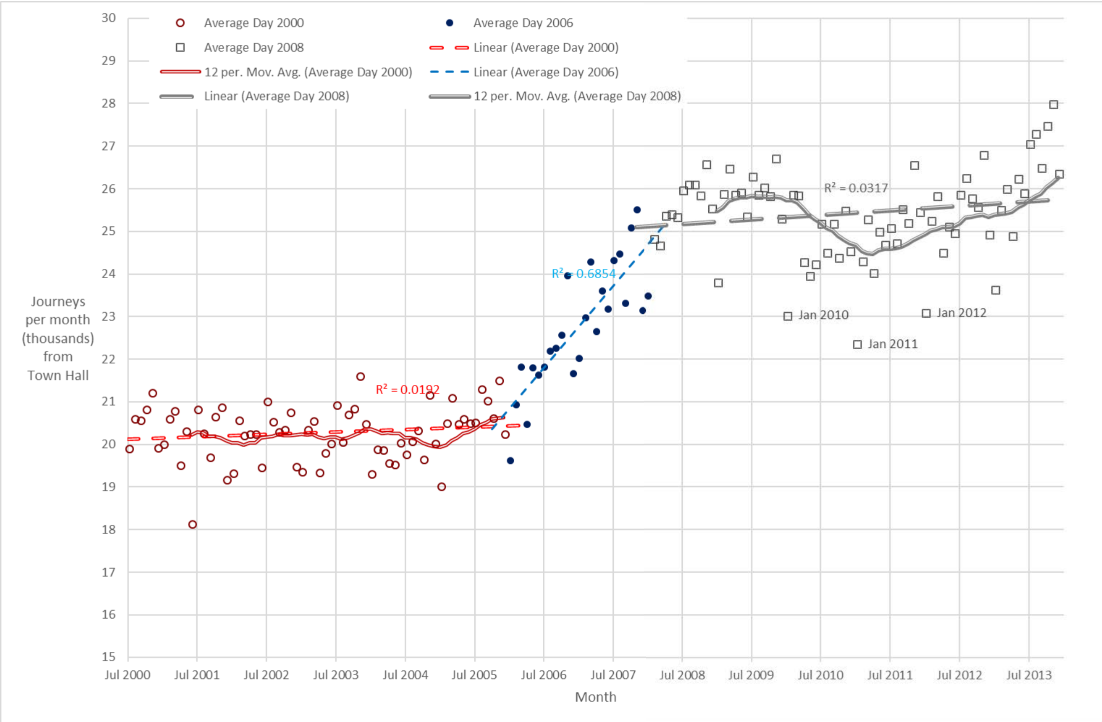
Source: • (TfNSW 2014a)

The figure shows that there is substantial variation between each month. There is an apparent surge in the number of journeys between Jan 2006 to Dec 2007. That change is also clearly shown in Figure 29 below, which plots the average journeys per day broken down into three linear regressions[9]. Before Dec 2005 shows very little growth, before a surge between Jan 2006 and Dec 2007 ($r^2 = 0.685$), then levelling off until Jun 2013.

Despite being higher resolution than an annual dataset, these figures show that monthly datasets still lack the resolution to detect behaviour change and passenger crowding. Monthly datasets are useful but insufficient to measure changes in passenger behaviour.

[9] Average journeys per day = average journeys for the month ÷ number of days for the month.

Figure 18: Estimated number of journeys for an average day from Town Hall (calculated from monthly patronage)



Source: • (TfNSW 2014a)

1.3.3.2 Workdays versus weekends

An obvious question to ask is what caused the growth in journeys at Town Hall station? The original monthly journey dataset has sufficient resolution to allow us to measure that a change occurred and estimate the period over which it occurred. However, the monthly dataset does not have the resolution to allow us to measure what changed.

Table 1 below summarises the number of embarkations / disembarkations recorded by Opal at locations in the NSW GMA for every day over a week in November 2016. This table shows that the number of weekend embarkations on the railway in the Sydney CBD is substantially less than on an average workday (Tuesday, Wednesday, & Thursday). Note that there was track-work on that particular weekend which caused Circular Quay, Museum, and St James stations to be closed.

To verify that weekend in November is indicative for the railways, a summary for a week in August is shown in Table 2 below. Unfortunately, those are the only two regular weeks for which weekend data is publicly available.

These tables show that there is a significant difference in the number of journeys on the NSW railways to Town Hall on these weekends in 2016. This fact raises interesting questions about the growth in monthly journeys at Town Hall between 2005 and 2008. Was the monthly increase caused by growth in weekday patronage, weekend patronage or did all weekdays grow at the same rate? Without a higher resolution daily dataset, it is impossible to be sure about the causes of the change detected.

Table 1: Embarkations from Sydney CBD Locations between 2016-11-21 to 27

Mode	Location	Workday (,000)	Saturday (,000)	Sat %	Sunday (,000)	Sun %
Train	Central	105	74	*70%*	64	*61%*
	Circular Quay	30	Track Work			
	Martin Place	24	8	*37%*	6	*26%*
	Museum	13	Track Work			
	St James	11	Track Work			
	Town Hall	112	75	*67%*	57	*51%*
	Wynyard	77	27	*35%*	20	*27%*
Total	*Train, Bus, Ferry, & Light Rail*	*535*	*294*	*55%*	*245*	*45%*

Source: • (TfNSW 2017b)

Table 2: Embarkations from Sydney CBD Locations between 2016-11-08 to 14

Mode	Location	Average Workday	Saturday	Saturday %	Sunday	Sunday %
Train	Central	117	61	53%	56	49%
	Circular Quay	29	26	90%	31	108%
	Martin Place	23	4	21%	3	16%
	Museum	12	5	40%	4	35%
	St James	10	4	42%	4	40%
	Town Hall	110	63	58%	53	48%
	Wynyard	77	18	23%	14	19%
Total	Train, Bus, Ferry, & Light Rail	547	281	51%	263	48%

Source: • (TfNSW 2017a)

Weekend connectivity

It seems obvious that the number of passengers using a public transport service during the weekend would be less than the patronage during the working week. In fact, many surveys indicate people consider public transport primarily as a mechanism for commuting. Often, service partners seem to approach weekend transport as merely an additional social service. The truism that patronage will always be minimal on the weekends is a given for many within the Australian transport industry. However, higher resolution TOTOR datasets record weekend patronage as accurately as weekday patronage, and it is now possible to determine that truism is in fact a Misknown-Known.

The weekend patronage is lower and that may be because the weekend public transport service level does not meet the customers' needs and the customers respond by driving. However, the reasons will be more complicated and further research is required into why weekend public transport does not meet the service quality sought by customers.

As shown later in this thesis in section 5.3.1 – Mobility Datasets, the Queensland dataset empirically shows that weekend passengers respond positively to the Gold Coast Light Rail (G:Link) which targets a 10-minute vehicle headway on the weekend. After Houston redesigned their bus-network, the weekend patronage increased immediately after weekend services were improved with a high-frequency grid network facilitating better access to a variety of destinations, suggesting those features attracted customers.

> 'Weekend ridership spiked immediately in response to the expanded weekend service, and numbers on the light rail system are way up since the new network better complements the three new lines. Most importantly, the service matches where many people are making trips, so more and more are likely to discover that transit is now a good option for them.' (Llamas 2016)

Do weekends always have lower patronage?

Table 3 below shows that weekend patronage on buses in the Sydney CBD drops to less than 60% of workday levels. However, the patronage on the light rail and ferries increases on weekends (especially Sundays), to exceed workday patronage levels.

Table 3: Embarkations (,000) from Sydney CBD between 2016-11-21 to 27

Mode	Location	Workday	Saturday	Saturday %	Sunday	Sunday %
Bus	2000 Series Bus Stops	135	79	*58%*	62	*46%*
Light Rail	Capitol Square	1	1	*93%*	1	*103%*
	Central (Light Rail)	8	6	*78%*	6	*83%*
	Paddy's Markets	1	2	*113%*	2	*155%*
Ferry	Circular Quay Wharves	17	20	*122%*	22	*132%*
	Darling Harbour Wharf	1	2	*126%*	3	*163%*

Source: • (TfNSW 2017b)

Sydney Harbour is regarded as one of the most beautiful harbours in the world, and locals and tourists alike use its ferry network to derive experiential and social value rather than pure mobility or access (IPART 2019; Rihova et al. 2015; Smith & Colgate 2007). As such, ferry patronage is higher on weekends than on weekdays.

Sydney light rail patronage on the weekend remains similar to workday patronage. Discussions with Newcastle and Canberra indicate their light rail networks also have substantial weekend patronage. Further research is required into the desirability of light rail on weekends, and whether that represents passengers responding to the higher levels of service or whether other factors are involved.

What is more, this pattern extends across the total patronage throughout the NSW GMA. The following table shows the patronage throughout the Opal area on trains and buses in the GMA does decline substantially on the weekend. However, patronage on the ferries increases and on the light rail remains steady over the weekend.

Table 4: Embarkations (,000) by mode in NSW GMA in August and November 2016

Passengers (,000)	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
August 2016	8	9	10	11	12	13	14
Bus	945	960	962	966	921	434	333
Ferry	33	33	33	34	39	43	62
Light Rail	37	27	26	26	27	23	25
Train	1,216	1,257	1,258	1,280	1,265	644	551
November 2016	21	22	23	24	25	26	27
Bus	913	936	924	959	940	485	358
Ferry	40	42	35	42	48	47	51
Light Rail	26	29	30	29	30	23	25
Train	1,195	1,232	1,239	1,295	1,302	654	503

Source: • (TfNSW 2017a) • (TfNSW 2017b)

The ferries operate on Sydney Harbour to key leisure destinations like Taronga Zoo and Manly. Is the increase in patronage purely due to changed mobility patterns, or are the ferries being used to meet other customer values? Further research into customer-centric transport is warranted as it may explain why the ferry and light rail examples are exceptions to the rule expecting lower weekend patronage.

Light Rail Daily Patronage

Compiling the daily patronage numbers for the IWLR from a higher resolution Opal dataset (January – June 2016) produces the results shown in Table 5 & Figure 19 below.

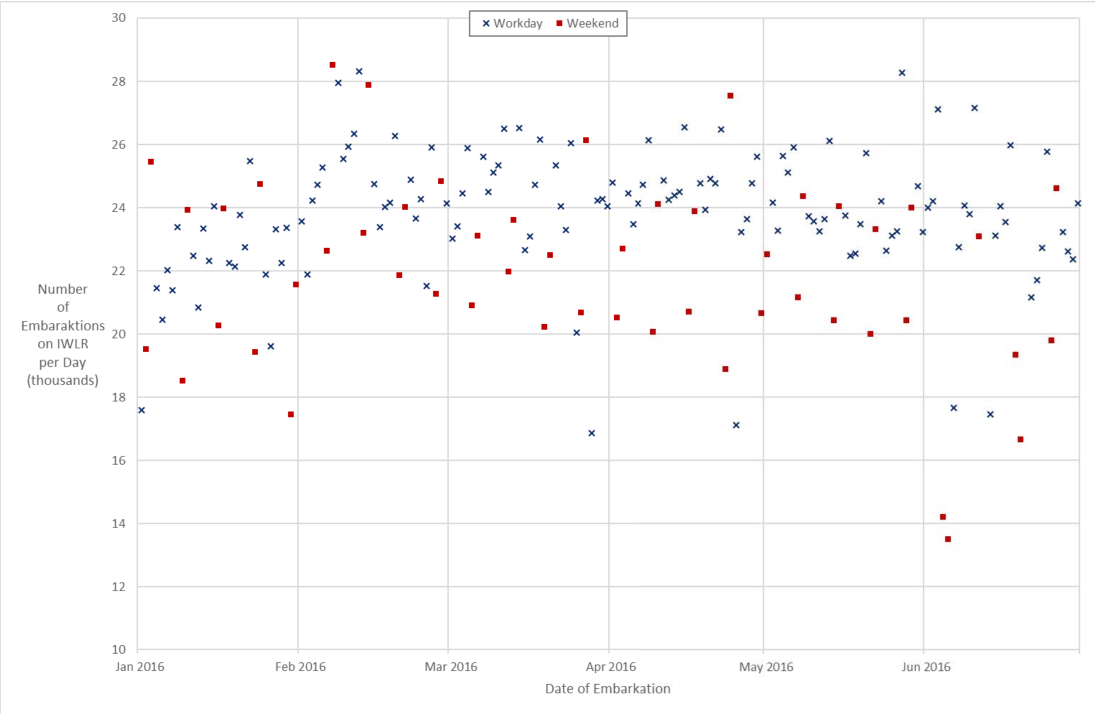
High resolution analysis allows us to empirically codify a Known-Unknown to answer the question – is IWLR patronage steady over every day of the week? The numbers show that the average weekend patronage on the IWLR is over 93% of the average workday patronage. In addition, the average patronage for all days is close to the average for the workdays and for the weekend days. As well, the median numbers are similar to the mean numbers indicating that the data is not skewed. The statistical measurements and the plot of patronage show that IWLR patronage does not vary substantially between workdays and weekends. It would not be possible to make that conclusion with a dataset resolution of a week, month, or year.

Table 5: Comparison of daily IWLR patronage from 2016-01-01 to 2016-06-30

Measurement	All Days	Workday	Weekend	Weekend vs Workday	Workday vs Weekend
Minimum	3,994	3,994	13,538		
Mean	23,278	23,721	22,171	93.5%	107.0%
Median	23,703	24,081	22,270	92.5%	108.1%
Maximum	32,825	28,364	32,825		
Standard Deviation	2,966.4	2,684.6	3,326.2		
Count	182	130	52		
Standard Error	N/A all dates were used so these numbers represent a census.				

Source: • TfNSW Light Rail Full Financial Year

Figure 19: Number of journeys on the IWLR between 2016-01-01 and 2016-06-30



Source: • TfNSW Light Rail Full Financial Year

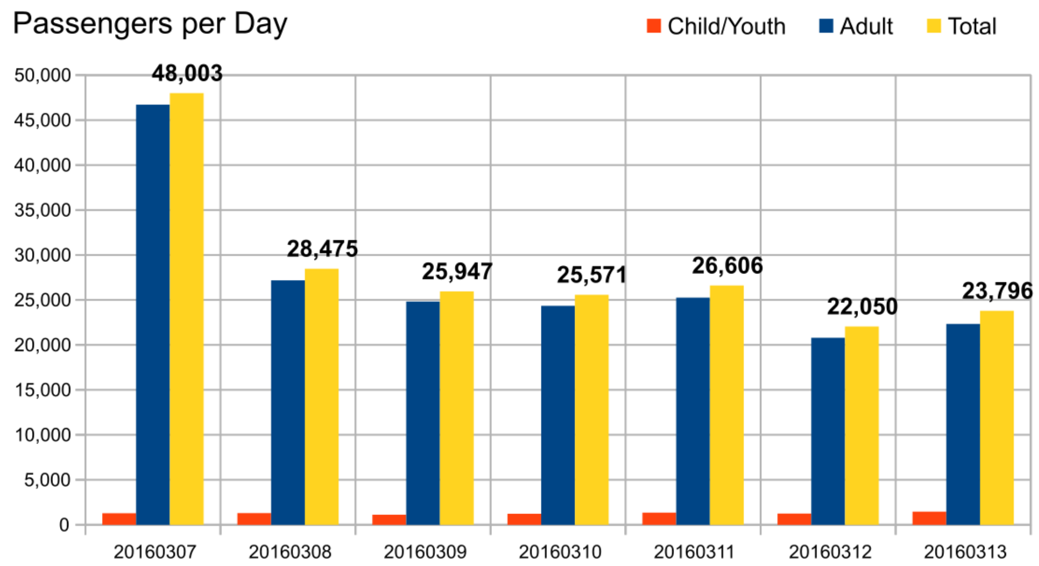
This high-resolution dataset, empirically disproved an assumption that patronage on transit would always be lower on weekends – thus a Known-Known was in fact a Misknown-Known. If passengers respond positively to a high frequency public transport network on weekends, it may be possible to achieve a shift to more sustainable public transport outside the peak by improving frequency, span, reliability and connectivity. Further empirical research is recommended to determine the factors that are encouraging the use of more highly polluting transport over sustainable transport.

1.3.4 Distortions in the data

In May 2016, a demonstration TOTOR analysis was prepared for the IWLR to be delivered at the launch of the UTS TRC. The analysis was conducted, and a full public presentation was prepared in one man-week using spreadsheets in LibreOffice Calc.

The demonstration analysis showed unusual passenger behaviour in Figure 20 below. The number of embarkations aggregated by day shows 48,000 Tap-Ons on Monday 7th March 2016 but only 26,000 Tap-Ons occurred on Wednesday. Figure 20, Figure 21, and Figure 22 are reproduced as-is from the original demonstration.

Figure 20: Embarkations by day per card type on the IWLR

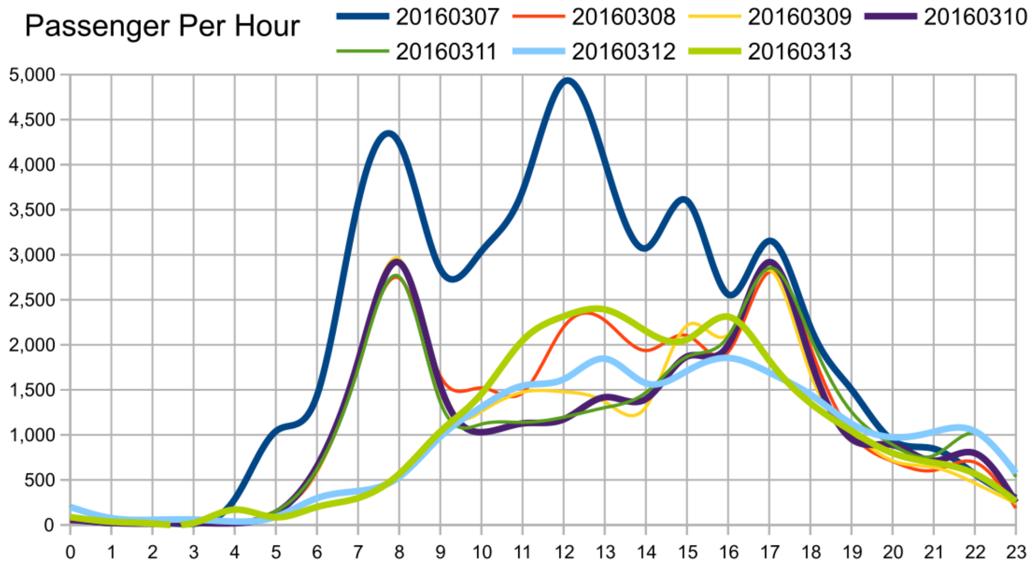


One advantage of the Opal dataset is each transaction records the timestamp to the nearest minute. Aggregating the dataset by hour of the day allows a more nuanced analysis of the passenger response to the transport system than was possible before these datasets.

In Figure 21 below, the count of Tap-On by each hour per day is plotted as a line graph. The dark blue line represents Monday Tap-Ons on the IWLR. There were substantially more taps during Monday morning, lunchtime and afternoon than on other days. This indicates there may be an environmental or behavioural artefact distorting the data.

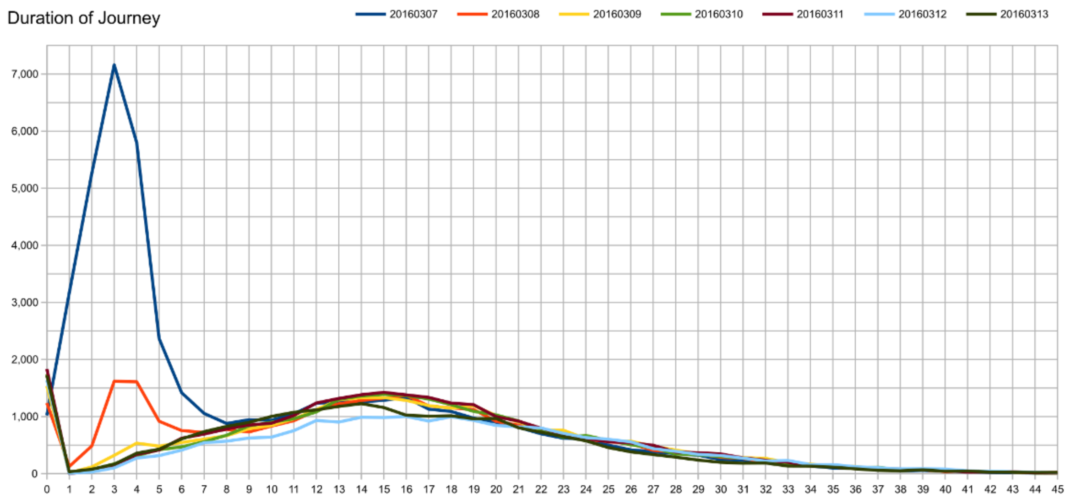
Since the Opal dataset is a high-resolution record of customer response to the transport network, operators and researchers are able to develop novel techniques to expand the operators' knowledge by addressing what is currently Unknown.

Figure 21: Embarkations by hour per day on the IWLR



In this case, the Opal dataset can be used to compute the duration of each journey segment by calculating the duration between the Tap-On and Tap-Off timestamps. In Figure 22 below the frequency of journey segment durations per day was plotted from the demonstration dataset. Note that for Wednesday through to Sunday the durations form a skewed bell curve around 15 minutes. However, for Monday there is a spike for journeys of five minutes or less and a similar spike on Tuesday.

Figure 22: Count of journey segment by duration per day on the IWLR



The average wait time for random arrivals would be 4 minutes – half of the vehicle-headway (at least 8-minutes). All journeys that take one minute or less on the light rail travel only one stop along the line. Therefore, that spike does not match the expected measurements of normal passenger behaviour. So, what Unknown caused that spike?

1.3.4.1 Fare Minimisation

To increase the speed of Opal card take-up, NSW offered free travel after eight trips in a week. This popular incentive ensured the majority of the population adopted the smart card tickets within a short time period. Once the roll-out incentive was no longer required, the government switched to a higher priced fare regime.

NSW operates a distance-based fares structure and for persons travelling long distances the price for their regular commute would substantially exceed the price of short trips. The Opal fare for an outer Sydney train commuter was $8.30 per trip but a one stop light rail trip was $2.10 (46.9% vs 11.9% of hourly minimum wage). The Paddy's Market and Capitol Square light rail stops are 250m apart, The Star and Pyrmont Bay light rail stops are 350m apart.

As such, many public transport users in NSW engaged in a fare minimisation strategy, whereby they would walk to one of these light rail stops to Tap-On and then walk to another stop and Tap-Off (Saulwick & Kontominas 2015). Since these stops have off-board ticket readers and were close together, a person did not need to wait for a vehicle to undertake fare minimisation, they would simply walk. Fare minimisation is the probable reason for the large number of short travel durations in the previous figure.

The high resolution of the Opal dataset allows us to test both this hypothesis and compensate for the artefact in future analysis. Table 6 below lists the Top-10 IWLR embarkation-disembarkation pairs by date sorted by the number of embarkations. It shows that during the week analysed, the four most popular pairs had the shortest walking distance and were those identified by the press as popular locations for fare minimisation. In addition, all of the 22,000 journey segments between the top four pairs were recorded on Monday.

Five of the next six most popular pairs were from Central to The Star and they occurred on Wednesday, Friday, Thursday, Tuesday, and Saturday, all days when there was a show at the Lyric Theatre. On Sunday, when fares were capped at $2.50 for the entire day, there was only 20 segments recorded between Pyrmont Bay and The Star, which is a stark contrast to over 8,000 segments on Monday. It is reasonable to conclude that these four combinations (highlighted in blue) are popular solely for fare minimisation.

Table 6: Top 10 travel pairs by count per date[10]

Tap On	Tap Off	Date	Rank	Count
Pyrmont Bay	The Star	Mon 7 Mar 2016	1	8,040
The Star	Pyrmont Bay	Mon 7 Mar 2016	2	7,820
Capitol Square	Paddy's Markets	Mon 7 Mar 2016	3	3,610
Paddy's Markets	Capitol Square	Mon 7 Mar 2016	4	2,860
Central Station	The Star	Wed 9 Mar 2016	5	2,090
Central Station	The Star	Fri 11 Mar 2016	6	1,920
Central Station	The Star	Thu 10 Mar 2016	7	1,870
The Star	Central Station	Wed 9 Mar 2016	8	1,780
Central Station	The Star	Tue 8 Mar 2016	9	1,770
Central Station	The Star	Sat 12 Mar 2016	10	1,760

Therefore, during the period when this offer was available, any analysis of the IWLR needs to compensate for this artefact. Due to the resolution of the dataset it is possible to explicitly exclude just those pairs used for fare minimisation. The effect of excluding the minimisation pairs is demonstrated in Figure 23 below.

Figure 23: Embarkations on the IWLR with probable Fare Minimisation highlighted.

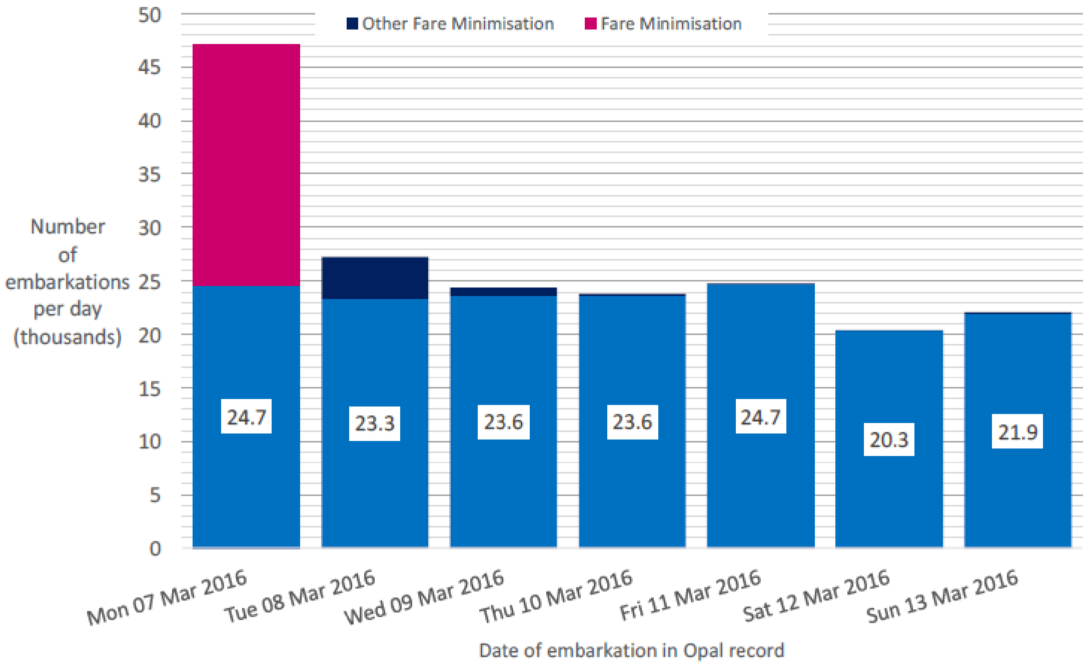


Figure 23 shows the number of embarkations by day colour-coded per categorisation – either as fare minimisation (pink/dark blue) or actual trips (light blue). The revised categorisations show a realistic pattern with the steady number of journeys during the workdays. While this fare discount was available the number of passengers on the

[10] On Thursday, Friday and Saturday there was 20 segments or less between Pyrmont Bay and The Star.

IWLR on Monday was substantially lower than the number of tickets purchased. This trend was detectable in the high resolution TOTOR dataset but not in low resolution monthly patronage data.

1.3.4.2 TOTOR versus TOR

Most smartcard ETS only record the Tap-On location for each journey segment creating a Tap-On Recording (TOR) dataset. If the Opal system only recorded the Tap-On location, it would not have been possible to isolate the fare minimisation because the passenger could have disembarked at any stop on the network. Analyses of passenger behaviour during that period using a TOR dataset would need to either estimate the level of fare minimisation or exclude all data from the affected stops or affected days.

When using a TOR dataset, excluding the Paddy's Markets stop entirely would be a reasonable technique to compensate for fare minimisation. A TOR dataset does not record the passenger's Tap-Off, therefore, we can assess the analysis limitations of a TOR dataset compared to a TOTOR dataset by "deleting" the Tap-Off data from the TOTOR dataset and then conducting the analysis.

Table 7 below shows the number of Tap-Ons at the Paddy's Market in the CBD at Haymarket. Table 7 shows that excluding the Paddy's Markets stop entirely, would underestimate the patronage on Monday by over 1,600 journey segments. Between Thursday and Sunday, less than 3% of the journey segments embarking from Paddy's Market could be classified as probable fare minimisation. Excluding Paddy's Market would alter the interpretation of the data because when fare minimisation is minimal – Thursday to Sunday – this stop is the 3rd or 4th most patronised stop on the IWLR.

Table 7: Tap-Off stop for a Tap-On at Paddy's Market stop by Fare Minimisation

Tap-Off stop from Paddy's Market in March 2016	Mon 7	Tue 8	Wed 9	Thu 10	Fri 11	Sat 12	Sun 13
All Stops	4,550	2,280	1,820	1,850	2,000	2,150	2,950
Capitol Square (Possible Fare Minimisation)	2,860	870	200	60	20	10	60
Not Fare Minimisation	1,690	1,410	1,620	1,790	1,980	2,140	2,890
Not Fare Minimisation %	*37%*	*62%*	*89%*	*97%*	*99%*	*100%*	*98%*

Source: • TfNSW Light Rail Full Financial Year

A TOR dataset would have proven insufficient to allow for the codification of this Unknown, preventing accurate measurement of actual passenger behaviour. Worse still, the compensation mechanisms may have resulted in a misinterpretation of actual behaviour due to the analyst being forced to exclude important but tainted data.

As described in the literature review for Stage 1 of this thesis (Hounsell 2017c), a lot of research has been conducted into inferring the disembarkation location from subsequent embarkations and other data sources. However, none of those methods are reliable enough to accurately diagnose the fare minimisation seen on the IWLR.

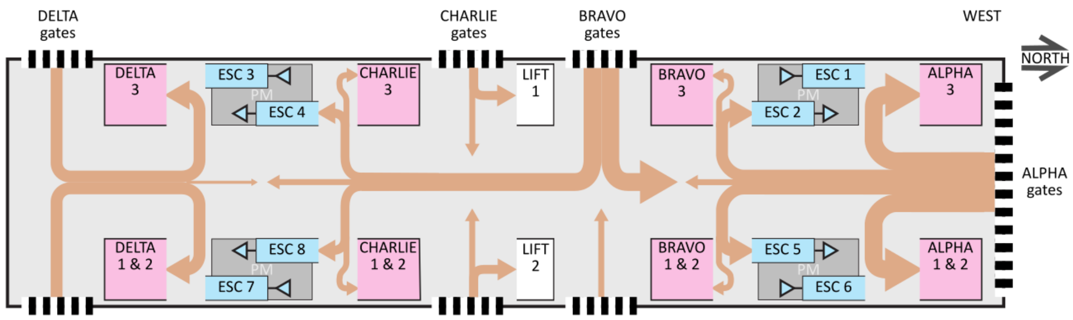
The NSW government suspended the eight trips offer in September 2016. In Table 2 above, the embarkations for Light Rail in August 2016 for Monday were 10,000 embarkations above Monday in November 2016 – the August data is problematic because of the aforementioned fare-minimisation. Without fare minimisation, the November patronage returned to the actual passenger levels.

The Opal dataset is a high-resolution dataset and so it was possible to identify fare minimisation. Accurately correcting for fare-minimisation would not have been possible with a low-resolution or TOR dataset.

1.3.5 Minute by Minute

UTS was engaged by the service operator – Sydney Trains – to assist in developing cyber-physical systems to identify and compensate for human behaviours that were impacting the delivery of reliable train services in the crowded core of Sydney's rail network, especially at Town Hall and Wynyard Stations. ISF collaborated with FEIT and other UTS faculties in the research and development of a Responsive Passenger Information (RPI) systems in conjunction with Sydney Trains and TfNSW.

Figure 24: Diagram of Sydney's Town Hall Station concourse



One of the key issues identified by Sydney Trains staff was the impact of 'The River'. Station staff relayed how starting at around 17:07, a surge of after-work passengers arrived at the station gates for the evening commute. The primary set of gates at which the passengers arrive is the alpha (northern) gates, which is connect via a passageway under the roads to King Street (via the QVB) and Pitt Street (via the Galleries Victoria).

Figure 25 below shows the number of passengers Tapping-On at Town Hall station was over 20,000 persons in the peak hour starting at 17:00 on an average workday in 2016. The high-resolution Opal dataset allowed counting the number of people entering the station per hour. However, the Opal dataset has a time stamp with the resolution of one minute. Figure 26 empirically demonstrates that the absolute peak time for persons entering Town Hall station was indeed within the first 10 minutes after 17:00. The absolute peak was empirically measured as 500 persons per minute.

Note this peak may not represent total possible demand for access to the station, it may represent only the physical limits of the infrastructure (the gates) in admitting persons. Figure 26 below shows the 30 minutes average for number of entries to the station reaches 300 persons per minute and is sustained at over 200 persons per minute for over 1½ hours starting at 17:00. These high-resolution empirical measurements are only possible through the high-resolution of the Opal dataset.

Figure 25: Number of taps at Town Hall per hour – Workday – Nov. 2016

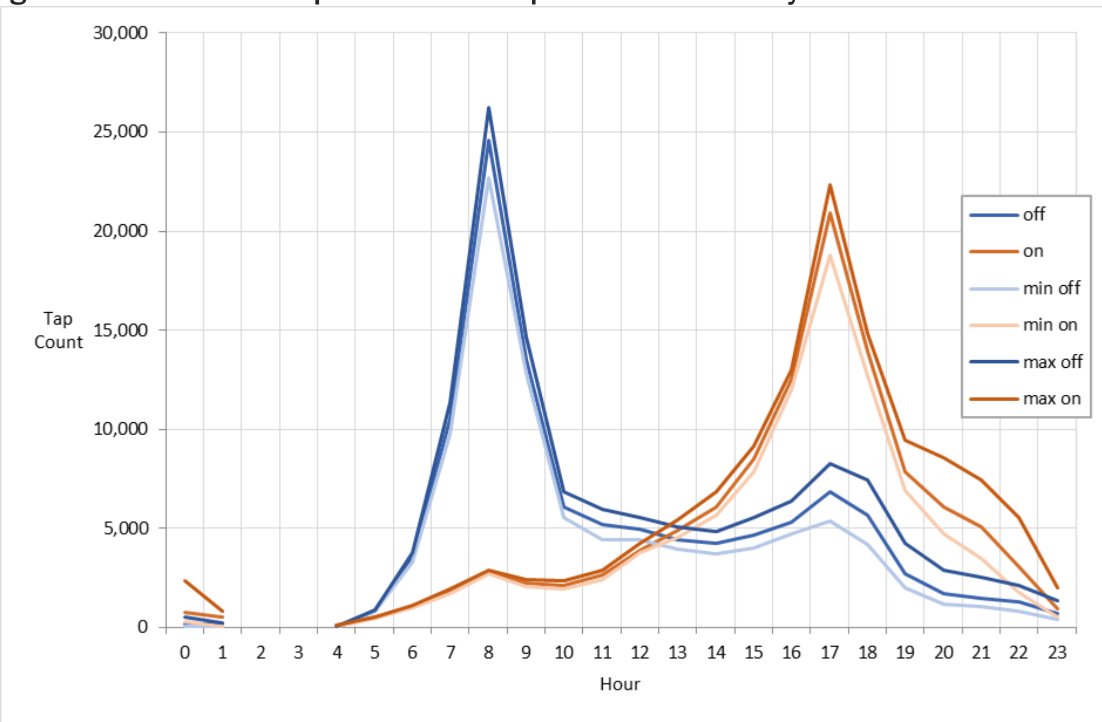
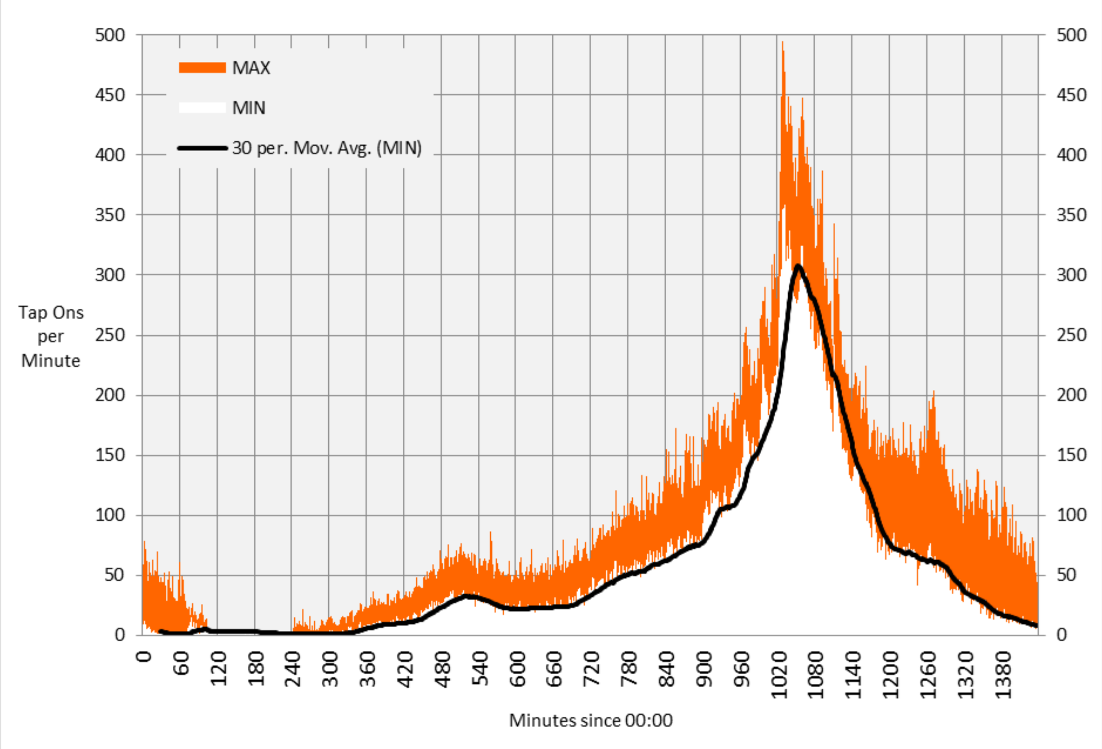


Figure 26: Embarkations (Tap-On) at Town Hall per minute – Workday – Nov. 2016



During the RPI system project, we investigated whether it was possible to resolve entries to the second to anticipate and predict surges. Unfortunately, the Opal system was only designed to support a resolution of one minute. Only once the ETS had a resolution of one minute, were we capable of conceiving uses for a dataset with the resolution of one second. During research and development, people rarely conceive of the uses for advanced technology, but once they have it, they conceive of ever more creative ways to use it (Spolsky 2004). Projection bias, say/do gap, and availability bias are three cognitive biases that limit research and development. Overcoming the veil of the Unknown that limits what individuals and organisations can conceive of is the work of researchers and good business analysts.

Ongoing research and development for the RPI system with Sydney Trains shows rich datasets enable researchers and operators to develop new tools for passenger behaviour analysis which were inconceivable before rich sources of empirical data were available.

1.4 The Knowledge Gap

The most important question in planning and engineering is the most difficult question to develop a complete and correct answer to: what do your customers need? The most important stage of any project is scoping and requirements gathering as they are used to identify customer needs (Pressman 1997, p. Chapter 2 & 11; Spolsky 2004, p. 53).

This question leads to two additional sub-questions: what service are you actually offering to customers? And how are customers responding to what you are actually offering? It is vital for organisations to remember the service they planned to deliver may not be the services they are delivering.

This section describes classes of knowledge and why the structure of knowledge result in organisations that are unable to answer those three important questions. To be an empirically driven organisation, the operator must first admit there will always be gaps in their knowledge, and to close those gaps an organisation must constantly seek to address those gaps. Understanding the different classes of knowledge allows an empirically-driven organisation to direct efforts to address the gaps in their knowledge.

An empirically driven organisation has a strategic advantage because it is driven by the assumption that it requires better information. Thus, it will have a commitment to continuously identify and address the limits of its knowledge and understanding through continuous engagement with customers and the industry to reduce its Blind Spot and Unknowns, as well as to seek any currently Hidden or Unexpressed truths.

Model-driven organisations will be less successful in the delivery of transport services because models are simplifications based only on current knowledge and assumptions. A model driven organisation will be unable to address the Blind-Spots & Unknowns that limit the organisations understanding of customer needs and industry best practice.

1.4.1 The Johari Window

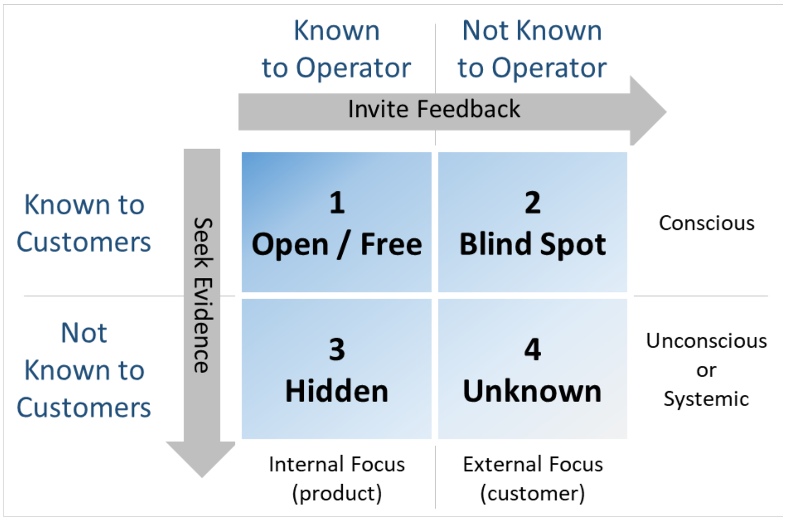
A Johari[11] Window is a tool used by psychologists to assess a person's understanding of themselves versus other persons' understanding of the person – see Figure 27 below. In the original exercise the subject and their peers are all given a single list of descriptors, then all participants in private pick descriptors they believe describe the subject.

Descriptors that are agreed by both the subject and reviewers is considered Open. Descriptors only picked by the reviewers are considered the subject's Blind-Spot. Any

[11] In Luft & Ingham (1961), they named their model "Johari" using a combination of their first names.

descriptors only picked by the subject are Hidden or Unexpressed, and unselected descriptors are considered Unknown or Undiscovered Potential. Traits selected only by the subject are Unexpressed and are an “opportunity for improvement” (Halpern 2009). Since its inception, this elegant technique has been expanded to a variety of fields such as organisational transformation and diplomacy / intelligence. The Johari Window is considered the origin of United States Secretary of Defence Donald Rumsfeld’s infamous Unknown-Unknown comments (Logan 2009) – see section 1.4.7.

Figure 27: Adapted Johari Window – Partners vs Customers



Source: (Luft & Ingham 1961, p. 1)

1.4.2 Knowledge of the operator and the customers

Organisations exist on a spectrum from product-centric to customer-centric (i.e. ②). A product-centric transport operator will have minimal shared knowledge with their customers and thus have a large Blind-Spot where they have not sought customer knowledge. In contrast, a customer-centric operator undertakes significant work to ensure that most of the customer knowledge is known to them, therefore they would have a very small Blind-Spot, as expressed in the Johari Window in Figure 27.

Humans are an interesting animal and not all of our actions are conscious. In fact, a significant body of research across psychology and neuroscience indicates that humans may physiologically respond to a situation and make a decision long before our consciousness has begun its assessment (Bechara & Damasio 2005). Although the fields of sociology, psychology, neuroscience, and other profilers of human behaviour continue their research, the complete truth about how and why humans engage with transport systems in the ways that we do is currently Unknown.

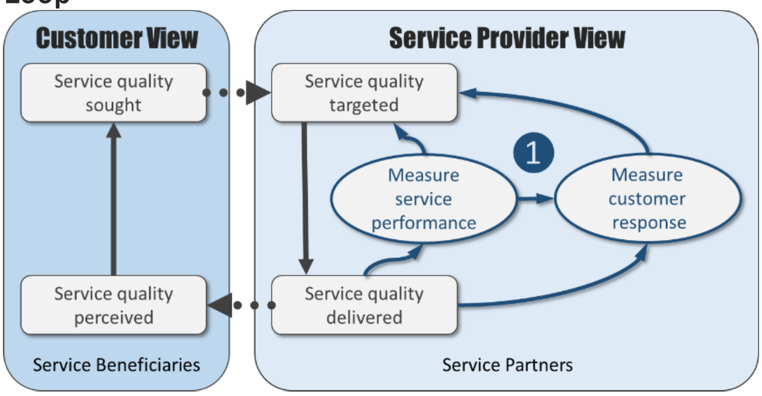
Customers do not intrinsically need transport / mobility, it is an extrinsic derived need used to satisfy their intrinsic needs. Service partners who are targeting the derived need for transport / mobility may misalign themselves with the customers actual desires. Service partners need to know their customers to serve their intrinsic needs.

Thus, the attempts to move from model-driven to empirically-driven (i.e. ①), seeks evidence to assist in the transition from serving what is consciously known and expressed (stated-preference) by customers, to also serving the unconscious expressions that arise from customers using transport to satisfy their desires (revealed-preference).

A transport operator planning to deliver (targeting) a service may not realise the service is not being delivered as planned (Blind-Spot), or that the customers do not perceive the service as the operator does (Hidden). As an example, an operator targets one bus every 10 minutes, but due to traffic there is bunching, and the delivery has 30-minute gaps. The failure to achieve the service targets will remain in the operator's Blind-Spot unless the operator is continuously measuring arrival times.

To overcome the Blind-Spot, the operator must measure the actual performance of the services being delivered, as well as the customer's response to both the promised services and the actual performance of the services delivered, as in Figure 28 below.

Figure 28: Reducing the Blind-Spot and Hidden areas using the Service Quality Loop



If the customers cannot easily interpret the service plan, they will not know how to use them to satisfy their desires. The more complex the timetable, the higher the cognitive load required to use the system and the less customers will engage with it. Legibility is part of the reason customers respond positively to high-frequency Metro networks (Rose et al. 2013; Walker 2012). Thus, significantly complex network designs can result in a significant proportion of the public transport network being in the Hidden quadrant.

1.4.3 Seven Demands

Walker (2012) summarises the desires and expectations customers repeatedly express to him and their order of evaluation as 'The Seven Demands of Useful Service' Walker (2012, p. 24). This framework is a simple description of passengers' stated desires that can be leveraged to interpret the customers response to the services targeted and delivered – Open quadrant.

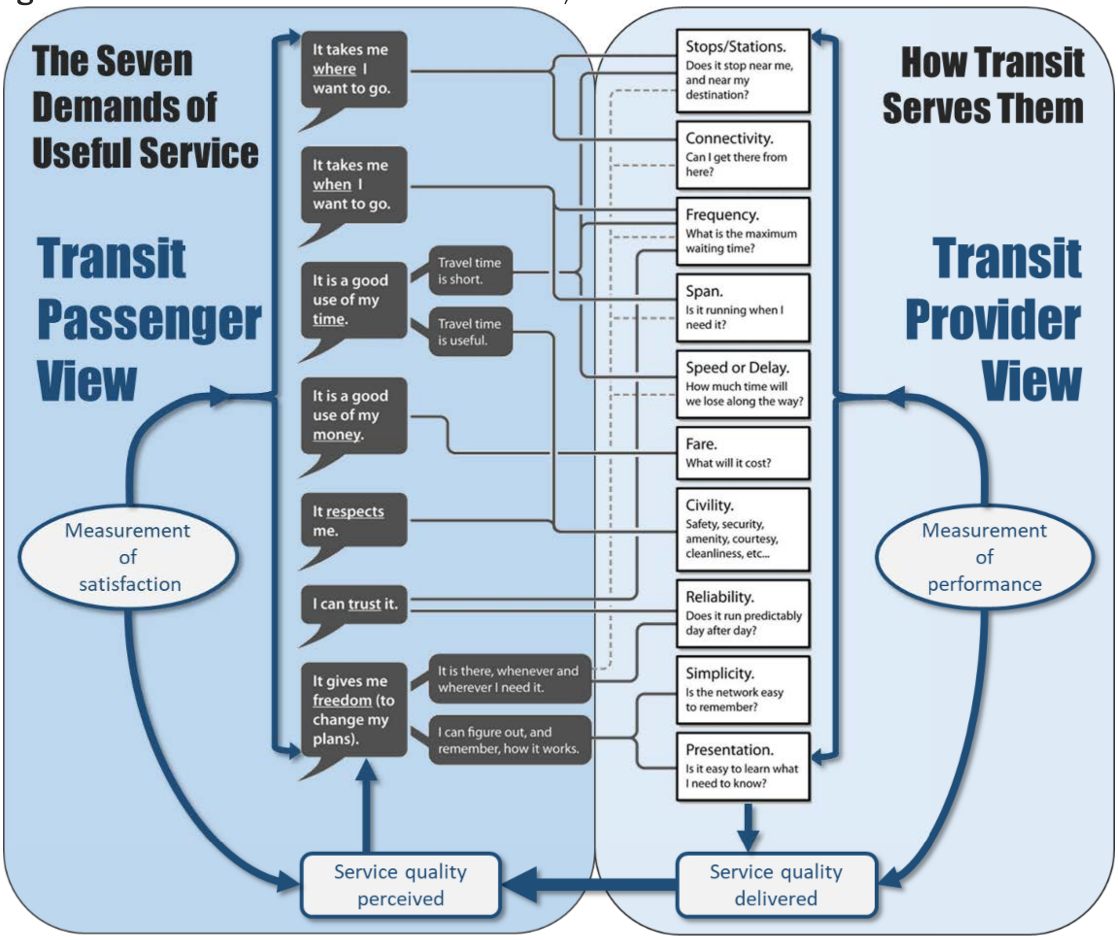
Note that these seven demands from the passengers are all self-centred and relate only to the intrinsic needs and desires of the customer — the subject of each statement is either me, myself, or I. None of the statements relate to the specifics of providing a public transport service – the product. All of the statements are based on the intrinsic needs of the passengers, but remember, they are not the only service beneficiaries.

Walker (2012) describes his framework and how it represents his interpretation of passenger behaviour quite concisely as follows:

> 'In the hundreds of hours I have spent listening to people talk about their transit needs, I've heard seven broad expectations that potential riders have of a transit service that they would consider riding:
>
> 1. It takes me where I want to go.
> 2. It takes me when I want to go.
> 3. It is a good use of my time.
> 4. It is a good use of my money.
> 5. It respects me in the level of safety, comfort, and amenity it provides.
> 6. I can trust it.
> 7. It gives me freedom to change my plans.
>
> 'I've listed the demands in the order in which you, as a customer, usually evaluate them. Generally, you would first evaluate transit in terms of whether it exists at all in the place where you need it (demand 1). Then you would consult a schedule and determine whether it is there when you need it (demand 2). Next, you might compare the cost of transit (in money and time) with the benefits (demands 3 and 4) to decide whether transit is worth trying compared to your alternatives.
>
> Now you are ready to try the service. You notice whether you feel comfortable and respected as a passenger, and whether you can put your travel time to good use (demand 5). If you become a regular customer, you start noticing whether the service works the same way day after day–in short, you decide whether you can trust it (demand 6).
>
> Finally, as your own needs vary from day to day, you begin to discover how well the service responds to those changes (demand 7).' (Walker 2012, p. 24)

Figure 29 below overlays Walker's tool (in grey) on the Service Quality Loop (in blue). 'The Seven Demands of Useful Service' correspond to the Service Quality Sought (and Perceived) under the passengers' view and 'How Transit Serves Them' corresponds to the Service Quality Targeted under the service partners view.

Figure 29: Seven desires for useful transit, and how transit serves them



Source: (Walker 2012, p. 27, Figure 2-1) . Credit: Eric Orozco

1.4.4 Frequency, Span, and Connectivity

Walker (2012) coined the phrase '*Frequency is Freedom*'. He states that frequency (vehicle-headway) along with span (hours-of-service) are fundamental for satisfying the passengers desires with public transport (Walker 2012, p. 85).

For example, Houston transitioned from a hub and spoke bus system to a high-frequency grid system with longer operating hours and more weekend services resulting in an increase in patronage. Similar changes were initiated in Seattle, Chicago, New York and in Los Angeles County (Hymon 2020). Los Angeles goal is that 'Buses would arrive every 5-10 minutes for 83% of current riders compared to 49% today'.

> ‘After years of declining bus ridership, last August Houston METRO overhauled service patterns around the city, updating the bus network for the first time since the 1970s. … From September 2015 … to July 2016… ridership on local bus and light-rail [added] an additional 4.5 million boardings – a 6.8% increase. … Local weekend bus ridership is one of the new system’s strongest areas, continuing a trend that begun [sic] almost immediately after the redesign was implemented. From June 2015 to June 2016 … local buses saw a 13% increase in ridership on Saturdays and a 34% increase on Sundays, according to METRO, with similarly strong numbers for rail as well.’ (Schmitt 2016)

Figure 30: Houston’s Revised Bus Network Map

Source: (Walker 2014)

NSW and Victoria talk about the need for all-day frequent transport services woven together to create a network like a web throughout the city. However, the service-quality targeted by the service-partners does not have the characteristics of service plans required to achieve those goals, especially in terms of frequency, span, and connectivity.

In 2004, Barrie Unsworth completed the *Review of bus services in New South Wales*, which recommended a network of bus-priority corridors with frequent all-day services. The majority of on-topic submissions from the public were about the need for a frequent and reliable bus network (Unsworth 2004).

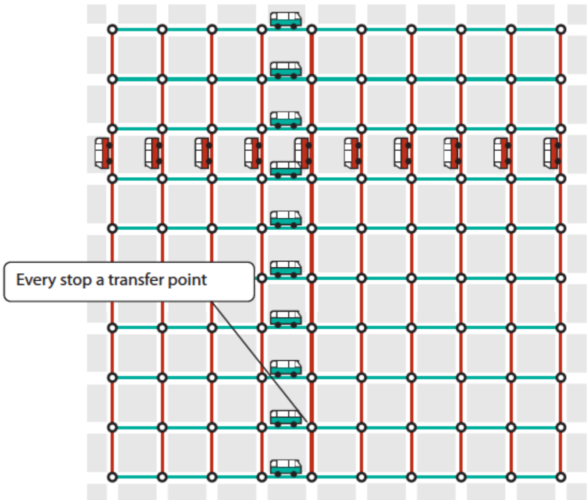
> ‘The major issues raised in submissions from organisations and individuals in metropolitan areas were the need to improve service planning and to ensure frequent and reliable bus services. … The Interim Report identified a network of strategic corridors in the [GMA]. These corridors reflect actual travel patterns, are anchored at regional centres and serve patronage generators such as district centres, campuses and hospitals. They will provide frequent, fast, convenient and direct services [to] integrate with local services, connect with other modes and provide a blueprint for improving bus priority measures.’ (Unsworth 2004, pp. 6, 8)

Unsworth’s recommendations led to the creation of the Metrobus routes, which began operation in October 2008 with the M10 running Leichhardt-Kingsford (discontinued).

In 2020, NSW is still considering plans to roll out a frequent bus priority network (TfNSW 2013). Meanwhile, the RMS has implemented a *Bus Priority Infrastructure Program* that has primarily involved removing bus stops – see Section 5.8

It has been empirically established that weekend trips are to a wider variety of locations than weekday trips (Hounsell 2017a; Martin & Thornton 2017; Thorpe 2017; Transurban 2016). Houston's focus on connectivity and frequency using Walker's Seven Demands may empirical demonstrate of Paul Mees' Squaresville thought-experiment in Figure 31. Mees (2015) postulated that rather than increasing the number of services on the existing north-south transport (red), the residents of Squaresville would be happier with *increased connectivity* by adding new services east-west (cyan). Further research, into the impacts of frequency and connectivity on customer response to a public transport is essential for developing customer-centric service targets.

Figure 31: Squaresville ultimate network



Source: (Bell 2015, p. 2, Figure 3)

1.4.4.1 The Seven Demands can be expressed as Levels of Service

The transport industry has looked at ways to measure and internalise the concepts of frequency, span and connectivity. The North American Transit Cooperative Research Program (TCRP) and Transportation Research Board (TRB) determined that vehicle headway (frequency) and hours-of-service (span) are important factors in passenger satisfaction. They codified their findings using a Level Of Service (LOS) scale as shown in Table 8 and Table 9. The TCRP and TRB found a qualitatively different passenger response to frequent services with vehicle-headways of 9 minutes or less.

Table 8: Fixed-Route Service Headway (Frequency) LOS

LOS	Vehicle-Headway (minutes)	Frequency (vehicles per hour)	Comments
A	<10	>6	Passengers do not need schedules
B	10-14	5-6	Frequent service, passengers consult schedules
C	15-20	3-4	Maximum desirable time to wait if bus/train missed
D	21-30	2	Service unattractive to “choice” riders
E	31-60	1	Service available during the hour
F	>60	<1	Service unattractive to all riders

Table 9: Fixed-Route Hours-of-Service (Span) LOS

LOS	Hours of Service	Comments
A	19-24	Night or “owl” service provided
B	17-18	Late evening service provided
C	14-16	Early evening service provided
D	12-13	Daytime service provided
E	4-11	Peak hour service only or limited midday service
F	0-3	Very limited or no service

Source: (Kittelson et al. 2003, pp. 3-30, Exhibit 3-13)

1.4.5 Knowledge of the operator and the industry

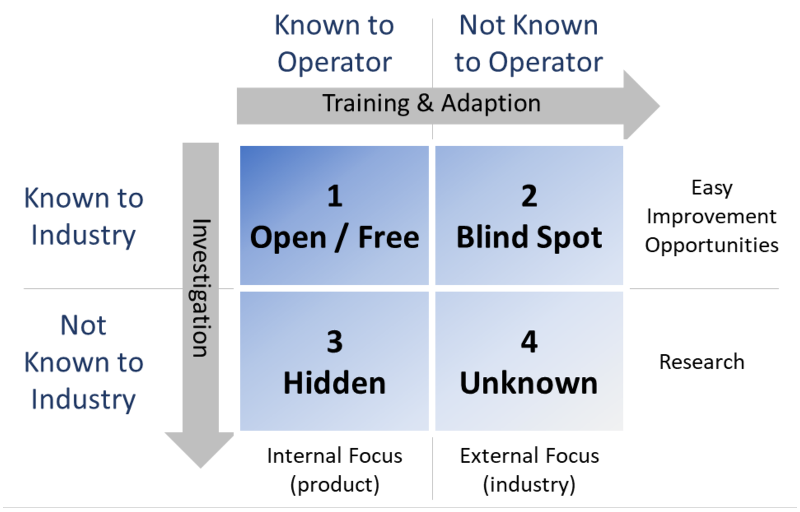
When transitioning to an empirically-driven customer-centric public transport operator, there are two obvious questions:

- Are there techniques that other operators use that we could adopt?
- Are there any easy wins?

According to the United Nations (UN) there are over 512 cities with more than one million inhabitants (UN 2016, p. 2). To sustain a population of that size invariably requires a public transport network. The Johari Window can be applied when considering skills and knowledge held by the subject versus skills and knowledge held by others in the industry. There are five hundred other transport operators in the world who may have knowledge that is in the Blind-Spot of Australian transport operators as conceptualised in Figure 32 below.

Figure 32 compares the skills and knowledge held by a transport operator in comparison to other operators or the transport industry as a whole. A skills-gap of an operator, but not the industry, could be overcome through training and adaption. Learning from the industry is a low-cost improvement opportunity for operators, even if contestability and competition are central to procurement policies.

Figure 32: Adapted Johari Window – Operator vs Industry



Adapted from: (Luft & Ingham 1961, p. 1)

The Hidden / Unexpressed quadrant can be considered as referring to trade secrets while the Unknown quadrant represents opportunities for further research.

An operator that is constantly assessing its Blind-Spot or looking for Unknown skills and knowledge can be considered to have an external focus. Whereas an operator focusing on what they already know can be considered to have an internal focus. Operators with an internal focus are unlikely to be adaptable or customer-centric because their focus upon their own internals creates a self-reinforcing cycle.

Other operators may have more skills and knowledge or better practises that allow a more empirically-driven (i.e. ①) or more customer-focused (i.e. ②) approach. Learning from the industry is one way to shrink an operator's Blind-Spot, meanwhile more comprehensive research is one way to shrink the Unknown.

1.4.6 Cognitive Bias

Achieving customer-centricity is a difficult task that requires a nuanced knowledge of customer needs that accurately represents the many and varied customers – thus maximising the operator's Open quadrant. Unfortunately, service operators are fallible and like most people suffer several forms of cognitive bias that increase the risk of developing an inaccurate picture of what customers do.

Before an operator can achieve a customer-centric approach, they must be able to correctly describe and interpret their customers' values. Values are the outcomes that are important to the entity (customer / operator), such as door-to-door travel or labour costs. Values are essential to decision-making because they enable evaluation and ranking of possible outcomes in order to select the preferred outcome. If the operators' values are not aligned with the customers values, when both are evaluating the same options, each may select very different options as their preferred option.

Cognitive biases limit the ability of people and organisations to clearly articulate their own values and increase the risk of decision-makers unintentionally misinterpreting and becoming misaligned with the customer's values.

Cognitive biases are an intrinsic feature of our species that impact our ability to think rationally. Liedtka (2015) states that "the literature on cognitive bias … provides a well-researched body of work over more than five decades delineating the flaws of human beings as information processors' (Liedtka 2015, p. 930). Emily Pronin coined the term Blind-Spot Bias to describe the phenomenon that 'Almost everyone feels that cognitive bias is a problem for other people. In contrast, almost no one appreciates the degree to which they suffer from cognitive biases' (Howard 2018, pp. 525-).

To deliver a customer-centric service, operators must accept that their knowledge is limited and so they need to work extensively with their customers to prevent cognitive biases from distorting their conception of customer values and biasing decisions away from the customers' needs.

The Egocentric Empathy Gap is particularly problematic for transport operations as it 'causes decision-makers to consistently overestimate the similarity between what they value and what others value' and it is a person's 'tendency to project their own thoughts, preferences and behaviours onto other people' (Liedtka 2015, p. 930).

Hypothesis Confirmation Bias and Endowment Bias predispose people to maintaining and arguing for a viewpoint despite evidence to the contrary (Liedtka 2015, p. 930). Humans have an innate need to maintain consistency in their public statements, which has been shown to influence decision-making (Cialdini 2007, p. 81).

The psychological effects of biases can leave a person arguing for an unsubstantiated viewpoint long after a more reasonable viewpoint has been adopted by the majority. Cognitive biases can impede our ability to accept new ideas. Being aware of the risk of cognitive bias is the first step in mitigating that risk.

Table 10: Flaws in Cognitive Processing and the Consequences for Problem Solving

Cognitive Bias	Description	Innovation Consequences
Projection bias	Projection of past into future	Failure to generate novel ideas
Egocentric empathy gap	Projection of own preferences onto others	Failure generate value-creating ideas
Focusing illusion	Overemphasis on particular elements	Failure to generate a broad range of ideas
Hot/cold gap	Current state colours assessment of future state	Undervaluing or overvaluing ideas
Say/do gap	Inability to accurately describe own preferences	Inability to accurately articulate and assess future wants and needs
Planning fallacy	Overoptimism	Overcommitment to inferior ideas
Hypothesis confirmation bias	Look for confirmation of hypothesis	Disconfirming data missed
Endowment effect	Attachment to first solutions	Reduction in options considered
Availability bias	Preference for what can be easily imagined	Undervaluing of more novel ideas

Source: (Liedtka 2015, p. 930, Table 3)

However, being aware of cognitive biases is not enough to overcome their risk entirely. Overcoming biases requires constant vigilance and working with customers to align values and minimise Blind-Spots (Branson 2008; Hines, Holweg & Rich 2004).

> ‘Being inclined to think that you can avoid a bias because you are aware of it is a bias itself. Laurie Santos and Tamar Gendler called this notion the GI Joe fallacy after a cartoon that claimed that “knowing is half the battle”. To the extent that learning about biases is half the battle, this is because your knowledge will help you to recognise cognitive errors in *others*. This can give you the false impression that your thinking has improved while other people remain hopelessly misguided. This can contribute to something called naive realism, in which people believe only they see the world objectively, as it truly is, while others are ignorant, misinformed, or deluded. Instead of making us wise, knowledge of cognition vices may make it easier to simply dismiss people with whom we disagree when we should actually be paying attention to their arguments.’ (Howard 2018, p. 527)

1.4.7 Requirements Elicitation

A significant cause of project failure is mistakes during the Requirements Gathering stage (Whittaker 1999). The Johari window is a useful technique, however its simplicity means that additional approaches are required to better target the analysis.

Figure 33: Classic Waterfall Software Development Life Cycle

The requirements gathering stage is when the business analyst consults with stakeholders to identify the actual needs the project is meant to solve. Unfortunately stakeholders don't always know exactly what they need, nor everything that is possible (Pressman 1997). Part of the difficulty in requirements gathering, is stakeholders will often describe their needs in terms of the existing systems, which is a side effect of cognitive biases like Availability bias. In many cases people confuse the derived desire for a tool (used to satisfy a desire) with the intrinsic desire itself (Liedtka 2015).

For example, when the flexible mobility provided by a car allowed teenagers to satisfy the intrinsic desires of freedom (autonomy), friendships (belonging), and relationships (love) many developed a life-long mis-association conflating those desires with the desire for a car (Abbott 2009) – a fact that advertisers still exploit today. Now the internet based Mobility-as-a-Service – MaaS – is disrupting conventional wisdom by proving people desire mobility/access not vehicles (Crayen 2018; Li & Voege 2017). Online shopping and dating have also proven that people don't require mobility/access, if they can satisfice their desires online (Rosenfeld, Thomas & Hausen 2019).

The careful elicitation of requirements from all beneficiaries is needed to deliver a desirable public transport service.[12] Over time, as project management has professionalised, there has been more research into improved techniques for gathering requirements and specifying projects. One identified limitation that introduces risk into the requirements gathering phase, is that individuals and organisations frequently have knowledge gaps. Knowledge gaps can be classified as follows:

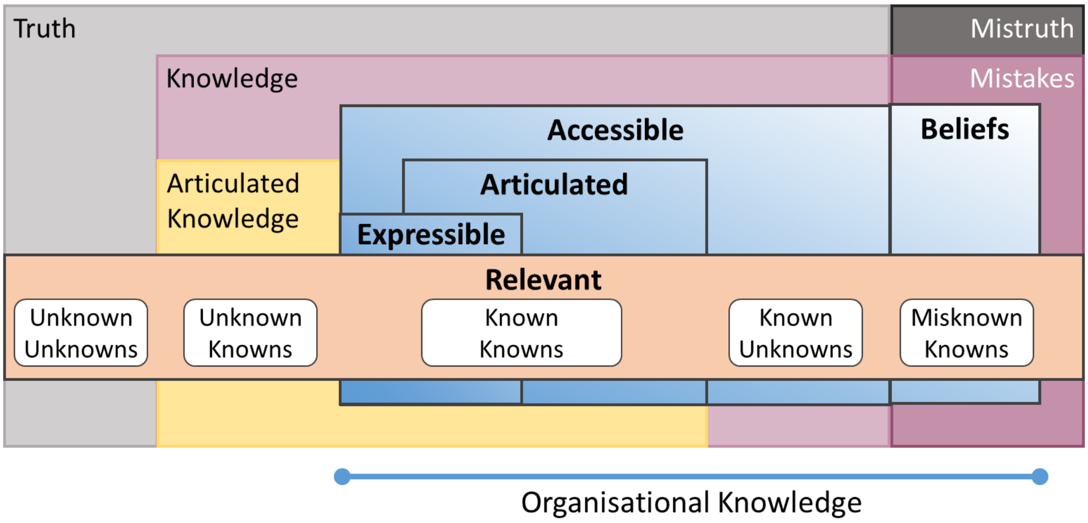
> 'Gacitua et al. [14] formalised the Rumsfeld taxonomy within a proposed Tacit Knowledge Framework using the properties of expressible, i.e. known knowledge; articulated, as documented known knowledge; accessible, which is known but not in the foreground of the stakeholder's mind and therefore a memory recall problem; and relevant to the project and domain. This produced definitions for:
>
> • **Known knowns**: expressible, articulated, and relevant.
> • **Known unknowns**: not expressible or articulated, but accessible & potentially relevant.
> • **Unknown knowns**: potentially accessible but not articulated.
> • **Unknown unknowns**: not expressible, articulated or accessible but still potentially relevant.'(Sutcliffe & Sawyer 2013)

[12] The business analyst's job is to find out what is in the stakeholder's Blind-Spots, Unknowns, and Hidden quadrants.

That framework has been translated into Figure 34 below. First there is the whole truth, shown as the bounding grey box. Humanity's knowledge exists as a subset of all truth (shown in pink). Articulated knowledge is all documented knowledge, which could in theory be accessed by the organisation – such as academic papers or German best-practice manuals (shown in yellow). The organisations knowledge is shown in blue and covers accessible knowledge (usually in the back of the minds of the staff), articulated knowledge (in the manuals) and expressible knowledge (in the fore-front of the minds of the staff). For delivering a project / transport-service, there is relevant knowledge (shown in the orange band) as it covers all other types of knowledge.

For a transport operator, the Known-Knowns is all the knowledge that the staff can immediately recall or quickly access in the documented procedures.

Figure 34: Elicitation Review Framework



Based On: (Based On Halpern 2009, p. 11)

The Known-Unknowns is what is known only to some staff, but also issues that staff know but don't really understand – often associated with little problems that go ignored.

The Unknown-Knowns are the Blind Spot of the operator, where they could expand their skills and knowledge by working with customers or industry. The unknown-knows are the areas where there is easy room for improvement.

In contrast, the Unknown-Unknowns are the areas where further research is required and is associated with problems that no-one in the world has solved, yet.

As a real-world example, senior staff within TfNSW were unfamiliar with exact relationships between fleet size, headway and runtime – see Appendix B.5 Transport Terminology. In that way, the relationship represents an Unknown-Known for senior staff in TfNSW. As stated previously, TfNSW is a diverse entity with senior staff from many disciplines who do not have direct experience in transit operations.

There is another very important type of knowledge that I have identified: Misknown-Knowns. This knowledge may have been carefully acquired and based on research, however, it may simply be incomplete, or worse incorrect. These Misknowns may be the result of cognitive biases unintentionally distorting analysis or research.

In some cases, Misknowns are side effects of an incorrect paradigm that has since been disproved, such as phlogiston chemical theory (Zeibots 2007). In other cases, they may be the side-effect of the research method, such as a stated preference survey over-inflating the cost of a transfer penalty (TfNSW 2001a, 2001b). Misknowns may be the failure to take into account an unconscious effect such as loss-aversion (Martin & Thornton 2017). Or, Misknowns may be the use of incorrect assumptions like linear elasticities or that the value of travel-time savings is continuous and not discrete (Carrel, Halvorsen & Walker 2013; Kitamura, Fujii & Pas 1997; Naznin, Currie & Logan 2016)

1.4.8 Putting it together

A model codifies the organisation's *interpretation* of the world within the context of their worldview or paradigm. As a simplification, a model will be built using the organisation's Known-Knowns and also its Misknown-Knowns. The model will represent the Open quadrants – the overlap between the customers' conscious knowledge and the operator's conscious knowledge, as well as the organisation's overlap with the industry's knowledge. Since the model is created by the organisation, it cannot represent the operator's Blind-Spots, including missing customer knowledge and missing industry's knowledge - i.e. the Unknown-Knowns. Also, a model will not represent the Unknown-Knowns that exist within the operator, such as the deep knowledge of passenger movements on platforms by station staff, and instead will rely on simplifications known to the modeller – like gravity models. No model can represent the Unknown-Unknowns, such as customers unconscious behaviour.

For these reasons, a model driven approach will miss many of the subtle nuances of the world in which the operations are undertaken. Models are a vital part of human thinking because all models represent an attempt to explain and codify the interpretive analysis being undertaken. However, that interpretive analysis will always represent our limited view into massive interrelated chaotic systems.

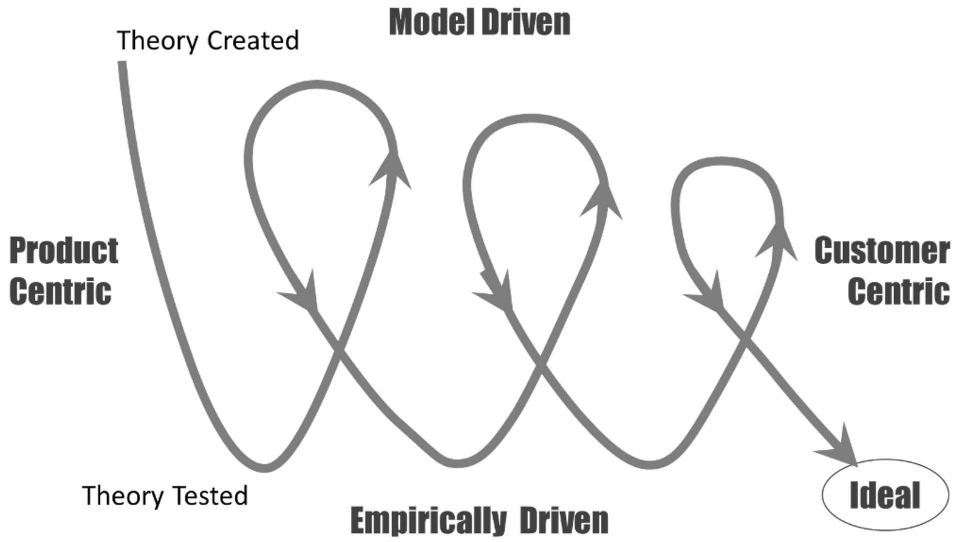
The move towards being empirically-driven can be seen as more ‘scientific’ within the terms of scientific falsificationism because any policy (or theory) for transport operations can be tested against reality to identify and eliminate what is not true (Bunge 2016; Chao, Chen & Millstein 2013; Thornton 2014).

An empirically-driven approach first accepts and embraces the limitation of models, persons and organisations. Thereafter the approach engages in a continuous iterative process of reassessing its models and processes. These iterations are used to improve the organisation, adopt industry best practice and realign it with customers. In the business literature that is described as a continuous-learning organisation (Branson 2008; Hines, Holweg & Rich 2004; Suárez-Barraza, Smith & Dahlgaard-Park 2012).

Figure 35 below shows the cycle of an organisation engaging in continuous learning as it transitions towards being an ideal transport operator through realignment, eliminate Misknown-Knowns, formalising Unknown-Knowns, and using research to reduce Unknown-Unknowns.

Returning to the model of the three stages of transformation from 1.2.2 Product-Centric Model- to Customer-Centric Empirically-Driven above, Stage ① of the empirically driven approach is to improve the completeness and the veracity of the operator’s measurement of their service delivery (α) and the customers response to that delivery (β). Stage ② improves the comprehensiveness and veracity of the operators understanding of their customers motivations and behaviours. The operator should continuously repeat the cycle, each time improving their empirical measurements before once again improving their customer centricity.

Figure 35: Ideal cycle of paradigm changes in service organisations



1.5 What do you prescribe?

An operator wants to become empirically-driven and has acquired high-resolution datasets measuring their service delivery and their customer's response. They are raring to go and ready to drive future service targets from empirical data. There is just one very large problem: *it is impossible to drive planning only from descriptive data.*

The following section describes why the operator's values will drive their service targeting, and why the customer's values will drive their perception of the services promised and delivered. When operators and customers do not share values, the operator cannot plan services the customers perceive as desirable. Values that are not shared are in the Hidden or Blind-Spot quadrants of a Johari diagram.

Referring to the Service Quality Loop, the service quality targeted is defined by the planners and includes parameters such as bus routes, hours-of-operation and vehicle headways. Engineers are responsible for delivering the service quality targeted, such as building stops, roadways, and programming signalised intersections. Planners determine what could be and the engineer determines how to achieve that.

The transport planning target for the IWLR optimisation programme was to reduce runtime by four minutes to save one vehicle for redeployment. The subsequent investigations on how to deliver reduced runtimes was transport engineering.

Planning and engineering are not just applications of science, they are an application of philosophy. Specifically, these professions are the assessment and evaluation of alternatives to select a preferred option. The option selected will be the one judged to best align with the values of the assessor. During project scoping and requirements gathering, the assessor should have aligned their values with the client and community.

> 'Ethical engineering practice requires judgment [sic], interpretation and balanced decision-making in context.' (EA 2018a)

Figure 36: Structure of the research programme - convergent vs divergent

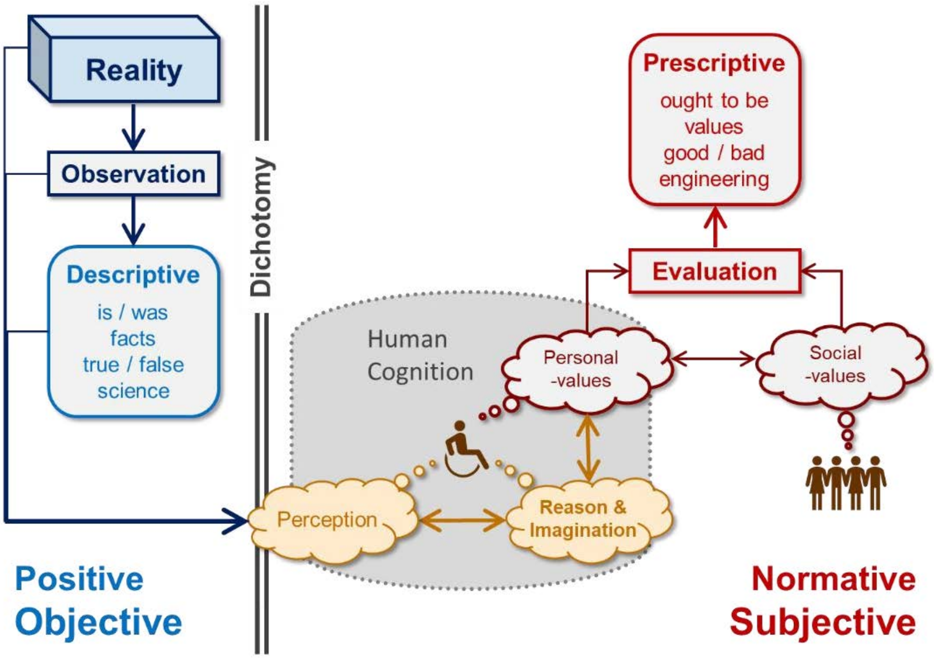
Source: Adapted from (Adapted from Hounsell 2018a)

1.5.1 The Is-Ought Dichotomy

One of the philosopher David Hume's great contributions to humanity was his is-ought dichotomy that describes how statements which prescribe what is good, right, and moral, cannot be logically derived solely from statements that describe reality (Blaug 1980; Cohon 2018; Fieser 2019; Moonan 1975; Morris & Brown 2019; Zeibots 2007). The dichotomy requires operators to achieve explicit values-alignment because the customer's values cannot be logically deduced purely from measurements of reality.

Assessments must be constructed with acknowledgement and reference to Hume's is-ought dichotomy explained there is no logical way to derive prescriptive statements of what ought-to-be (preferred state) from descriptive statements about what is and was (current and past states). Overcoming this problem is particularly difficult for engineers and planners as their work requires design trade-offs, which are expressions of what ought-to-be and thus cannot be logically derived solely from descriptions of what is. To fully exercise their judgement, professionals need to understand that their values are a construction of humanity and not derived solely from science (observations of reality).

Figure 37: Is–Ought Dichotomy as informed by perception and values



Hume's is-ought dichotomy and the difference between object and subjective reality are shown in Figure 36 above. There is an objective reality that exists outside of human cognition (top left). That objective reality can be observed using sensors in controlled experiments to develop descriptive statements that records the facts of what was. Each descriptive statement can be tested for its truth value – i.e. is it true or false. Those descriptions provide an objective record of reality and allow the development of positive (scientific) models of reality.

Humans are able to perceive the objective reality, to use experimental observations, and consume descriptive statements. In this way, humans are able to use their perception of reality to feed their cognition. Humans will use their perception to drive their reason and their imagination. However, a person's reasoning and imagination will also affect their perception. We can all think of a time when our imagination ran away with us and we mistook reality – such as seeing a spider in a shoe or a crocodile in a river.

Human cognition also allows us to construct personal-values. Our values reflect our goals and needs. A person's values will impact their imagination and thus their reasoning. Since we are social animals, our personal-values will be significantly influenced by our group's social-values (Victor & Cullen 2008).

Evaluation is a fundamental characteristic of intelligence. Any time an entity is required to rank alternative scenarios to select a preferred scenario, that entity is relying on its values to evaluate the alternatives (Bunge 2003, 2012). Values require cognition and the ability to use reason and imagination to assess the desirability of possible impacts from our actions. For example, a tourist at the Mary River in the Northern Territory on a 35°C day might evaluate the option of swimming in the river. The tourist would consider the relative values of a refreshing swim against their goals, which will include not being eaten by a two-metre long salt-water crocodile.

Using values and evaluation, an entity is able to derive prescriptive statements about what ought to be (what is good or desirable) and what ought not to be (what is bad or undesirable). In this way, an entity is able to develop a normative conceptual model that includes an evaluation framework.

The material objective reality does not have desires, only organisms emerging from the material reality have desires. Thus, the material reality cannot have values. Without universal values, it is impossible to construct a prescription of what ought to be from a description of what is – thus the Is-Ought Dichotomy.

Transport planners will use values to determine the balance of capital-cost, operating-cost, and service levels. Engineering is the evaluation of possible alternatives to assess the preferred alternative, thus how the various options are evaluated will be based on values including the trade-offs in the trinity of time, cost, and quality/capability. For example, an engineer will select the materials to use for a new tram bridge across the Parramatta River. They observe that previous bridges have been masonry arch, wooden, steel, and steel-reinforced-concrete. The engineer will have to evaluate each option based on cost of construction, lifetime, maintenance cost, and carrying capacity.

It is vital that during requirements gathering the partners determine the customers values or use requirements already known to the industry – such as the aforementioned ‘Seven desires for useful transit, and how transit serves them’ (Walker 2010).

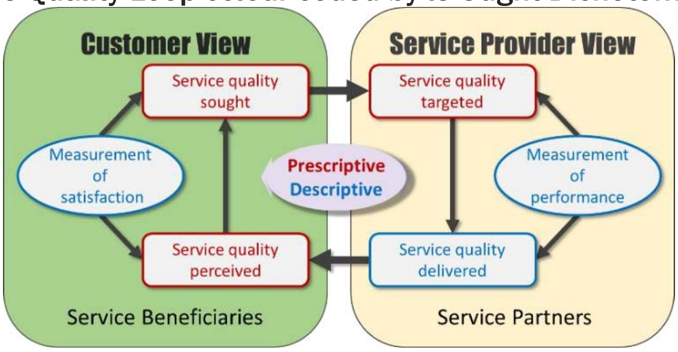
1.5.2 Service Quality Loop

Figure 38 below shows the elements of the Service Quality Loop colour-coded to indicate the is-ought dichotomy at the heart of successful transport operations. The service quality targeted will use the operator’s values-framework while the service quality sought and perceived will use the customer’s values-framework – therefore they are all prescriptive thought (red). None of these prescriptive elements can be logically derived solely from those elements measuring and describing reality (blue).

The service partners will have delivered an objective service in the real world by running trains or buses for example. The measurements of satisfaction and performance must be objective properties to allow their quantitative or qualitative description.

However, to evaluate if the service delivered was adequate, the providers and customers must reference their social, organisation, and personal values-frameworks. As such, it is impossible for the service provider to target a service quality matching that sought by the customer if they both have different (misaligned) values-frameworks.

Figure 38: Service Quality Loop colour coded by Is-Ought Dichotomy



For the values-frameworks to be aligned between the customer and operator, all the values must be Known-Knowns and shared by both parties. The service quality targeted (timetable) will be shared so that customers can access the services. Customers withholding their perception of the service would prevent improvement, therefore their feedback should be shared. It would be advisable for customers to have full visibility of the measurements used so they may provide feedback to identify Misknowns and prevent unintentional values misalignment.

Therefore, we can say that in order to fully implement the Service Quality Loop the operator and the customers need to work together to ensure both parties share all relevant knowledge, including values-frameworks. That all knowledge is shared and in the Open quadrant, ensures there are no Blind-Spots or Hidden areas that prevent relevant knowledge from being used during the development of performance measurements and service targets.

1.6 Summation

This chapter has shown the need for continuous empirical measurement in order to ensure that an organisations models are informed by accurate measurements of the transport network delivered and the customers responding to those services. Without accurate measurement it is possible for models to severely underestimate or overestimate the customers response to planned services.

Before the IWLR extension, two of the requirements (values) given to the engineers were a quick project delivery and the $120 million budget. Considering the patronage forecasts for 2026, the engineers decided that a single-track stub terminus would be sufficient to support a peak vehicle-headway of 8 minutes. The model indicated that terminus would allow effective operation for 10 years before patronage levels required more capacity. The final cost for the IWLR extension (third stage) was $176 million. However, the model was inaccurate and the patronage levels in 2016 reached the capacity limits of the system. TfNSW began investigating costly options to increase capacity, such as rebuilding the stub terminus with double track (Samra 2017).

The service partners had significant Blind-Spots, due to the limitations of their data gathering and the limitations of their model. They did not completely understand their customers, and thus, did not share their customers values. Due to incorrect assumptions, the department chose to construct lower-cost infrastructure which in practice did not meet the operational requirements for the actual level of passenger demand.

These forecasts assumed minimal induced patronage and did not assume passengers would trade pleasantness for speed on discretionary journeys. These factors represent Unknows-Knowns that were in the Blind-Spot of the model.

By moving towards empirical measurements and a customer-centric approach, it is possible for a transport operator to maintain values alignment with their customers, so that they may accurately understand and predict customer behaviours in response to the transport network they actually deliver. In addition, by understanding the customers, the transport operator will be able to accurately predict the future and correctly plan the necessary infrastructure to meet the coming demand. Finally, by staying empirically driven, the operator can accurately measure the customers changing responses over time and adapt services to meet customer needs while at the same time controlling costs.

CHAPTER 2 – PRIVACY

This chapter outlines the issues that ensure privacy protection is a primary requirement for analytics; then Chapter 3 outlines approaches to protect privacy during analyses.

Privacy is an interesting word; it is routinely used but very rarely defined. Louis Brandeis was one of the first to discuss the right to privacy, which he defined as a general ‘right to be let alone’ (Heawood 2014; Warren & Brandeis 1890). Merriam Webster defines privacy as – the quality or state of being free from observation or intrusion. To paraphrase an old axiom - we may not know what privacy is, but we know it when we have lost it.

Privacy is essential to the intrinsic needs of security and self-actualisation for humans (Maslow 1943; Warren & Brandeis 1890). Privacy is essential for individual mental health and the exercise of liberty; therefore, privacy is an intrinsic human right. Privacy enables democracy through freedom of expression, knowledge, and association. Since privacy is an intrinsic right, it cannot be surrendered, traded, or curtailed by contract, employment, venue entry, or regulation. Additional reading on privacy is important for professionals as ethics require us to challenge our assumptions and cognitive biases.

Privacy is usually defined in the negative, unfortunately that results in a perceived amorphous quality to privacy which allows frequent infringement. In some cases, privacy is defined only by the ways it can be infringed. For example, the Commonwealth of Australia Privacy Act 1998 does not define what privacy is, instead it defines privacy by specifying mechanisms for how it should be protected.

> CoA Privacy Act 1988 - Section 2A - ‘The objects of this Act are:
> (a) to promote the protection of the privacy of individuals;’

A product-centric transport operator may not deem privacy to be a central consideration; after all, privacy has nothing to do with the operation of trains or buses. However, since privacy is central to the wellbeing of humans, protecting privacy is essential to customer-centric operations. Without embedding privacy risk management in corporate culture, transport operators cannot be allowed access to large empirical datasets on customers, such as Opal. A lack of a privacy protection method prevents the shift of operators from a model driven to an empirically driven approach in service delivery.

Before operators can move to an empirically driven approach, the methodological issues of protecting individual privacy and safety must be realistically and comprehensively addressed. A privacy or safety violation can never be undone and the damage to an individual, or to the organisation's reputation, may be irreparable or unrecoverable. It is impossible to retrospectively add a requirement for privacy protection after the requirements gathering stage (Petersen Peter 1999; Whittaker 1999). Therefore, privacy protection must be a primary requirement before the design of new analytics methods and operational approaches, such as those proposed by this thesis in Chapter 4.

Historically the labour and space required to efficiently store information, as well as the slow search times, ensured that only the most well-funded 'security' agencies were capable of storing significant volumes of data. That all changed as Moore's law and economies of scale enabled the fourth Industrial Revolution, facilitating the widespread deployment of Information Systems throughout the world. This industrial revolution is why in 1980, the OECD identified that Privacy Protection would soon become a matter of significant and international concern (OECD 1980; OECD 2013).

Transport operators relying on surveys could never obtain data their customers wish to keep Hidden and were able to protect the privacy of survey responses using established confidentiality practices. However, ubiquitous sensors and TOTOR ETS create a situation in which the privacy of all customers could be violated irreparably without a carefully considered approach with privacy protection as a primary requirement.

The NSW Opal ETS creates a census of the city's populations travel behaviour, with TOTOR datasets providing an excellent source of empirical data on transport operations and customer response. Before this empirical dataset can be used however the risk of privacy breaches must be carefully controlled.

During the start of the research programme, operators and bureaucrats expressed significant concerns that Big Data analysis could be misused to compromise individual safety. Halfway through the programme a significant test case was decided in the NSW Civil And Administrative Tribunal – Waters v Transport for NSW [2018] NSWCATAD 40 (15 February 2018) (NSWCATAD 2018). In this case a person's travel histories were found to be biographical and thus personal information and due to certain practices

TfNSW was found to have breached NSW privacy law. TfNSW has been reticent about any persons using the Opal datasets since the finding against them.

Public Transport Victoria was less cautious and released a transaction dataset to a “hackathon”. They were then condemned in the media for breaching the privacy of Victorians, for releasing individual travel histories (Gearin 2019; Taylor 2019)

In response to these events, as a reasonable institutional risk control response, the operators have locked down their datasets. However, this is preventing the realisation of opportunities to improve services by using TOTOR datasets to empirically measure transport operations and the subsequent customer response.

On the other hand, the public have a Right-To-Know what their bureaucracies are doing and how their money is being spent. Departments cannot justify withhold information if there are ways to release the information while protecting privacy— NSW GIPA Act.

In addition, the release of privacy-safe empirical data to the public allows third parties (such as universities) to undertake research at their own cost. Third-party research will be conducted in areas that represent the operator’s Blind-Spots, in quantifying the Known-Unknowns, or discovering the Unknown-Unknowns. Publication of third-party research gives operators free access to more and broader analysis. A wide range of published empirical analysis from differing perspectives also allows thorough analysis of the transport network delivered and the resulting customers response.

This thesis uses the Information Protection Principles (IPP) from OECD, CoA, and NSW. IPP are based upon universal human needs. Therefore, the principles of privacy protection used in this thesis, the methodological concerns, and methods proposed would be applicable to NSW and other jurisdictions.

2.1 Operating ethically as an Engineer

It is important for responsible professionals to embrace a culture of privacy to ensure their work can never be used by individuals or the state to violate the innate dignity of individuals by infringing the right of everyone to live unobserved and unmolested.

In the Appendix are extracts from Engineers Australia General Regulations – Schedule 1 – Code of Ethics and Guidelines (EA 2018a, pp. 88-91) and (EA 2018b) . These are provided as clear examples of how the institutional body of engineers has identified the need for practitioners to protect the community during the execution of their duties.

An engineer acting ethically must '1.3 respect the dignity of all persons' and '3.2 support and encourage diversity'. This chapter explains why protecting privacy is essential to protecting the dignity of individuals and especially minorities, which is essential to supporting diversity. An engineer must act ethically to act professionally. Risk Management is a significant part of ethical action and is explained next.

2.2 Risk Management Guidelines

Communities, organisations, and individuals 'face external and internal factors and influences that make it uncertain whether they will achieve their objectives.' – that is they face risks (ISO & AS 2018). The modern approach is to build Risk Management process within an organisation to modify its risk profile and ensure opportunities are not forgone in favour of the fantasy of forgoing risk.

> '… managing risk is … a process of optimization that makes the achievement of objectives more likely. Risk treatment is then concerned with changing the magnitude and likelihood of consequences, both positive and negative, to achieve a net increase in benefit.' … 'In the past, it has been common for risk to be regarded solely as a negative concept that organizations should try to avoid or transfer to others. However, it is now widely understood that risk is simply a fact of life and is neither inherently good nor inherently bad. To avoid [risk] entirely is to forgo the opportunity of pursuing objectives. If we can successfully detect and understand risk, including how it is caused and influenced, we can, if necessary, change it so that we are more likely to achieve our objectives and might even do this faster, more efficiently, and with improved results' (Purdy 2010, p. 882).

Consider the Service Quality Loop (1.5.2) coded as positive and normative components, the objectives of community, organisations, and individuals are expressed in the service quality sought and the service quality targeted. An implicit objective in the Service Quality Loop is expressed in the Engineers Australia ethics guidelines as '*4.2 practise engineering to foster the health, safety and wellbeing of the community*'. To protect the safety and wellbeing of the community, the service partners must minimise the risks created by their actions, and that includes minimising the risks of violating privacy.

2.2.1 Terminology

In 1999, the Risk Management subcommittee of Standards Australia released AS/NZS 4360:1999 – Risk Management Guidelines.[13] These guidelines outline processes of continual organisational engagement in risk management. Defining common terminology harmonises interpretations and facilitates professions working across organisations and stakeholders in collaborating on risk management (Purdy 2010).

Table 11: ISO Risk Management Guidelines – 2018 Terms & Definitions

3.1	**risk**	Effect of uncertainty on objectives. *†
3.2	**risk management**	Coordinated activities to direct and control an organization with regard to **risk**.
3.3	**stakeholder** / interested party	Person or organization that can affect, be affected by, or perceive themselves to be affected by a decision or activity.
3.4	**risk source**	Element which alone or in combination has the potential to give rise to **risk.**
3.5	**event**	Occurrence or change of a particular set of circumstances. ‡
3.6	**consequence**	Outcome of an **event** affecting objectives. §
3.7	**likelihood**	Chance of something happening. |
3.8	**control**	Measure that maintains and/or modifies **risk**. ¶

Source: (ISO & AS 2018, p. 1)

> * An effect is a deviation from the expected. It can he positive, negative or both, and can address, create or result in opportunities and threats. Objectives can have different aspects and categories, and can be applied at different levels.
>
> ‡ An event can have one or more occurrences and can have several causes and several consequences. … something that is expected which does not happen, or something that is not expected which does happen. An event can be a risk source.
>
> § A consequence can be certain or uncertain and can have positive or negative direct or indirect effects on objectives. Consequences can be expressed qualitatively or quantitatively. Any consequence can escalate through cascading and cumulative effects.
>
> | In risk management terminology, the word “likelihood” is used to refer to the chance of something happening, whether defined, measured or determined objectively or subjectively, qualitatively or quantitatively, and described using general terms or mathematically (such as a probability or a frequency over a given time period). The English term “likelihood” does not have a direct equivalent in some languages; instead, the equivalent of the term “probability” is often used. However, in English, “probability” is often narrowly interpreted as a mathematical term. Therefore, in risk management terminology, “likelihood” is used with the intent that it should have the same broad interpretation as the term “probability” has in many languages other than English.

[13] Standard 4360 was further updated with a release in 2004. In 2009 an ISO subcommittee leveraged that work to release Risk Management guidelines as ISO 31000:2009. These guidelines were further updated by an ISO subcommittee in 2018 to create AS ISO 31000:2018 (ISO & AS 2018; Purdy 2010).

¶ Controls include, but are not limited to, any process, policy, device, practice, or other conditions and/or actions which maintain and/or modify risk. Controls may not always exert the intended or assumed modifying effect.

It is useful to remember this particular note as a shorthand for risk analysis:

† **Risk** is expressed in terms of **risk sources**, potential **events**, their **consequences** and their **likelihood**.

The following table has useful terms from the 2004 Australian & New Zealand Standard that were not carried over to the international standard.

Table 12: AS/NZS – Risk Management – Guidelines – 2004 Terms and Definitions

1.3.6	hazard	A source of potential harm
1.3.8	loss	Any negative **consequence** or adverse effect, financial or otherwise. *
1.3.9	monitor	To check, supervise, observe critically or measure the progress of an activity, action or system on a regular basis in order to identify change from the performance level required or expected
1.3.10	organisation	Group of people and facilities with an arrangement of responsibilities, authorities and relationships. † EXAMPLE: Includes company, corporation, firm, enterprise, institution, charity, sole trader, association, or parts or combination thereof.
1.3.12	residual risk	**Risk** remaining after implementation of **risk treatment**

Source: (AS & NZS 2004)

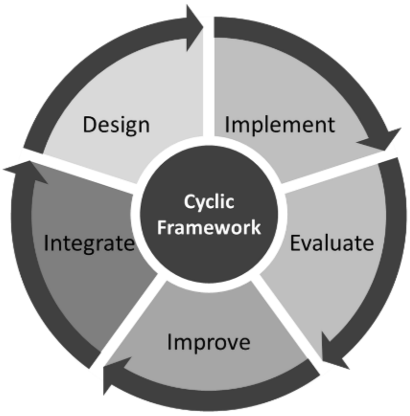
* This term has been replaced with negative consequence.
† An organization can be public or private.

2.2.2 Cyclic Processes of Continuous Risk Management

The standard defines several cycles that can be generalised in the framework shown in Figure 39 below. The standard acknowledges that the world is constantly changing, and risk management is a continuous process requiring continuous monitoring & adaptation.

The cycle begins with a design phase that consists of generating options, evaluating options, selecting a preferred option, then preparing an implementation plan. The second phase is a trial implementation and the third is evaluation. The fourth is improvement and the fifth is full integration into the organisation’s systems and operations.

Figure 39: Cyclic framework



Source: (ISO & AS 2018, p. 4, Figure 3)

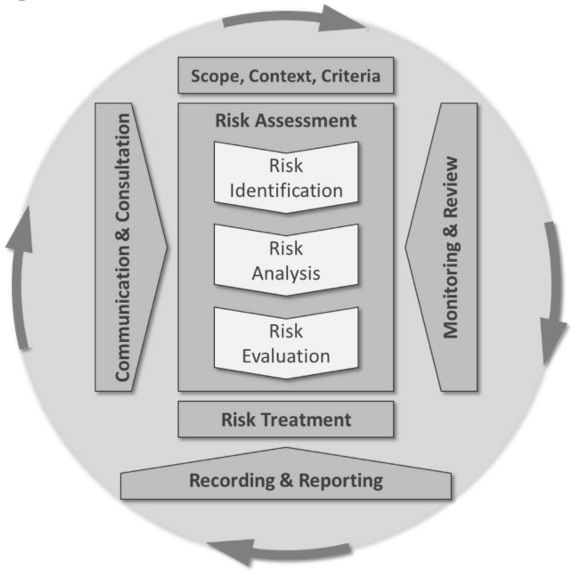
Due to the evaluation and selection of options, Risk Management is a prescriptive task. An evaluation method to assess the success of the risk controls is developed during the design phase. Thus, all Risk Management strategies and plans are value-laden and reflect the personal-, social-, and organisational-values of the risk management team.

2.2.3 Risk Management Process

The first step in a standard risk management cycle is to identify the scope of the plan. Risk is the 'effect of uncertainty on objectives' (Purdy 2010, p. 882) and since objectives drive all levels of an organisation, risk occurs and requires management at all levels, including the strategic, operational, programme, project, or activity levels. The second step is to describe the internal and external context. The third step defines the criteria used to determine the level of acceptable risk for the organisation. The fourth step is a complete risk assessment which includes the development of a risk management plan. The management plan outlines the fifth step which defines the controls & treatments designed to minimise either the likelihood of an event or the negative consequences if the event does occur.

In addition, a risk management cycle will have continuous subprocesses to facilitate consultation and subsequent communication of the strategy, to facilitate the monitoring of events, and to review the strategy, as well as record and report what has occurred for use during future iterations of the risk management cycle. Note that these continuous sub processes are used during all phases of the cyclic framework but especially in the design, evaluation, and improvement phases.

Figure 40: Risk Management – Guidelines – Process



Source: (ISO & AS 2018, p. 9, Figure 4)

2.2.3.1 Risk Analysis

The core of risk assessment has three subsections. First, during risk identification, the risks that might impede the organisation from taking advantage of opportunities and achieving its objectives are identified. There are a range of techniques to identify the uncertainties that may impede an organisation. Organisations should ensure they identify the risks from sources outside their control, as well as events with multiple consequences, and also identify intangible consequences such as reputational damage. ISO 31000:2018 identifies the following factors to consider (ISO & AS 2018, p. 1):

- tangible and intangible [risk sources]
- causes and events
- threats and opportunities
- vulnerabilities and capabilities
- changes in the external and Internal context
- indicators of emerging risks
- the nature and value of assets and resources
- consequences and their impact on objectives
- limitations of knowledge and reliability of information
- time related factors
- biases, assumptions and beliefs of those involved.

The second risk assessment step is explaining the risks in detail. The level of detail for each analysis is dependent on the organisation and the character of the risk. The degree of analysis will be dependent on the level of planning, the possible consequences (especially losses & human hazards), and the likelihood of occurrence. Consequences may act as events for risks, triggering a cascade of additional consequences.

Special care must be taken when analysing highly uncertain events and to control for personal bias. Quantifying some consequences may not be possible or informative, so an analysis can be purely qualitative, like a list of consequences.

Risk evaluation is the final step in risk analysis, where for each risk the organisation determines if the risk needs further action. The organisation could maintain existing controls, undertake further analysis, or consider additional risk treatments. Risk controls & treatment have costs, so it is reasonable for an organisation to take no action, especially for unlikely risks or those with minimal consequences. If a risk is likely and high impact it is prudent for the organisation to reconsider and adapt its objectives.

2.2.3.2 Risk treatment

Risk treatment is a continuous iterative process of selecting options to address risk in a cyclic framework. The process is repeated until the residual risk is acceptable.

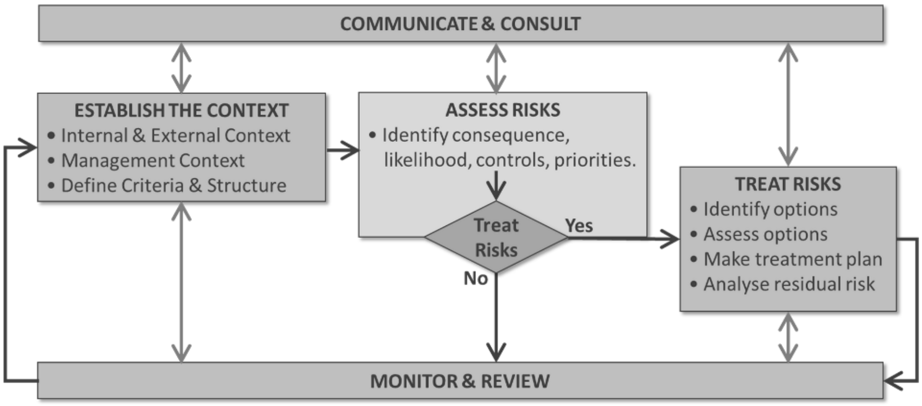
> ‘ISO 31000:2009 gives a set of general options to be considered when risk is treated. The order of the list reflects preference. … The options are:
> a) Avoiding the risk by deciding not to start or continue with the activity that gives rise to the risk;
> b) Taking or increasing the risk in order to pursue an opportunity;
> c) Removing the risk source;
> d) Changing the likelihood;
> e) Changing the consequences;
> f) Sharing the risk with another party or parties;
> g) Retaining the risk by informed decision.’ (Purdy 2010, p. 884)

Risk treatments will give rise to additional consequences, and may create new risks, as such, the impact of treatments need to be monitored and addressed. It is important to remember that a decision to forgo an activity because of risk is a decision to forgo an objective and its reward in order to avoid the risk.

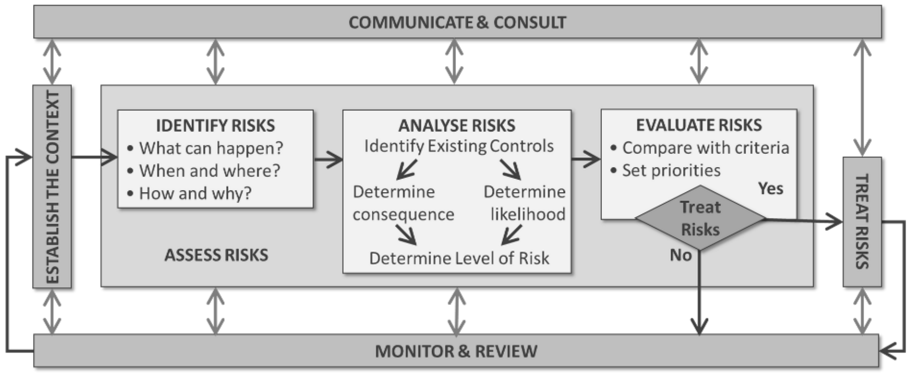
2.2.3.3 Risk Management and Opal datasets

When TfNSW began discussions with the UTS regarding the usage of Opal datasets, we all began an informal risk management cycle that was informed by the formal risk management culture. The informal risk management leveraged treatments and controls that had already been implemented by Cubic (the Opal system’s operator) and TfNSW.

Figure 41: Risk management process



Assess Risk Subprocess



Source: (AS & NZS 2004, p. 13, Figure 3.1)

The Opal system allows an adult to associate an Opal card with their name and credit card to facilitate automatic billing, ensuring they always have sufficient funds to buy a ticket. To obtain the Seniors Concession card, to which they are legally entitled, TfNSW forces seniors to associate their name with a card. As an example of existing risk controls, Cubic stores the names and card numbers in a separate database and does not store the card numbers with the trip records. As such, trip records are extracted without any personal identifiers such as name or credit card number. This risk control ensures that analysis of the trip records will never include personal identifiers.

It must be noted that the NSW Government allows members of the NSW Police Force to access the Opal records without a warrant, including searches based on card number or personal identifiers (Knaus 2016).

Due to the anonymisation of the data by Cubic, there are minimal direct hazards for individuals from the use of the data; see below. The primary losses would be reputational for TfNSW and UTS – these consequences would most likely be actualised by the public and media misunderstanding the actual risk to individuals.

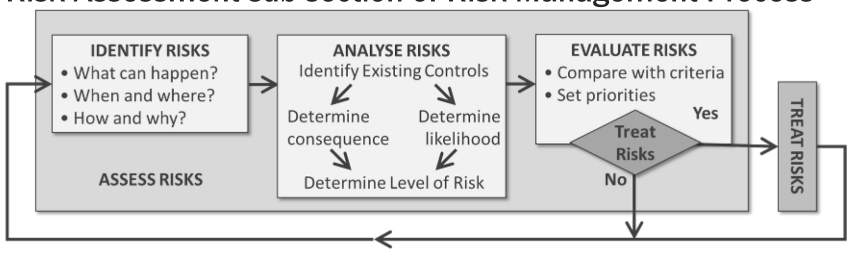
2.3 Safety of Individuals

The following sections identify the negative-consequences (hazards) possible from the misuse of a TOTOR dataset. In other words, why people are legitimately concerned about being observed. First the perception of direct threats by individuals is discussed, then, in the next section, the structural threats from society are examined.

Stakeholders are concerned about crimes being assisted by TOTOR datasets or their derivatives. This section, examining the hazards of theft and sexual violence, is included because these particular hazards were raised repeatedly by stakeholders (including within TfNSW) and demonstrate a significant perceived hazard.

Stakeholders assumed that using these brand-new Opal datasets presented a significant risk of hazards to individual safety. It was assumed that travel histories in the hands of stalkers or criminals could imperil individual safety. The second step of risk assessment is to analyse the risk thoroughly and empirically without assumptions.

Figure 42: Risk Assessment sub section of Risk Management Process



Source: (AS & NZS 2004, p. 13, Figure 3.1)

2.3.1 Steal from person

The stakeholders identified an event which is categorised in the crime statistics (in decreasing order of occurrence) as “Steal from person”, “Robbery without a weapon”, “Robbery with a weapon not a firearm” and “Robbery with a firearm”. The thesis will now discuss the risk source, consequence, and likelihood.

The consequence for individuals is loss of assets, personal belongings, and a degree of trauma depending on the details of the event. The consequence for the service partners is reputational damage, with the potential for reduced patronage and revenue if services become perceived as unsafe. The risk source was identified as – Big Data analysis demonstrating patterns of behaviour. What is the likelihood of this event, and does the dataset make that higher?

Stakeholders often cited a hypothetical mugger identifying a person regularly travelling alone at a given time. First, the majority of stealing offences are from a property – with the largest target being shops. Shoplifting is less effort and risk with higher rewards than stealing from a person. That is probably why the rate of stealing from a shop is 6.5 times higher than from a person, and 30 times higher than robbery with a weapon (BCSR 2019). In addition, it is perhaps counter-intuitive but bag-snatching, muggings, and other forms of stealing from a person are more likely to occur at locations that are busy. This is further complicated by less people carrying cash. The Opal dataset does not have any record of income to help thieves identify targets.

Thieves do not waste their time hanging around in deserted alleys where there are no victims. Instead thieves tend to prefer areas with a steady supply of people, but with an irregular flow, which allows them to quickly isolate a victim. Thieves will generally avoid constant crowds because they do not wish to be detected or recognised. Busy Streets theory is based on the principle of passive surveillance — a street with enough engaged persons increases the likelihood of detection and identification, as such thieves will go elsewhere (Devlin 2019; Heinze et al. 2018; Salmelainen 1992).

In other words, Busy Streets change the risk profile of crime. The physical layout of a location is also central to its ability to host crime without detection – that truth is the central reason for approaches such as Designing Out Crime (LAPD 2019). That is why, going to a location to assess its risk profile for crime is more productive for a thief than undertake a Big Data behavioural analysis.

2.3.2 Sexual violence

Another stakeholder concern was violent attacks especially sexual assault. For this identified hazard, the significant risk factors here are societal and environmental. For this risk analysis, the event is sexual assault and the risk source is Big Data analysis demonstrating patterns of behaviour. The consequences for the service partners are the same. The consequences for the victim are severe trauma and lifetime emotional scars. What is the likelihood of this event, and does the dataset make that higher?

The rate of reported Sexual Assault in NSW for FY 2018-19 was 73.5 per 100,000 persons. With indecent assault being 103.1 per 100,000 persons. The analysis of this hazard indicates that the most likely places for sexual assault where at home by a person known to the victim (see below):

> '■ The vast majority of victims of sexual violence are female and most perpetrators are male.
> ■ In most cases of sexual assault, the perpetrator is someone the victim knows, and perhaps knows well, such as a current or former intimate partner, or a relative.
> ■ Sexual assault is an aggressive act motivated by power & control.' (WHO 2004)

The dataset could be used to identify a person regularly travelling alone at a given time. Females (all ages), male children, and male youths make up the largest proportion of victims of sexual violence. Opal datasets do not contain information on gender, but card type is a proxy for age. The likelihood is highly situational so the actual numbers on sexual assault require further analysis.

(Cook, David & Grant 2001) noted that 47% of sexual assaults were by a family member or a person known to the victim and 16.8% were by a stranger. In the 2001 survey 36.2% of respondents did not state their relationship to the perpetrator. Since 2010, the ABS have published yearly a comprehensive survey of recorded crimes.

Table 13: Relationship of Sexual Assault Victim to Offender, 2018

Sexual Assault Relationship to the Victim	2018 Victims	Proportion (%)
Known to victim	7,688	75.1%
Stranger	1,195	11.7%
Total	10,241	

Source: • (ABS 2019, Table 11) – 45100DO003_2018 Recorded Crime – Victims, Australia, 2018 – Table 11 VICTIMS, Relationship of offender to victim by offence, Selected states and territories, 2018

There has been a reduction in the proportion of recorded Sexual Assaults by Strangers in NSW between 2010 and 2018, although the number of victims has remained steady at 1,200 indicating a reduction in the victimisation rate. Unfortunately, for the 26,000 persons reporting a sexual assault in NSW in 2018, the total number of assaults has increased and the rate of victimisation has also increased (BCSR 2019). Further research is required; but the data indicates increased reporting of sexual assaults, especially when the assaults were perpetrated by person's known to the victim.

The trend in the proportion of reported sexual assaults in community locations has fallen over the last two decades from 1997 to 2018. Further research is required, but this may reflect increasing awareness, ubiquitous recorded video surveillance, or increased reporting of domestic assaults (ABS 2019; Cook, David & Grant 2001).

Table 14: Sexual Assault Victims, location where offence occurred, Australia 2018

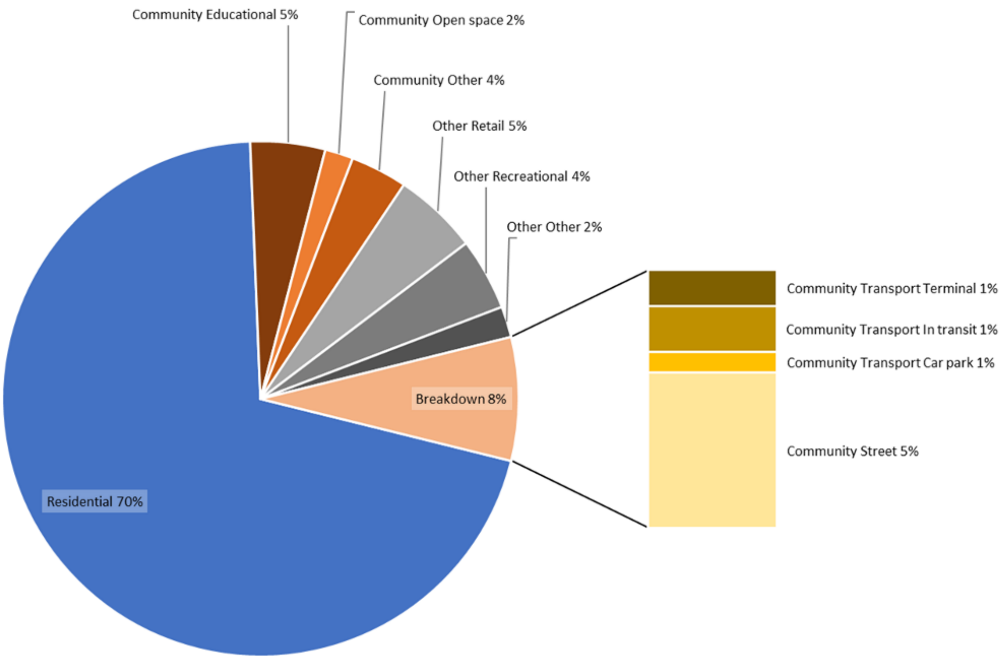
Sexual Assault Location	2010	%	2018	%
Residential	12,532	66.4%	17,006	64.6%
Community	3,574	18.9%	4,382	16.7%
Other location	1,768	9.4%	2,919	11.1%

Source: (ABS 2019`Table 3)

The proportion of sexual assaults on the street fell from 6.5% in 2010 to 4.3% in 2018, open space assaults fell from 3.1% to 1.6%, while transport assaults fell from 3.3% to 2.9%. The rate of sexual assault in transit conveyance or at a terminal was 2.3 in a year per 100,000 in population. There were 595 assaults on transit in Australia in 2018. In the FY 2017-18 the transit patronage in the five large states and ACT was 17.134 billion. Therefore, the rate of sexual assault per 100,000 public transport embarkations was 0.0347, which translates to a tiny rate of 3.47e-07 per embarkation.

That is not to diminish the severity for those 595 victims of sexual assault, nor is it meant to suggest that there should be no risk treatment. However, in order to assess the actual risk, the actual likelihood of the event must be assessed.

Figure 43: Sexual Assault Victims in CoA, location where offence occurred, 2018



Source: • (ABS 2019, Table 3) – 45100DO001_2018 Recorded Crime – Victims, Australia, 2018 – Table 3 VICTIMS, Location where offence occurred by selected offences, 2018

Table 15: Sexual Assault Victims, location where offence occurred, Australia 2018

Sexual Assault Location	Number	Proportion	Rate per 100,000

Residential			17,006	64.6	68.0
Community	Educational		1,130	4.3	4.5
	Transport	Terminal	262	1.0	1.0
		In transit	333	1.3	1.3
		Car park	149	0.6	0.6
		Other	11	0.0	0.0
	Open space		425	1.6	1.7
	Street		1,123	4.3	4.5
	Other		860	3.3	3.4
Other	Retail		1,277	4.9	5.1
	Recreational		1,092	4.2	4.4
	Other		467	1.1	1.9
Total			26,312		105.2

Source: • (ABS 2019, Table 3)– 45100DO001_2018 Recorded Crime – Victims, Australia, 2018 – Table 3 VICTIMS, Location where offence occurred by selected offences, 2018

People are still most likely to be murdered, raped, or assaulted by a person they know than by a stranger, and the proportion of reported crimes by strangers has been reducing in NSW since the start of this survey in 2010. "Appendix D.3 – The rates of crimes in NSW" lists the breakdown of crime rates per 100,000 persons in NSW for FY 2018-19. In NSW people are far more likely to be a victim of fraud than theft, assault, or sexual assault, and that excludes the modern plagues of bank-fraud and wage-fraud.

However, intimidation, stalking and harassment in NSW is reported at a rate of 24.8 per 100,000 persons, and has trended up over the last 5 years. If the victim is known to the person then they are at some risk from access to the PAS database or the Opal datasets.

2.3.3 Comparative likelihood of hazards

> 'There is now ample evidence that people overweight low probabilities, and this explains the widespread desire to gamble on low-probability events (e.g. lottery tickets) and to insure against low-probability catastrophes. … consumers' decisions about insurance are impacted by distortions in their perceptions of risk and by alternative framing of premiums and benefits. … When subjects were first asked how much they would pay for insurance against any disease and then any accident (thus isolating vivid causes), the price reported was more than twice that reported … for protection for any reason.' (Metcalfe & Dolan 2012, p. 510)

From the data on crime in NSW above and in the references, the actual likelihood of being a victim has decreased over time. To assess the actual likelihood, it is necessary to assess the likelihood of an event relative to the likelihood of another comparable event.

Comparing likelihoods is essential to comprehend the quantitative analysis. However, humans are notoriously bad at assessing relative probabilities. Metcalfe & Dolan (2012) notes humans show more concern when discussing a detailed event (e.g. a particular mugging scenario) than when discussing a general risk (e.g. rates of theft in society).

In 2017, there was 509 deaths per 100,000 persons in NSW. The top three causes were malignant neoplasms (cancers) at 147, circulatory diseases at 134, and respiratory diseases at 51 deaths per 100,000 persons. Injuries and poisoning were the fourth highest cause with 34 deaths per 100,000. Fatalities are not a comparable statistic as few crimes cause fatalities (only 1.1 crimes per 100,000 are murder or manslaughter).

However, hospitalisation is a severe event that can precede lifelong consequences – what does not kill you is actually more likely to leave you far weaker than before. So, hospitalisation is a useful comparison to interpret the likelihood of an event.

Table 16 below lists the leading causes for hospitalisation from injury or poisoning in NSW. In NSW, 348 persons per 100,000 population were hospitalised from injuries caused from using motor vehicles (cars etc) during FY 2017-18. In total 27,545 persons were hospitalised during just one financial year by motor vehicles.

Table 16: Leading causes of injury or poisoning hospitalisations, NSW FY 2017-18

Cause of Hospitalisation	Persons	Rate per 100,000 population
Fall	168,486	1,803
Exposure to unspecified factor	34,038	420
Other injury/poisoning	28,667	365
Motor vehicle transport	27,545	348
Cut/Pierce (unintentional)	21,537	275
Struck by/against (unintentional)	19,829	252
Self-harm	15,712	208
Natural/environmental factors	12,297	153
Interpersonal violence	9,477	124
Overexertion/repetitive movement	8,073	101
Poisoning (unintentional)	7,051	85
Fire/Burns	4,100	52
Machinery	3,142	40
Threats to breathing (unintentional)	1,571	17
Water transport (excluding drowning)	625	8
Drowning	472	6
Rail transport	132	2
Air transport	75	1
Firearm (unintentional)	68	1
Total	362,897	4,258

Source: • (NSW 2019b)

There is a higher rate of harm from unintentional poisoning (85) than all sexual assaults (74) in NSW. The rate of sexual assault on transit in Australia (2.38 per 100,000) is roughly similar to the rate of hospitalisation caused by railways themselves.

2.3.4 What should be done?

'Risk is … expressed in terms of risk sources, potential events, their consequences and their likelihood' (ISO & AS 2018, p. 1). The central question in any risk management plan is: do the planners value the reduction in likelihood or reduction in the severity of the consequence over the cost / disbenefit of the additional risk controls / treatments?

The likelihood of direct threat to individuals from using the Opal dataset is quite low. The removal of biographical data and keeping unusual patterns private will further control the risk to protect individuals.

However, that is not how bureaucracies perceive risk. If one person is robbed, or worse sexually assaulted due to the use of the Opal dataset, that would result in damage to the reputation of the department and more importantly to the government. Damaging the reputation of the government could cost them an election, and in modern Australian bureaucracies, such damage could result in the end of a career for the responsible public servants. Therefore, those undertaking risk planning for Opal datasets will identify a significant personal risk from the dataset to the risk management team.

In comparison, there will be no damage to the reputation of the government if the dataset is not used. The status-quo public transport will serve customers, even if the service partners operate the public transport network from a model-driven product-centric approach rather than an empirically-driven customer-centric approach.

The state may be unable to achieve their objective of optimisation of the public transport network without using the dataset, but the consequences of that are unnecessary operating expenses, unnecessary pollution, reduced international competitiveness, and triggering feedback loops that result in catastrophic climate change. When their career or reputation could be destroyed by the risk, people may be happy 'to forgo the opportunity of pursuing objectives' (Purdy 2010, p. 882). Therein lies the problem, in Australia careers are never destroyed by maintaining the status-quo.

Thus, the question becomes how to control and treat the risk to individuals from ‘analytics with malevolent intent’ in order to control and treat the risk of reputation damage. Chapter 3 - Privacy Protection and Chapter 4 - Data Preparation discuss the risk control framework developed by this research.

2.4 Systemic Risks

The hazards from privacy violation are not just the direct consequences from the action of individuals. NSW specifically identifies ‘*information about ethnic or racial origin, political opinions, religious or philosophical beliefs, sexual activities or trade union membership*’ in its privacy legislation because such information has historically proven a source or risk to individuals. History shows that some of the most significant hazards to individuals come from other groups of people or from the state itself. This section describes those systemic risks.

2.4.1 Social Norms

It is incorrect to assume that what seems good, right, virtuous, proper, and moral always has been and always will be unchanged. The temptation to think that your current worldview is right is simply the cognitive bias of ethnocentrism (Liedtka 2015).

Sometimes the most radical insurrectionist act is actually quite a simple thing — rejecting a social-norm. The clearest example in transport is Rosa Parks rejecting the bus driver’s order to relinquish her seat in the ‘coloured section’ to a white passenger and the subsequent Montgomery bus boycott that initiated, (Wolf & Zuckerman 2012).

Laws that are presented as being about ‘protecting morality’ or preserving ‘national-security’ may in reality be designed to harm a minority — those who are deviant. Recently, the threat of terrorism has been used to justify bans of burqas in public places, such as public transport. The ban in France was widely criticised as being designed to harass a religious minority. French politicians clearly understood banning the burqa would force those Muslim women who considered the clothing essential to forgo public transport rather than be immodest (Diallo 2020; Quinn 2016). CIA World Factbook notes France has a Christian Majority (63-66%), the next largest group is No Religion (23-28%) followed in a distant third by a Muslim minority (7-9%), (CIA 2019).

A key reason for bans on culturally, ethnically, or religiously significant clothing (such as headscarves) is the ability of those bans to effectively ban people of certain minorities from accessing services and participating in society. Humanity has a long history of banning significant ethnic or religious practices to persecute minorities.

2.4.2 Social Values

Universal-values exist across many organisms, such as the desire for energy. Personal-values are purely internal and often subjective — such as favourite colour. Social-values are those values shared by most members of a social group. Groups with shared values have stronger intra-group bonds and thus retain group coherence.

Social-values are shared through processes such as socialisation. For example, traffic engineers are indoctrinated on the primacy of vehicle flow rates during their education.

> 'Social psychologists have found that value sharing is critical in the emergence and maintenance of social groups, families to gangs to political factions.' (Bunge 2016, p. 383)

A social-norm is the expectation of compliance with a certain social-value. Since social groups are dynamic combinations composed from weak interpersonal bonds, conflict can create disequilibrium which can result in the group losing coherence. Long lasting social groups tend to consciously and unconsciously limit conflict by requiring conformance to social-norms. Non-conforming behaviours — i.e. deviance — can result in disagreement and thus conflict between individuals, or between individuals and the group. Deviance can be behavioural, such as wearing inappropriate clothing, or existential such as being a sexual, political, cultural, ethnic, or religious minority.

2.4.2.1 Social Control

As such, social groups use social-control to ensure members comply with social norms. Social-control is the regulation of a person's thoughts and behaviours in ways that prevents, limits, or punishes deviance. Social-control can be active, external, and imposed by 'force' on members. For example, using police against women wearing head scarves, or coloureds for refusing to surrender their seat. Social control can also be passive by using the process of socialisation to disperse social-values within a group. Passive social-control ensures that members police their own actions, e.g. using advertising and fashion to ensure women believe only certain clothing is acceptable.

Social-norms are defined by those within a group that have social-power. In this way, social-control ensures the compliance of members in a group to the will of the most powerful members of that group. Some social-norms may seem quite simple and reasonable - such as 'appropriate' clothing. However as shown above the socially powerful may have ulterior motives, and enforcement of social-norms can be used as mechanism to control individuals. Extreme examples of social-control can be found in Saudi Arabia where women were until recently prohibited from driving.

Socialisation enforces passive social-control thus allowing the socially-powerful to ensure that the members of the group share the professed social-values of the powerful. As such, a person deviating may offend socialised members of the group, even their closest family. Therefore, a non-compliant person can face informal sanction, even exclusion, from all members of the social-group, even their closest family.

A common tactic of those with social-power is to create a perception that deviance from their strictures is unethical and thus immoral. The power of socialisation to create exclusion and thus punishment is clearly a primary motivation for anti-Muslim rhetoric in certain newspapers and affiliated cable television channels. The power of excluding others using social-values has long been demonstrated in France since it helped the rise of those using anti-immigrant rhetoric (Diallo 2020; Quinn 2016).

For more discussion on the topics of privacy, social-control, legal positivism, and human rights see (Carbery 2010; Council 2016; Croome 1995; Crowe & Stephenson 2014; Devlin 2019; ECJ 2007; ECtHR 2010; Gilbert 2014; Green 2018; Himma 2019a, 2019b; India & Ambedkar 1949; Jones 1994; Keoghan & Mcdonough 2017; Kyodo 2019; MacAskill 2016; Marr 1999; Maybee 2016; NAA 2009; Oshinsky 2012; Paine & Philp 1998; Robertson 2006, 2009; Tarafder & Basu 2017; UK 1998; UN 1948; USA 1776; VEOHRC 2019; Warren & Brandeis 1890; Wolmar 2008, p. 157; Wolmar 2009) and Chief Justice of Australia Dixon in Polites v Commonwealth [1945] HCA 3, as well as C.J. Gleeson and J. Kirby in Al-Kateb v Godwin [2004] HCA 37.

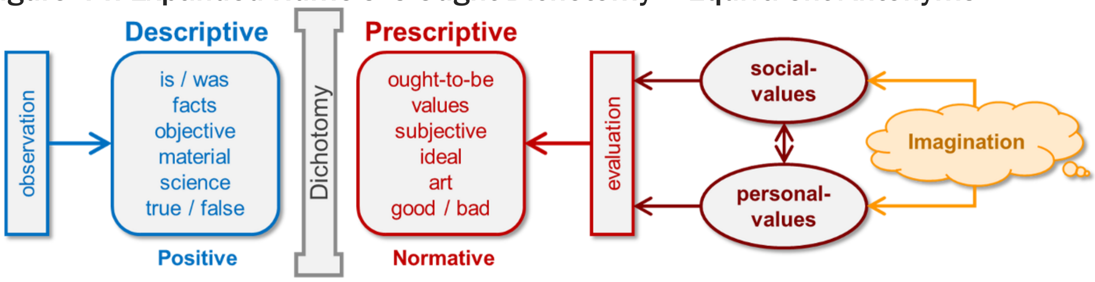
The Engineers Australia (EA) – Guidelines on Professional Conduct was adopted in 2018. EA identified the need for engineers not to discriminate '*in respect of race, religion, gender, age, sexual orientation, marital or family status, national origin, or mental or physical handicaps*' (EA 2018a, p. 89, 1.3 (b)).

The insidious nature of discrimination means that it occurs frequently and uncritically during the evaluations of options for budget expenditure. Despite the NSW Transport Access Program having a budget of over $700 million, after eight years there are still many train stations in Sydney which are only accessible by stairs because much of the funding was spent on car parking. Inaccessible stations are discriminatory because persons with mobility impairments, including the elderly and the disabled, are prevented from accessing the services delivered (Auditor-General 2019; Saulwick 2018).

2.4.3 Liberalism

Social-values are used to evaluate possible behaviours and outcomes to determine the preferred alternative (the social-norm). Observation merely records and describes reality. Evaluation requires imagination to posit alternatives to the current observation and thereafter values to choose a preferred alternative. Thus, social-norms require imagination and are prescriptive thought. As Hume noted, prescription cannot be logically derived solely from descriptions of reality (science).

Figure 44: Expanded Hume's Is-Ought Dichotomy – Equivalent Antonyms



Liberalism is the idea that since social-norms are derived from imagination it is immoral for any person or group to force their values onto a whole society.

After years of religious intolerance, the English developed a theory of religious tolerance based on the theory of 'liberalism'. Religious tolerance was championed by leading thinkers during the Enlightenment, such as Voltaire. What sociologists now describe as social-norms and social-control was identified by John Stuart Mill as 'the moral coercion of public opinion'. In his essay on liberty he argues the key principle of liberalism - that force or social-control should only be used to prevent harm (Anderson 2017; Mill 1859; Murphy 2019; Shafer-Landau 2012).

> 'The object of this Essay is to assert one very simple principle, as entitled to govern absolutely the dealings of society with the individual in the way of compulsion and

> control, whether the means used be physical force in the form of legal penalties, or the moral coercion of public opinion. That principle is, that the sole end for which mankind are warranted, individually or collectively, in interfering with the liberty of action of any of their number is self-protection. The only purpose for which power can rightfully be exercised over any member of a civilised community against [their] will is to prevent harm to others. [Their] own good, either physical or moral, is not a sufficient warrant. [They] cannot right-fully be compelled to do or forbear because it will be better for [them] to do so, because it will make [them] happier, because, in the opinions of others, to do so would be wise, or even right. These are good reasons for remonstrating with [them], or reasoning with [them], or persuading [them], or entreating [them], but not for compelling [them], or visiting [them] with any evil in case he [does] otherwise. To justify that, the conduct from which it is desired to deter [them] must be calculated to produce evil to someone else. The only part of the conduct of anyone, for which [they] is amenable to society, is that which concerns others. In the part which merely concerns [themselves], [their] independence is, of right, absolute. Over [themselves], over [their] own body and mind, the individual is sovereign.' (Mill 1859, p. 13)

In theory, Australia is a secular humanist liberal democracy in which a person's liberty should only be violated to prevent harm to others. In liberal theory the verb violate is chosen intentionally to convey the belief that an act that breaches a person's liberty is an act that violates that person and thus the act is harmful and thus it is unconscionable and immoral. As John Stuart Mill makes clear any means to control a person is morally equivalent, be it social-control or through the state's monopoly on violence.

Liberalism is a logical approach. Self-preservation is a universal value. It can be described from observation. A group can then assume that intervening to prevent another from doing a person harm would be valued by the victim. However, no person can know the internal thoughts and feelings of another. Therefore, a competent person is the best judge of their own self-preservation. Therefore, no group should interfere to prevent what they perceive as a competent person being privately 'immoral'.

The reason privacy is so central to a liberal democracy is that privacy is essential for the exercise of liberty. The loss of liberty eliminates autonomy and negatively impacts an individual's mental health and thus harms them. Thus, violating privacy is illiberal.

2.4.4 Panopticon

Inescapable and unceasing observation, results in a total loss of privacy, creating a state where all individuals are constantly self-policing because any acts of deviance could be observed and punished. As an example, Bentham designed the Panopticon — a circular prison with open cells arranged around a central tower (Reveley 1791). The central tower had narrow observation windows allowing the guards to observe any cell without being seen, therefore, the prisoners would never know if they were being observed. Since non-compliance was punished, all prisoners would have to police themselves because they would never know if their non-compliance (deviance) was being observed. Modern CCTV was initially not recorded but was developed to leverage self-policing because people could not be sure if a guard was watching. Forcing self-policing is behind the development of modern electronic Panopticons.

If social-control, either active or passive (self-policing), prevents a person from acting in a particular way (that does not harm others) due to the will of the socially-powerful, then those persons prevented from acting have been literally subjugated to the will of socially powerful. If a person cannot undertake harmless acts without fear of correction or punishment, they will be unable to live a full life. Thus, unceasing observation removes liberty and is a form of imprisonment.

Historical outrages and atrocities have crafted British and Australian law so that the power of punishment is reserved to the courts in Section 71 (Judicial Power and Courts) of the Constitution of Australia. Imprisonment is considered such a severe punishment that the authority to imprison is only available if no other alternative is available.

> **NSW – Crimes (Sentencing Procedure) Act 1999 – Section 5 – Penalties of imprisonment**
> '(1) A court must not sentence an offender to imprisonment unless it is satisfied, having considered all possible alternatives, that no penalty other than imprisonment is appropriate.'

Winston Churchill wrote in *Closing the Ring*, '*Nothing can be more abhorrent to democracy than to imprison a person or keep him in prison because he is unpopular. This is really the test of civilization*' (Churchill & Churchill 1986, p. 679). As such, the use of unceasing observation to imprison the general populace is unconscionable.

2.4.5 Democracy

Finally, privacy enables democracy through freedom of expression, knowledge, and association. If an individual is not free to think, express, or associate with political or cultural ideas that the socially-powerful disapprove of then the society is not free and certainly not democratic. Inevitably, the ideas that the powerful are likely to disapprove of the most, are any ideas that would reduce or remove their power. Witness the continuing actions by NSW, QLD, CoA, etc to limit the right to protest.

A clear example of freedom of association is the Chartists (1838 -1857) a civil-disobedience movement in the United Kingdom who sought reform to ensure the House of Commons was representative of all the people, not just the powerful. However, the socially-powerful in the UK chose to exercise the ultimate instrument of social-control — the armed forces (specially drafted into the constabulary) — to maintain their position of power until reform was forced upon them thirty years later in 1867. Many political prisoners from this period were transported to penal colonies in Australia.

> ‘who [was] actually transported? You can summarise them into four categories. There were rural convicts, essentially those from a farming background, political prisoners or exiles such as 19th century riotists, significantly the swing rioters of 1830 and those involved in the Chartists Movements of the 1830s and 1840s. There were female convicts, out of the 162,000 it is estimated that 26,000 were women, so it was probably about a sixth of all those who were transported. … between 15,000 and 20,000 children were believed to have been transported between 1787 and 1868.’ (NAA 2009)

It can be reasonably said that ‘rural convicts’ is a tactful euphemism for poor agricultural workers who were dispossessed by changing social-norms championed by the socially-powerful, such as enclosure. Adam Smith notes that since welfare was tied to the parish, not the nation, parishes were eager to use any mechanism available to remove the destitute from their corporations budget (Smith 2002).

Another example of the necessity of freedom of association is the history of the Union Movement. Workers of Great Britain (including on the railways) were not legally allowed to organise until 1824 (Wolmar 2008, p. 157; Wolmar 2009). Since the election of John Howard, the Commonwealth of Australia has increased social-control measures to limit the actions of unions, including banning the display of union symbols — just as French cities tried to ban religious symbols.

> ‘The railways exceed the limits of this unwritten contract and nothing could excuse the steadfast refusal to recognise even the most basic rights of the worker in their quest to maintain dividends to their shareholders.’ (Wolmar 2008, p. 156)

In summary, the socially-powerful use social-control through social-norms to ensure the social-group comply with the values of the socially-powerful. History has shown that the powerful repeatedly use their position to punish those who are not-compliant or deviant. ‘First they ignore you. Then they ridicule you. And then they attack you and want to burn you. And then they build monuments to you.’ Nicholas Klein (1914) to the Amalgamated Clothing Workers of America (Often mis-attributed to Gandhi)

As Hegel noted, society often swings widely between polar extremes – like a pendulum. We may currently enjoy protection of human rights from the powerful, but the pendulum will inevitably swing backwards. For example, the continuing movement in France to increase social isolation of a religious minority.

Therefore, it is important for Engineers and other responsible professionals to embrace a culture of privacy so that their work can never be used by individuals or the state to infringe the right of everyone to live unobserved and unmolested.

2.5 Summary

Privacy is essential for individual mental health, and the exercise of liberty. Therefore, privacy is an intrinsic human right. Since privacy is an intrinsic right, it cannot be surrendered, traded, or curtailed by contract, employment, venue entry, or regulation.

Privacy enables democracy through freedom of expression, knowledge, and association.

Imprisonment is a violation of a person's rights and is only permissible to protect others from harm. Inescapable and unceasing observation, resulting in a total loss of privacy, is a form of imprisonment which is a violation of a person's right to liberty. Therefore, all persons have the right to be unobserved.

These are key points for the operation of a society where information technology allows the collection and collation of vast amounts of biographic information on individuals.

As such, free societies have instituted privacy protections, such as PPIPA in NSW and the CoA's Privacy Act. Under Australian and NSW law transport service partners are legally required to protect the privacy of the individuals they serve or encounter. This legal requirement extends to the universities and to individual researchers who would engage with research into transport customers. This is discussed next, in Chapter 3.

Professional engineers, researchers, and service providers are duty-bound to prevent harm to the public and to clients that would arise as a consequence of their actions or inactions. As such, they should undertake a risk management process such as ISO 31,000 (AS/NZS 4360) to identify and control the risks — either by minimising the probability of an event or by minimising the severity of a consequence.

Deliberate or accidental breach of a customer's privacy is a risk event that could lead to direct negative consequences for the individual and will most certainly result in indirect negative consequences for the organisation, including reputational damage and possible revenue loss. If no risk controls are used to prevent the event of a privacy breach, the likelihood of such an event occurring becomes an almost certainty.

Without undertaking a risk management strategy that aims to protect the privacy of the public and the customers of transport services, researchers and transport operators cannot be allowed access to large empirical datasets, such as Opal. As such, they cannot shift from a model driven to empirically driven approach in service delivery.

Therefore, the first step in the transition to a more empirical approach to service provision is to understand and mitigate the risks to privacy created by the use of large biographic datasets. This thesis examines those risks and proposes a method to control the risk of privacy violation while still allowing measurement of transport service deliver, and customer response to that service delivery as described in Chapter 4.

CHAPTER 3 – PRIVACY PROTECTION

This chapter describes why and how the risk of privacy breaches are a potential barrier that limits the adoption of transport analytics using TOTOR datasets. It establishes the scope for privacy violations and principles of privacy protection before broadly describing the risks in order to establish a risk control framework that complies with privacy legislation. In this thesis, establishing a risk control framework is critical for moving towards an empirically driven approach to public transport analysis.

Chapter 4 will describe a method that uses the privacy protection principles, frameworks, and methods outlined in this chapter to control risk and develop privacy-protecting and-privacy safe datasets that facilitate transport analytics.

3.1 Information Protection Principles

A central component of the privacy protection legislation in NSW, CoA and Canada are Information Protection Principles (IPP) (ARLC 2010; Canada-OPCC 2019; CoA-OAIC 2019; NSW.IPC 2017). Similar principles serve as the basis for the privacy policies of private companies from Japan, Canada, USA and Australia. This legislation and these policies appear to arise from a common framework.

The Organisation for Economic Cooperation and Development (OECD) claims they were '*the first intergovernmental organisation to issue guidelines on international policy for the protection of privacy in computerised data processing*' (OECD notes on OECD 1980).

Table 17: OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data – Part Two – Basic Principles Of National Application

	Principle	Description
1	Collection Limitation Principle	There should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject.
2	Data Quality	Personal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete and kept up-to-date.
3	Purpose Specification	The purposes for which personal data are collected should be specified not later than at the time of data collection and the subsequent use limited to the fulfilment of those purposes or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose.
4	Use Limitation	Personal data should not be disclosed, made available or otherwise used for purposes other than those specified in accordance with [#3 Purpose Specification] except: a) with the consent of the data subject; or b) by the authority of law.
5	Security Safeguards	Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorised access, destruction, use, modification or disclosure of data.
6	Openness	There should be a general policy of openness about developments, [practises] and policies with respect to personal data. Means should be readily available of establishing the existence and nature of personal data, and the main purposes of their use, as well as the identity and usual residence of the data controller.
7	Individual Participation	An individual should have the right: a) to obtain from a data controller, or otherwise, confirmation of whether or not the data controller has data relating to [them]; b) to have communicated to him, data relating to [them] i) within a reasonable time; ii) at a charge, if any, that is not excessive; iii) in a reasonable manner; and iv) in a form that is readily intelligible to [them]; c) to be given reasons if a request made under subparagraphs (a) and (b) is denied, and to be able to challenge such denial; and d) to challenge data relating to him and, if the challenge is successful to have the data erased, rectified, completed or amended.
8	Accountability Principle	A data controller should be accountable for complying with measures which give effect to the principles stated above.

These principles are found in (OECD 1980) Part Two, paragraphs 7 through 14 of Annex to the Recommendation of the Council of 23rd September 1980: Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data. Further discussion of the principles is in the accompanying Explanatory Memorandum, under section II, part B, paragraphs 50 through 62.

The CoA passed the Privacy Act No. 119, 1988 which contained IPP in Section 14[14]. In 1998, '*NSW was the first state to enact public sector privacy laws*' a decade after the CoA, and two decades after the OECD recommendation, (ARLC 2010).[15]

Each of these legal instruments referred to a set of legal principles that have common themes with the principles outline by the OECD in 1980. The shared themes and principles across acts are outlined in the following matrix – Table 18.

Table 18: Matrix of privacy principles from OECD, NSW, and Australia.

OECD	NSW IPP	CoA APP
1 Collection Limitation	1 Lawful 2 Direct	5 Notification 3 Solicited Information
2 Data Quality	9 Accurate	10 Quality
3 Purpose Specification	4 Relevant	6 Use or Disclosure
4 Use Limitation	10 Limited	6 Use or Disclosure
5 Security Safeguards	5 Secure 11 Restricted	11 Security
6 Openness	3 Open 6 Transparent	1 Open & Transparent
7 Individual Participation	7 Accessible 8 Correct	12 Access 13 Correction
8 Accountability	PPIPA	Privacy Act 1988 Privacy Amendment Act 2012
	12 Safeguarded	
		2 Anonymity and pseudonymity
		4 Unsolicited information
		7 Direct Marketing
		8 Cross-border disclosures
		9 Government identifiers

The Australian Privacy Principles (APP) provide the most rights — such as ensuring individuals the right to interact anonymously with entities and prohibiting the use of government identifiers by unauthorised entities. Those two rights were considered requirements for preventing panopticon systems that surveilled the entire populace as a tool of social-control. These rights may have been in response to the outcry over the Australia Card — a commonwealth identity card that was proposed in 1985 by Bob Hawke's government and pursued until 1987 (Greenleaf 2007).

14 The Act has been substantially amended since then in 1990, 2000, 2004, 2012, and 2017. http://www.austlii.edu.au/cgi-bin/viewdoc/au/legis/cth/num_act/pa1988108/s14.html

15 Canada passed the Personal Information Protection and Electronic Documents Act in 2000, (CoA-OAIC 2019).

Table 19: NSW additional Information Protection Principles (IPPs)

	Principle	Description
12	Safeguarded [emphasis added]	An agency cannot disclose **sensitive personal information** without a person's consent, for example, **information about ethnic or racial origin, political opinions, religious or philosophical beliefs, sexual activities or trade union membership**. It can only disclose sensitive information without consent in order to deal with a serious and imminent threat to any person's health or safety. [emphasis added]

Source: (NSW.IPC 2017)

Table 20: Additional Australian Privacy Principles (APPs)

	Principle	Description
2	Anonymity and pseudonymity	Requires APP entities to give individuals the option of not identifying themselves, or of using a pseudonym. Limited exceptions apply.
4	Unsolicited information	Outlines how APP entities must deal with unsolicited personal information.
7	Direct Marketing	An organisation may only use or disclose personal information for direct marketing purposes if certain conditions are met.
8	Cross-border disclosure	Outlines the steps an APP entity must take to protect personal information before it is disclosed overseas.
9	Government identifiers	Outlines the limited circumstances when an organisation may adopt a government related identifier of an individual as its own identifier, or use or disclose a government related identifier of an individual.

Source: (CoA-OAIC 2019)

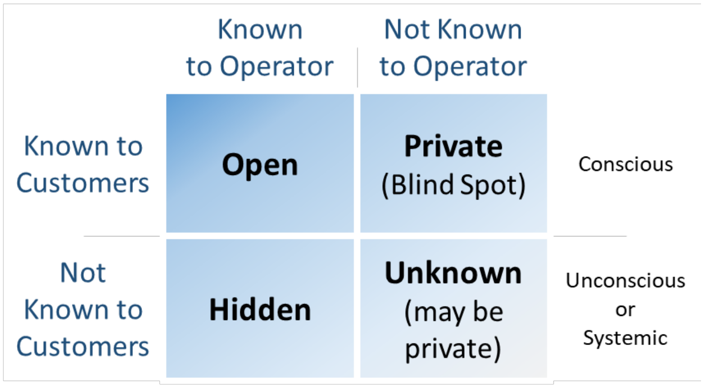
When the Commonwealth Privacy Act was introduced in 1988, it would have been inconceivable for a private corporation to create a surveillance network as large and comprehensive as that of a state. At that time practical surveillance was limited by the capacity of information system technology. Surveillance entities such as the State Security Service (Staatssicherheitsdienst, SSD) of East Germany (Deutsche Demokratische Republik) primarily used paper records. There was also limited public information on the Five-Eyes (Australia, Canada, New Zealand, UK, USA) intelligence alliance. During that time the Five-Eyes officially would not use the intelligence apparatus to monitor citizens of the five participating countries (Bamford 2002).

Each of the IPP from the sources outlined represent a risk management strategy, the most notable is the 12th principle from NSW — an agency cannot disclose sensitive information without consent. The 12th principle is an acknowledgement of the consequences of discrimination that arise from the revelation of sensitive information. Such revelations should therefore be considered a risk event.

The 12th principle has identified databases with sensitive information as a risk source to the individuals detailed within and specifically identifies the storage of information on politics, religion, ethnicity, and sexuality as a significant risk to individual privacy. The guidelines suggest two control mechanisms: the fall-back being the prevention of disclosure, while the preferred option is that organisations should never store sensitive information on individuals, unless critical to providing services.

The concept of privacy can be visualised using a Johari window as shown in Figure 45. The customer's information they want the operator to know would be considered Open. All other matters would be private and should remain in the operators' Blind-Spot.

Figure 45: Privacy expressed as a Johari Window



Remember, there will also be unconscious behaviours of the customer that they would never want the operator to know (currently Unknown but actually private). As such, the operator must be careful to limit information they collect when using technology to observe customers and monitor operations – as in OECD IPP #1.

3.1.1 Waters v Transport for NSW

The following section describes the most important legal case regarding the Opal ETS, derived datasets and their use in NSW. TfNSW and UTS are bound by the Privacy and Personal Information Protection Act 1998 (PPIPA)[16]. For any analysis of the Opal dataset to be legal it must comply with PIPPA, and with the common law established in judgements such as for Waters v Transport for NSW. This particular case is also important because the original determination triggered some changes in behaviour and policies within TfNSW.

16 http://www.austlii.edu.au/cgi-bin/viewdb/au/legis/nsw/consol_act/papipa1998464/

3.1.1.1 Round One

In 2016, Mr Nigel Waters lodged a complaint with the NSW Civil and Administrative Tribunal (NSWCATAD) against TfNSW. Mr Waters represented himself, TfNSW was represented by the NSW Crown Solicitors Office. The NSW Privacy Commissioner registered as an interested party to the case that was heard before Justice McAteer with a decision handed down on 15 February 2018. The case is known as Waters v Transport for NSW [2018] NSWCATAD 40 (15 February 2018) (NSWCATAD 2018)[17].

The case and the decision represent an adjudication over the differing interpretations of the terms '*directly related to an activity of the agency*' and *'reasonably necessary*' from PIPPA under Part 2 – Information protection principles, Division 1 – Principles:

> '**8 Collection of personal information for lawful purposes** [emphasis added]
> (1) A public sector agency must not collect personal information unless:
> (a) the information is collected for a lawful purpose that is *directly related to a function or activity of the agency,* and
> (b) the collection of the information is *reasonably necessary* for that purpose.'

Mr Waters alleged that TfNSW was contravening the IPP under PPIPA. TfNSW contended the data collection was reasonably necessary for revenue protection, through detecting fraud committed using Senior concession cards. J. McAteer decided:

> '1) The decision [in the internal review] of the respondent [TfNSW] is set aside.
>
> 2) Pursuant to section 55 (2) (b) of the Privacy and Personal Information Protection Act 1998, the respondent [TfNSW] is to refrain from the conduct in breach of Information Protection Principle 1 concerning any collection of personal information relating to *travel movement history* of the applicant in contravention of section 8 (1) (b) of the Act.' (NSWCATAD 2018) [emphasis added]

In their submission, TfNSW describes the structure of the data storage in the following excerpts. Since the Opal Database stores a unique card identifier plus the timestamp and location of each segment's Tap-On (embarkation) and Tap-Off (disembarkation), this database contains the complete travel history for each card.

> '73. Personal information … is stored in a dedicated database which is compliant with applicable safety and security standards. ["PAS Database"]
>
> 74. Travel data collected from Opal readers … stored on a separate database within a data centre, managed by a contracted third party, Cubic Transportation Systems Australia ("Opal Database") ….

[17] http://www.austlii.edu.au/cgi-bin/viewdoc/au/cases/nsw/NSWCATAD/2018/40.html

> 75. The Opal Database records each transaction involving an Opal card. Each transaction record contains information such as the card number, time, date and location of the tap on or off, the mode of transport, the value of the journey and any discounts applied. The Opal Database does not contain links to the personal information in the PAS Database and cannot link the Opal card used to the person issued with the card without access to the PAS Database. [emphasis added]
>
> 76. The information collected from the Opal readers is not collected in real time and "live tracking" cannot be performed' (NSWCATAD 2018).

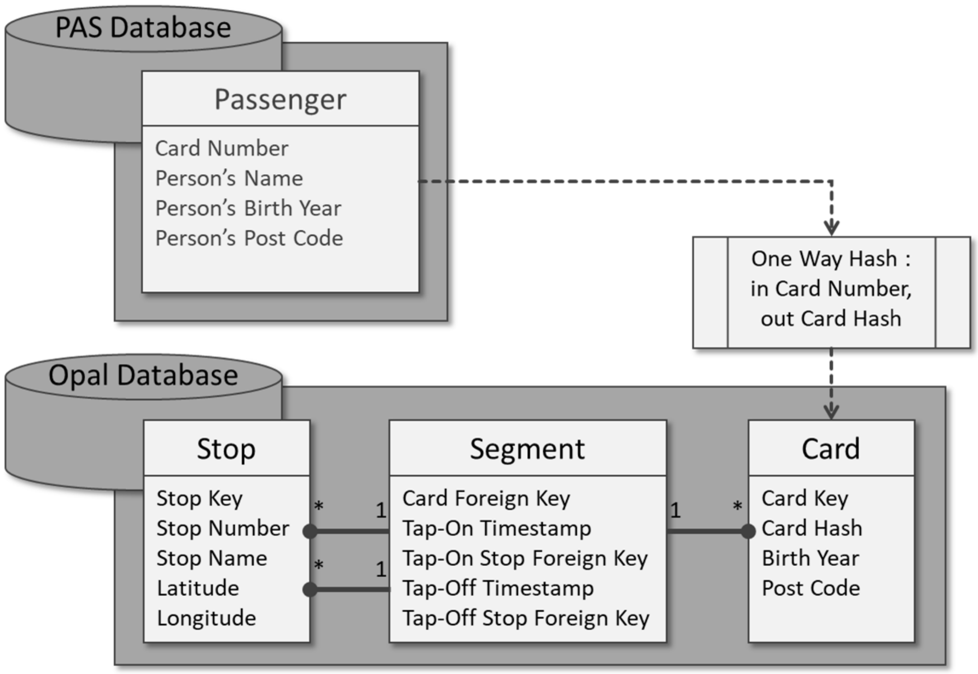
Travel histories are biographical, thus the underlined statement from TfNSW is incorrect because the histories in the Opal Database could be used to identify a person.

Figure 46 below shows a conceptual outline of the structure of the Passenger (PAS) and Opal Databases used by TfNSW. The database documentation suggests that the card number is passed through a one-way hash before it is stored in the Opal Database:

> '*A unique identification number for the Opal card, hashed/masked for privacy purposes and different from the card number printed on the card. Enables the tracking of a card across modes and over time*' (TfNSW 2015).

If the hash is non-reversible, it is not possible to link the travel history directly to a card without access to the PAS Database.

Figure 46: Conceptual structure of PAS and Opal Databases



Source: (NSWCATAD 2018), (TfNSW 2015)

Justice McAteer noted that '141 [TfNSW] has never strongly submitted that it desires the travel history of the Gold Opal card holders, merely it would appear that this is a necessary attendant function intertwined with the technology and the system which the respondent enables and desires in order to monitor the eligibility of these cards.'

In his submissions, the Privacy Commissioner observed that 'The registered card holder's travel history data is about them even on the restrictive approach to the definition, which says the data must be biographical about the person' (NSWCATAD 2018). Justice McAteer agreed. Perhaps his most important finding is that 'travel history… meets the definition of personal information' thus they are covered by the IPP of PPIPA. This finding requires travel history to be treated as personal information and also protected under the Australian Privacy Principles of CoA Privacy Act 1988.

> '84 … The ability of all Opal cards (registered or otherwise) to allow the seamless identification of the recent travel history through the respondent's website is one such observation.
>
> (1) I make a finding consistent with the evidence as set out later in these reasons, and the submissions, arguments and consideration above, that the information provided by the applicant at the time of application and registration of the Gold Opal card is the applicant's personal information.
>
> (2) I make a further finding that the *travel history* as recoded and accessible from the applicant's registered Gold Opal card amongst other things (accessible from the card) *meets the definition of personal information* and is the personal information of the applicant.' (NSWCATAD 2018) [emphasis added]

This common law finding requires that engineers and professionals treat travel histories as personal information and do their best to protect the confidentiality of those histories. Protecting confidentiality prevents engineers from personally examining travel histories.

My reading of the applicant's case was that seniors are prevented from using public transport anonymously as would be required if TfNSW was covered by the CoA under Australian Privacy Principle #2 – anonymity and pseudonymity. The applicants case rests on the contention that at any time the government has complete access to his biographical travel history because it is stored in the PAS and Opal Databases. This is proven by the NSW Police issuing 327 data access requests, without a warrant, in the first six months of 2015 after the full roll-out of the system (Knaus 2016).

> '1 … The dominant concern is that the introduction of electronic ticketing removed the ability of certain concession entitlement holders to travel anonymously under that entitlement, with their movements tracked by the respondent agency (as an arm of the Government), contrary to the privacy protections of citizens under the PPIP Act. …'

> ‘6 The applicant’s general grievance is that this change in the policy has introduced an effective form of surveillance over his ingress and egress within the relevant parts of the State by the lack of any equivalent option for anonymous travel. The applicant ties this grievance to various IPP’s but predominantly his grievance is that the ‘requirement’ of collection of his personal information is not reasonably necessary for the unstated purpose of travel on public transport as an eligible Senior. This central argument equates to a breach of IPP 1 and as a result is contrary to the requirements in s 8 of the PPIP Act.’
>
> ‘19 Other jurisdictions provide anonymous travel for entitlement based discounts’ (NSWCATAD 2018).

3.1.1.2 Round Two

TfNSW appealed the original decision and the three-person panel sided with TfNSW: ‘The decision by Transport for NSW on internal review is affirmed’. The appeal was heard before Hennessey, Seiden, and Lucy as Transport for New South Wales v Waters (No 2) [2019] NSWCATAP 96 (18 April 2019)[18]; (NSWCATAD 2019).

> ‘35 We agree with the appellant that the issue is whether the collection of the particular information that constitutes “personal information” is reasonably necessary for the lawful purpose for which it has been collected. Section 8(1)(b) does not require us to answer the question of whether the collection of information as personal information is reasonably necessary. Mr Waters’ submission focuses on the consequences of collection, not on whether the collection is reasonably necessary for the ticketing purpose.
>
> 36 We find that the collection of the Travel Data is for the ticketing purpose (calculation and collection of the correct fare) and is *reasonably necessary for that purpose.’* (NSWCATAD 2019) [emphasis added]

It is important to note the panel only found that the collection of travel histories in this case was reasonably necessary for the purpose of calculating fares. As such, the agency was not in breach of IPP #1 of the PPIPA. The appeal did not disagree with the finding that travel histories are personal information and are covered by the IPP. The appeal panel acknowledges it was not necessary to prevent concession card holders from having anonymous travel histories (as available in other jurisdictions), and that removing their privacy was only a matter of TfNSW policy.

The appeal finding focuses on a narrow definitional issue and does not assess the long-term impacts of increasing surveillance. The panel ignored the vast difference between 1) collection of information and 2) indefinite storage of information using an ever-increasing database. TfNSW does not provide a compelling reason why fare collection

18 http://www.austlii.edu.au/cgi-bin/viewdoc/au/cases/nsw/NSWCATAP/2019/96.html

requires indefinite storage of travel histories. They provide no reason why identifiers in all travel records could not be anonymised after a reasonable period, such as 91 days.

Nonetheless, it is now established common law in NSW that travel histories are personal information and must be kept private by any government or public institution or any corporation covered by the Commonwealth's Privacy Act.

3.2 Biography with or without consent

There is an old axiom in data science — that it takes filtering, combination, and analysis to turn data into information. Before proceeding, some terminology needs to be introduced from computer science to clarify how these findings have been used.

It is inefficient to repeatedly store data that does not change in a database, especially as computers are designed for fast information retrieval, which is why computer scientists use Entity Relationship Models when designing databases. The Entity Relationship Model allows records (rows) in one database table to refer to individual records in others through the use of unique identifiers.

Table 21 for example, shows that the CBD's railway stations were used for 180,000 disembarkations during the morning peak period, with Town Hall and Wynyard stations used for over 53% of those disembarkations. The latitude and longitude of Town Hall or Wynyard stations does not change during the day. As such, there is no point in repeatedly storing their coordinates (stable data) in each disembarkation record.

Table 21: Average workday CBD peak period disembarkations in Nov. 2016

Mode	Location	Workday [07:00-10:00)
Bus	2000 Series Bus Stops	60,700
Ferry	Circular Quay Wharves	5,500
Light rail	Capitol Square	300
	Central Station	800
	Paddy's Markets	200
Train	Central	42,800
	Circular Quay	10,900
	Martin Place	17,400
	Museum	7,100
	St James	6,400
	Town Hall	48,600
	Wynyard	48,500
Grand Total		249,200

Source: (Hounsell 2017a)

3.2.1 Identifiers

A unique identifier is a value that allows the extraction of a single record from a collection of records (rows) stored in a table. Each primary table in a relational database will be designed to have a single field (column) that contains only unique values to serve as identifiers. This field will be referred to as the key for that table. Creating a key allows other tables to refer back to a single record in the original table by using *foreign* identifiers, so other tables will create a foreign-key field to hold those references. This allows a database to minimise storage and bandwidth requirements.

For example, Figure 46 shows the structure of the Opal Database. The stop table has a key, and the segment table refers to those in two foreign keys. This allows each record in the segment table to refer to two unmoving stations (Tap-On and Tap-Off).

Table 22 below has an extract of a NSW public transport stops database table while Table 23 provides a demonstration of what the travel histories in the journey segments table could look like. The first two record (rows) in Table 23 demonstrate Card 1 & Card 2 travelled from Stop 5 to Stop 1 at 12:03. Using the Tap-On Stop Foreign Key to refer back to the stops table retrieves a single record for the Tap-On stop. Therefore, the first two records demonstrate that those cards travelled from Central to Circular Quay. The next two records demonstrate the two cards embarked at Circular Quay at 13:17 but the first card travelled to Museum and the second card to Central.

Table 22: Extract of NSW stops table (NSW GTFS & Opal)

Stop Key	Stop Number	Stop Name	Latitude	Longitude
1	200020	Circular Quay Station	-33.861324	151.21039
2	200030	Martin Place Station	-33.867871	151.211591
3	200040	Museum Station	-33.8757	151.210106
4	200050	St James Station	-33.870697	151.211837
5	200060	Central Station	-33.883882	151.205829

Table 23: Fake NSW journey segments table for the above example

Card Foreign Key	Tap-On Time Stamp	Tap-On Stop Foreign Key	Tap-Off Time Stamp	Tap-Off Stop Foreign Key
1	2021-01-02 12:03	5	2021-01-02 12:13	1
2	2021-01-02 12:03	5	2021-01-02 12:13	1
1	2021-01-02 13:17	1	2021-01-02 13:24	3
2	2021-01-02 13:17	1	2021-01-02 13:27	5

Although we can assume the cards were being used by humans to travel on the trains, we cannot assume the same person made both journeys in the dataset. That is why the analysis refers to cards travelling and not persons, as there is insufficient evidence in the dataset to prove a card's travel history was created by the actions of just a single person.

3.2.2 Composite keys

In the real-world, it is not possible to assign a unique identifier to each individual. Instead, it may be possible to combine several fields to create a unique tuple to identify an individual. A unique identifying tuple is called a composite key. For example, a Gregorian date is a composite key composed of three numbers used to create a unique identifier for every day in history. Similarly, the geo-location tuple of latitude and longitude could be used to create a composite key for a transit stop.

```
Date ::= ( Day of the Month, Month, Year )
Location ::= ( Latitude, Longitude )
```

3.2.3 Privacy budget

Since several data points can be combined to create a composite key, each additional data point describing an individual increases the potential for uniquely identifying an individual. That is because more data points reduce the search space in a non-linear way by allowing correlation across multiple data sources to describe an individual.

For example, Mohammad is reputed to be the world's most popular first name. However, such a popular first name still reduces the list of possible individuals within the list of Australian citizens. Two more data points, birth year and suburb of residence, may be all that is required to uniquely identify one Mohammad from the list of all Australians. Since no data release can be retracted, the release of information describing an individual reduces the privacy budget of that individual as each additional data point released decreases the possibility space that must be searched to isolate that individual. To control the risk of a privacy violation, it is best practice to store only the *necessary* information, and to prevent the release of data that could be used in a composite key.

Privacy budgets are a special concern for large organisations, such as governments and mega-corporations, because coordination and double-checking all data released across the entity is not feasible. In these organisations, the principles of safety through design must be used to prevent the release of biographic information about individuals.

At this point, it is essential to draw a distinction between biographical data describing individuals, datasets describing corporations (which are not people and so do not have human rights), and datasets listing activities. The next sub-section describes biographical datasets and the subsequent sub-section describes activity datasets.

3.2.4 Biography

Data that create a picture of the characteristics, beliefs, and behaviour of an individual is considered biographical because it constructs a detailed picture of the person and their life[19]. An identifier (such as a driver's license number) alone is not biographical, because it provides no data on the person's characteristics, beliefs, or behaviour[20]. Identifying information tells you unambiguously who a person is, whereas biographical information tells you how the person behaves.

A central principle of privacy protection is that biographical information is personal information and should be protected. The more data stored about a person, the more detailed the picture it paints, and so the data becomes ever more biographical.

The NSW Privacy Commissioner in their submission to Waters vs TfNSW, noted that the PAS Database contained identifying and biographical information, whereas the travel histories in the Opal Database are biographical information. Biographical information like travel histories is a greater risk to a person's privacy, as it records actual behaviours and documents their expressed preferences. As such, it is unethical to use technology to unnecessarily document a person's actual behaviours.

Timestamped embarkation-disembarkation travel pairs from a TOTOR dataset are biographical. Extra care must be taken to ensure the privacy risks are controlled when handling travel pairs. All timestamped locations are biographical, however travel pairs are certainly more biographical than just embarkations recorded in a TOR dataset.
A clarifying example of biographical information can be provided by the judgement for Waters v TfNSW [2018] NSWCATAD 40. The full name (identifier) Nigel Waters is a single data point on the individual. Using only that data it is not possible to write a biography on that person. However, we can add additional data from the judgement,

[19] In this thesis, the term biography is used in the broad sense of privacy theory. In privacy law, the term biography has developed specific meaning, especially after the UK case of Durant v Financial Services Authority (Linsay 2004).

[20] Some names clearly have religious and ethnic origins, such as the many variants of Mohammad.

first Nigel Waters identified himself to the court with the pronouns he/him, and that he is aged over 60 years old. He represented himself to the NSW Civil and Administrative Tribunal and he presented a compelling case on the difficult and obscure field of privacy law suggesting experience in the field. A quick search for ‘Nigel Waters privacy law’ and the top result is a journal article listing Mr Waters as the Associate Editor: *Waters, N. 1998, ‘New health privacy law in Canberra’, Privacy Law & Policy Reporter, vol 121* retrieved from the excellent AustLII[21]. Now we have a more informative biography of the individual — he is a retired lawyer specialising in privacy law. If Mr Waters biographer was then furnished with his complete Opal travel history, it may be possible to write a comprehensive biography of Mr Waters characteristics, beliefs, and behaviours. The comprehensive biographic power of travel histories appears to be his primary complaint against the long-term storage of travel history in a system where histories can be retrieved using just a name.

Returning to 3.1 – Information Protection Principles, NSW identifies a special class of biographical information called *‘sensitive personal information’*[22] that includes data on ethnic or racial origin, political opinions, religious or philosophical beliefs, sexual activities or trade union membership. As discussed in 2.4.1 – Social Norms, humans discriminate against and persecute individuals because of certain categorisations. Unfortunately, throughout the world there are still instances of people being beaten, maimed, or killed because they were perceived to belong to a certain category. Therefore, the categories frequently used for persecution were identified as sensitive information. Two horrible examples from 2019, were the massacre of Muslims in Christchurch, New Zealand and the massacre of Christians in Colombo, Sri Lanka.

A person’s travel histories may be further analysed to estimate sensitive personal information. Consider if the travel history showed a person regularly travelling to a bus stop located near a prominent mosque, union hall, or club. Now add accurate timestamps showing them arriving before services/meetings, and then showing them departing after those services/meetings. A biographer could conclude that person

[21] *‘The Australasian Legal Information Institute (AustLII) provides free internet access to Australasian legal materials [including legislation, judgements, reports, and law journals]. AustLII’s broad public policy agenda is to improve access to justice through better access to information.’* See http://www.austlii.edu.au/about.html

[22] NSW PPIPA forbids a public-sector agency from releasing sensitive information without specific.

attended those services/meetings. In this way the person's union / club membership, religious beliefs, or political opinions could be revealed by their travel history.

3.2.5 Activities

A dataset can be created to just record the activities of a collection of people rather than provide biographies on individuals. Public transport patronage is an activity dataset because it records the collective response to the service provided. An example of activity datasets is the number of embarkations at a station, or the volume capacity ratios for a network link. Activity datasets can also contain anonymous descriptions of behaviours, such as movements of 'blobs' through a train door.

Activity datasets usually use the four techniques of privacy protection – deidentification, disassociation, aggregation and elimination

As an example, a dataset indicating that one particular card Tapped-On at Bell Station in NSW between 06:50-07:00 every day of one given week is clearly a biography of an individual card. However, a record indicating that there was one Tap-On at Lithgow between 06:50-07:00 on Monday to Sunday provides information on activities. The difference is the second activity dataset uses deidentification to create the possibility that the observations were just the coincidental action of seven different individuals.

Records of activity on transport services are about crowds and are not biographical. However, if the activity were only by one person then the data could be used in composite keys to facilitate collation and create a biography. For example, at remote stations like Bell there may be only one person using the station, thus descriptions of activity at that station risk being biographical. Eliminating the records on very small crowds is a common technique to prevent accidental biography.

However, once the crowd size reaches a reasonable threshold, a dataset recording the activities of a crowd is not biographical because the number of individuals described make it impossible to isolate the behaviour of any one individual - that is the benefit of aggregation. This thesis does not attempt to specify the level of that threshold. As a thought experiment, consider the relative difficulty in identifying persons embarking every workday between 09:30-09:40 at Bell if there is just one person versus the difficulty if there is twenty persons.

There is an important distinction between travel histories and activity patterns – travel histories are unambiguously a biography of an individual, whereas activity patterns record activities occurring at a location or on a service. Travel histories are definitely biographical and thus personal information, whereas activity patterns are less descriptive and increase the search space. Thus, using activity patterns instead of travel histories in measurements is more likely to protect the privacy of individuals and control the risk.

Breaking up a TOTOR dataset into activity patterns removes the travel histories, thus transforming the dataset from personal information that require privacy protection into activity datasets documenting crowd sizes, which control the risk of privacy violations by removing the biographies.

3.3 Not what you think

Empirical datasets represent a complex reality and require careful interpretation. That is true for Opal datasets, and especially true for derived activity datasets. However, researchers must remember to guard against assumptions that are not practically true.

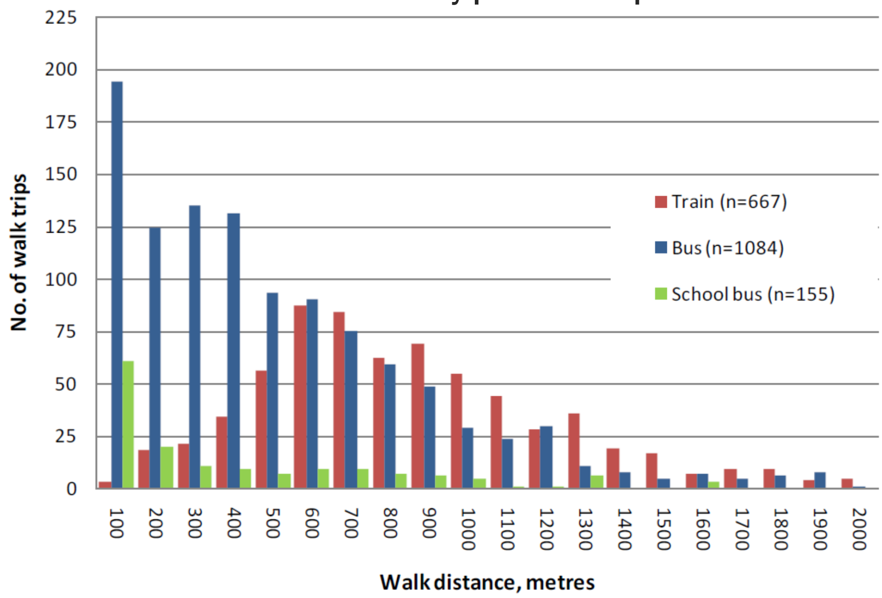
For example, the assumption that each Opal card represents an individual is empirically false and should not be used during analysis. Although over 90 percent of public transport users in NSW comply with the fare regulations on-board a service, there is nothing physically preventing them from sharing cards. It is also incorrect to assume that any individual person is only using one Opal card. There is nothing practical preventing a family from sharing one or more cards.

3.3.1 Walking distance

Disembarkation location is biographical — but human behaviour limits its descriptive potential. This section describes one simple reason why determining the origin and destination of a journey segment is not possible using the TOTOR records alone.

As described in Daniels & Mulley (2013) the stated walking distance for the majority of persons accessing buses and trains is up to 1,500 m. There is significant variation with the most common distance being between 500 and 600m.

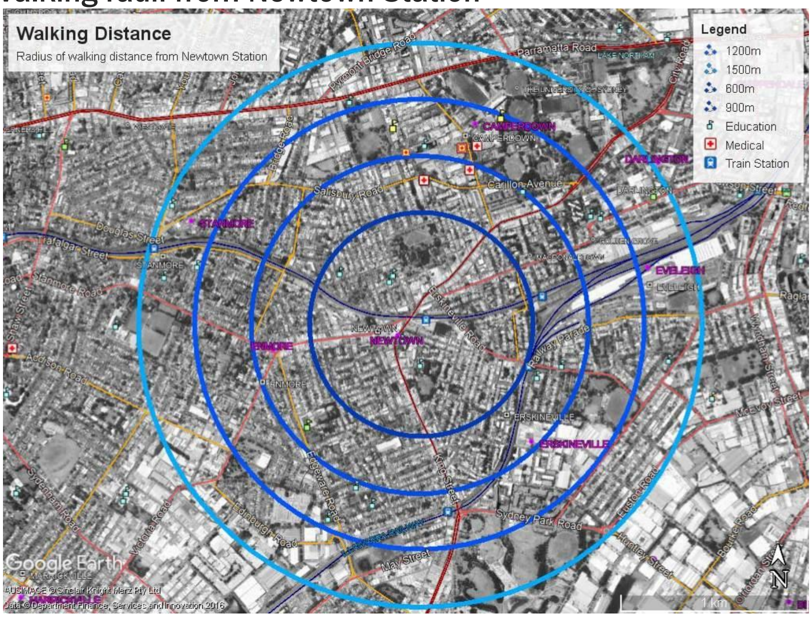
Figure 47: Distribution of walk distance by public transport mode



Source: (Daniels & Mulley 2013, p. 12, Figure 3)

As an example, Figure 48 below shows ‘Walking Radii’ — circles overlaid on an aerial photograph with the centre on the forecourt of Newtown station. [23]

Figure 48: Walking radii from Newtown Station



Using: (using NSW 2019c)

The circles have radii of either 600, 900, 1,200 or 1,500 metres. From the surveys, we know that some passengers will walk up to 1,500 m to and from a train station, especially if that station has a high service frequency like Newtown or Redfern. As such, a card disembarking at Newtown station could be travelling to locations as

[23] A 1 km walking radius creates a probability space of 3.14 km^2 while a 1.5 km walking radius covers 7.1 km^2.

disparate as the Parramatta Rd McDonald's in Stanmore, Newington school in Stanmore, the Factory Theatre in Marrickville, the trapeze school in St Peters, the Bunnings in Alexandria, or many buildings within the University of Sydney in Camperdown and Darlington.

Consequently, the probability space for determining the ultimate destination of a person disembarking at a particular train station within an urban environment is too large to draw significant conclusions from a single data point. This caveat is important when considering the differences between biography and activity datasets, as a disembarkation stop only describes a place a person was, not where they went.

However, the more data points available to describe an individual's behaviour, the more comprehensive an analysis of their behaviour could be and the more biographical a dataset could be. For example, given a year's worth of train and bus travel histories for a card, it would be possible to apply clustering and triangulation algorithms to develop a higher probability estimate of their home based on their first journey of the day, see (Hasan et al. 2013; Long & Thill 2015; Ma et al. 2013; Zhong et al. 2014).

The risk of biography created by a person's travel history is proportional to the number of records contained. People with unchanging travel patterns will be exposed to less risk than a person with geographically varied travel patterns that centre around their home. However, even with a large number of data points it may be impossible to identify an individual's house.

When assessing the consequence to individuals of the revelation of a single travel record, it is important to remember that the embarkation and disembarkation stop provide information on public transport use, they do not provide information on the origin or destination of a journey because many destinations are reachable from a stop.

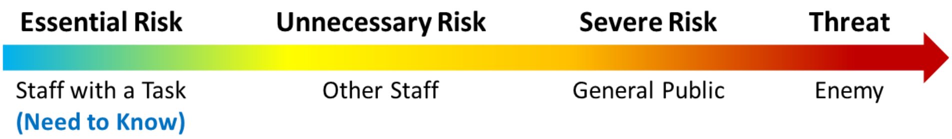
Although walking distance mitigates much of the risk to individuals from datasets, it does not mitigate the reputational risk to an organisation from the reaction to the perceived risks which individuals will assume is higher than the actual risk.

3.4 Need-to-Know

The Need-to-Know principle is central to information security and operational security. The Opal Database hashes the Card Identification Number (CIN) to disassociate it from the identity provided during card registration, as the ticket processor using the Opal Database does not require names or credit card numbers to function. This is the first layer of privacy protection and represents the Need-to-Know framework in action.

The Need-to-Know framework was used extensively through World War II to combat espionage. Simply, if a person knows a secret there is a risk they may intentionally or (more likely) unintentionally reveal that secret. The most effective risk control is to minimise the number of people who know that secret. Therefore, a secret was only revealed to the persons who needed to know about it in order to undertake their tasks.

Figure 49: Need-to-Know Spectrum



> 'When access to covered data is broader than what is required for legitimate purposes, there is unnecessary risk of an attacker gaining access to the data.' UC Berkley - Information Security Office - Need to Know Access Control Guideline[24]

The Need-to-Know principle is expressed in UTS' internal guidelines as follows:

> 'To reduce the possibility of an accidental breach of confidentiality or privacy, UTS Researchers SHOULD minimise their use of the Opal Data to a minimal subset required to answer their research question'.

A dataset can be considered as a collection of records (rows) that contain a required set of parameters or fields (columns), like a well-structured table or spreadsheet. The Need-to-Know risk control process aims to ensure researchers focus their investigations on the minimum number of fields required. In other words, fields which the researchers do not Need-to-Know are excluded from the dataset they have access to.

3.4.1 UTS Opal Data Access Guidelines

When TfNSW began discussions with UTS and myself regarding the use of TOTOR datasets, we began an informal risk management cycle. UTS developed an internal Opal Data Access Protocol to assure TfNSW of our commitment to protect customer privacy. This protocol has several goals that can be summarised as:

[24] https://security.berkeley.edu/need-know-access-control-guideline

1.1 — Protect the privacy of Opal customers. The raw Opal datasets contain travel histories. If data relating to individual travel histories was made public that data could be used to transgress the privacy of Opal customers, therefore UTS will ensure all researchers understand how to execute their legal duty to protect the customers privacy.

1.2 — Advance knowledge of NSW transport systems by developing theories and methods to analyse customer and system behaviour and aid in optimising the performance of transport services to enhance customer experience.

1.3 and 1.4 — Reduce demands on TfNSW staff for specific research from UTS researchers. The method proposed in the Chapter 4 will substantially reduce the demands on TfNSW staff by creating datasets that are privacy-protecting and privacy-safe by design.

1.5 — Increase opportunities for innovation and development of new applications and technologies using Opal datasets by exposing UTS researchers to the needs of TfNSW to improve the quality and performance of public transport services and the ensuing research opportunities featuring Opal datasets.

UTS undertook a cyclic process to integrate privacy risk management into the process of using datasets that relate to travel histories. I was the primary author of the guidelines for controlling the risk arising from Opal datasets. The guidelines require researchers to keep confidential all information received about individuals.

To comply with the University's existing precedent and practises, researchers have to treat an individual's travel history as confidential information and cannot access it without consent. Therefore, researchers cannot extract an individual's travel history from any dataset for examination. This is stated in the guiding principles as:

> 'UTS Researchers MUST respect that an individual person's travel history is personal and identifying information. As such, UTS Researchers MUST NOT remove a person's anonymity or breech their privacy by examining individual travel histories; or associating those travel histories with other personal or identifying information. UTS Researchers MUST NOT ever publish or release individual travel histories.'

The Need-to-Know principle forms the basis of the Opal Data Access Guidelines. The guidelines recommend research is designed to avoid datasets with travel histories. A simple first step to protect privacy is removing fields that identify individual cards to make the extraction of a biographic travel history for an individual impossible.

> 'Derived Opal Information is data or descriptions about the world that no longer contains any individual's ticketing information. It is RECOMMENDED that UTS Researchers preference Derived Opal Information for research.'

3.5 Derived Datasets

The Need-to-Know principle can be implemented by transforming raw biographical dataset into derived activity datasets though data treatments that eliminate biographical data. These are referred to as derived datasets in this thesis. This section describes the methods of privacy protection for creating derived datasets.

3.5.1 Identifying Persons

A person can be identified by:

- Identifiers – such as Name, Date of Birth, or Opal Card Number

A person can be described biographically by:

- Unique Behavioural Patterns – especially repeated spatial, temporal, and spatio-temporal patterns.

The objectives of privacy protection methods are removal of identifiers and biographies from the dataset. Removing identifiers and biographies is achieved by using:

- Deidentification
- Disassociation
- Aggregation
- Elimination

Deidentification is the first step in privacy protection. It is possible to remove identifiers (such as names) from a dataset and replace them with a descriptor that allows for further analysis without identifying a specific user. For example, a card identifier can be replaced with a unique randomly generate key or by a hash.

These four methods work by increasing the probability space in which an action could have occurred to reduce the confidence interval for identifying the participants in that action. There are over five million residents in the NSW Greater Metropolitan Area (GMA). After deidentifying the ticketing dataset, the probability of linking an individual data point in the dataset to an individual is 1 in 5 million (2.0e-7).

For example, a dataset revealing that three thousand trips occurred from Epping to Macquarie Park on Monday provides no biographical information about the actual behaviour patterns. A dataset revealing that two trips occurred from Epping to Macquarie Park at 07:34 on Monday gives unnecessary information and clearly describes individuals. A derived dataset that shows five hundred trips were made between Epping to Macquarie Park between 07:30 and 08:00 balances the need to protect privacy and the operator's and community's need for rich empirical analysis.

There are good reasons to undertake analysis on the complete travel history of a card, primarily to search for patterns across the entire population (Hounsell 2017a). However, extreme care must be taken when conducting such analysis. That includes resisting the temptation to examine the history of interesting cards, such as the card that undertook over 900 journey segments in 42 days.

3.5.2 Privacy Budget

Identifying patterns may not be immediately obvious in a single released dataset. However, patterns may become obvious with the release of more datasets. For example, if there were 51 trips from Bankstown to Granville on Monday and every day after there was only 50, the missing trip is now a unique event.

The biographic power of uniqueness of events revealed by time can be prevented by using aggregation and elimination to remove noise that creates identifying patterns. For example, always rounding counts down to a multiple of 10 (or 1,000) is an elimination strategy long employed in NSW (BTS 2014).

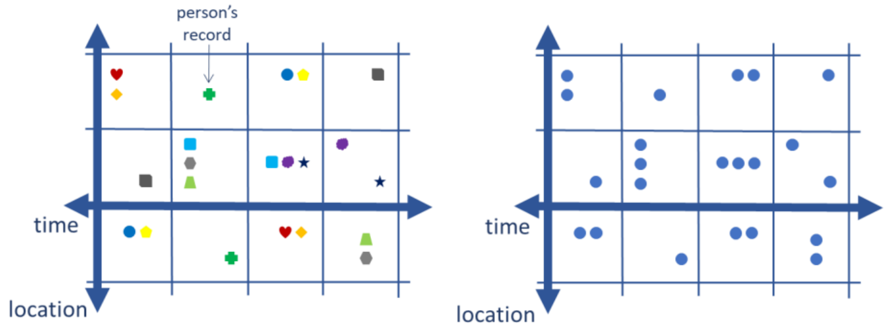
3.5.3 The four methods of privacy protection

Disassociation involves removing a particular dimension from a dataset, such as time or location. The more dimensions on a data point, the more precise a description it provides — in other words the more fields a record has the more biographical it is.

Disassociation works by separating fields from a data point that would be used to detect a pattern, such as breaking apart embarkation and disembarkation pairs or removing time. For example, a minute by minute count of the embarkations at Town Hall during peak-hour is an activity dataset and would assist operators but not describe individuals.

Disassociation can also be achieved by changing the type of measurement, such as percentages or rankings. Deidentification is technically another form of disassociation.

Figure 50: A diagram of disassociation by removing identifiers



Aggregation works by consolidating multiple records into a single point within a larger probability space. The most common form of aggregation is a simple count, such as counting passengers per hour, or per station. A dataset counting embarkations per Local Government Area (LGA) is another example that would cover many possible stations. In this case, the LGA does not narrow the probability space for the actual origin.

Figure 51: A diagram of aggregation in time

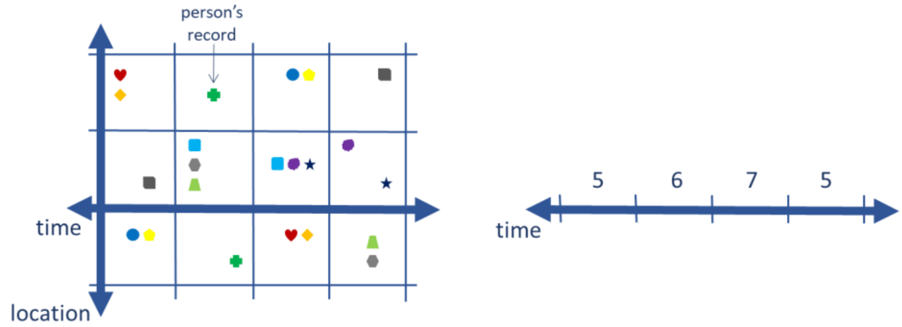
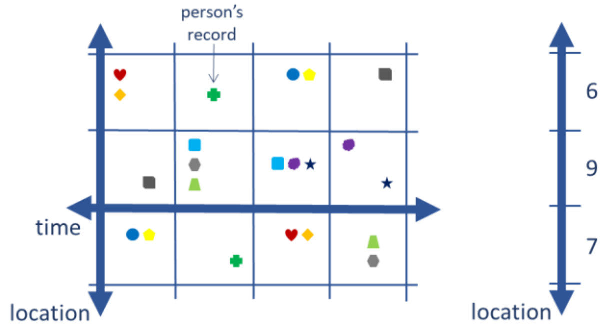
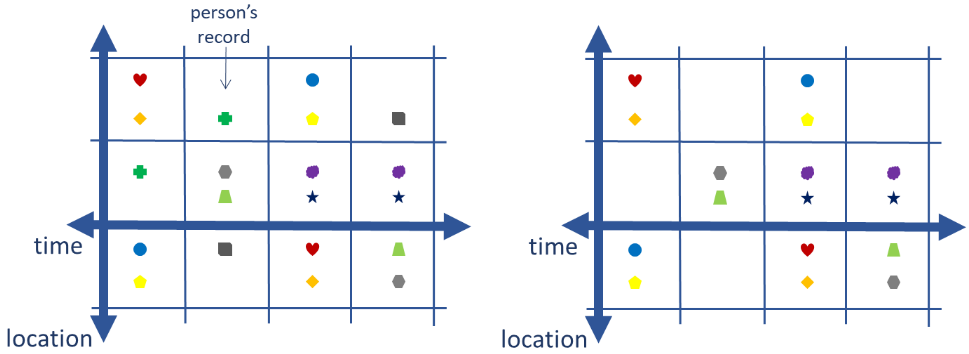


Figure 52: A diagram of aggregation in space



Elimination is a fall-back strategy where rare or unique (thus biographical) patterns are eliminated from the dataset. Elimination may be required for areas with limited services because a minimal number of potential passengers creates a small probability space that would render disassociation and aggregation ineffective. Elimination as privacy protection would mean a dataset is no longer exact, however, the community will know the actual patronage is greater than or equal to the values released.

Figure 53: A diagram of elimination of unique records



Count at Location Before Elimination					Count at Location After Elimination			
5	4	6	5		4	2	6	4

All derived datasets should be generated from the original data set because error accumulation will generate inaccuracies if derived datasets are used to generate subsequent data sets. For example, in Figure 53 if the resulting dataset after elimination was then aggregated over time, the error rate would be either 14.3% or 33.3% per location, which is unacceptably high.

3.5.4 Spatial aggregation

One method of privacy protection commonly used in Australia, including by the ABS, is spatial aggregations where results are aggregated into well-known areas.

ABS provide many aggregate statistics, such as the census Journey to Work results, using the Australian Statistical Geography Standard (ASGS). This standard provides a hierarchy of geographical areas, each higher level is built from the levels below.

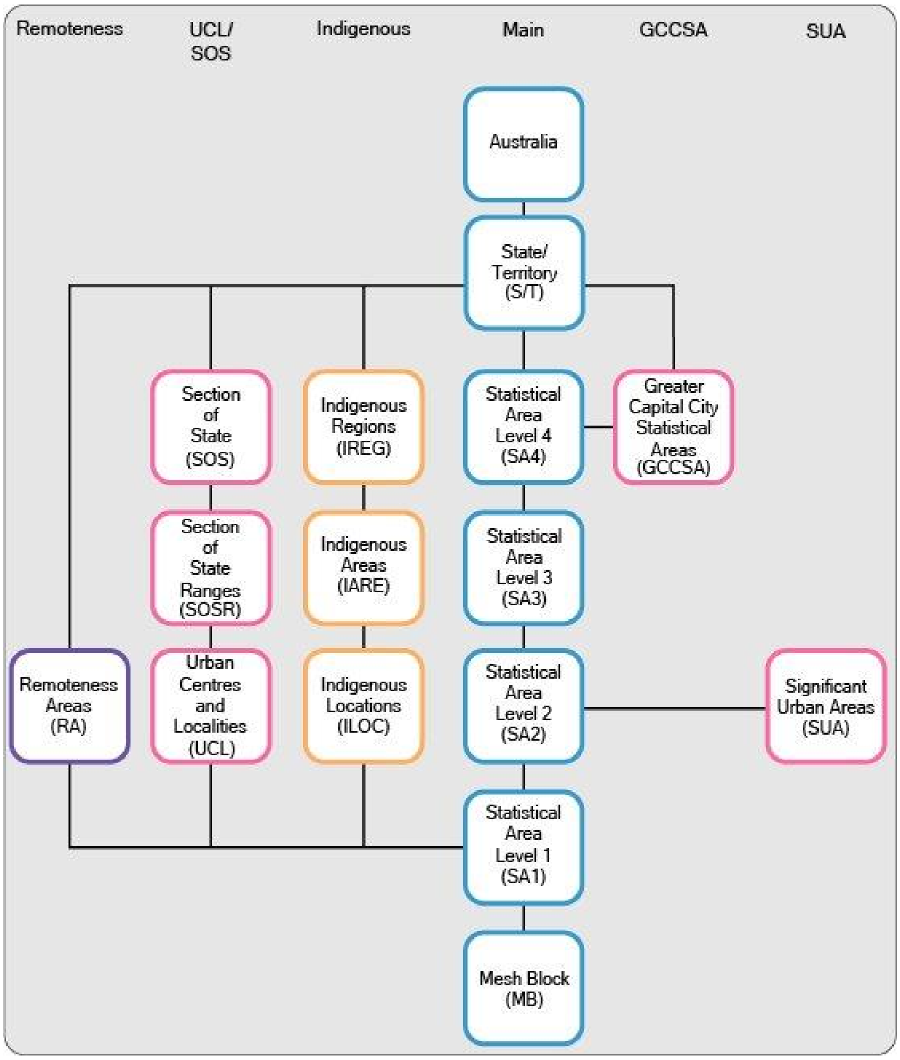
Since SA1 are quite small, that level of aggregation could risk privacy, therefore aggregations from datasets should generally occur at the SA2 level or above, because the area covered is defined by population where there are usually several stops per area.

Table 24: ABS ASGS Statistical Area Levels

Abbreviation	Name	Regions	Population Range
SA4	Statistical Area Level 4	106	100,000-500,000
SA3	Statistical Area Level 3	351	30,000-130,000
SA2	Statistical Area Level 2	2241	3,000-25,000
SA1	Statistical Area Level 1	54,805	200-800

Source: (ABS 2016, Diagram 1)

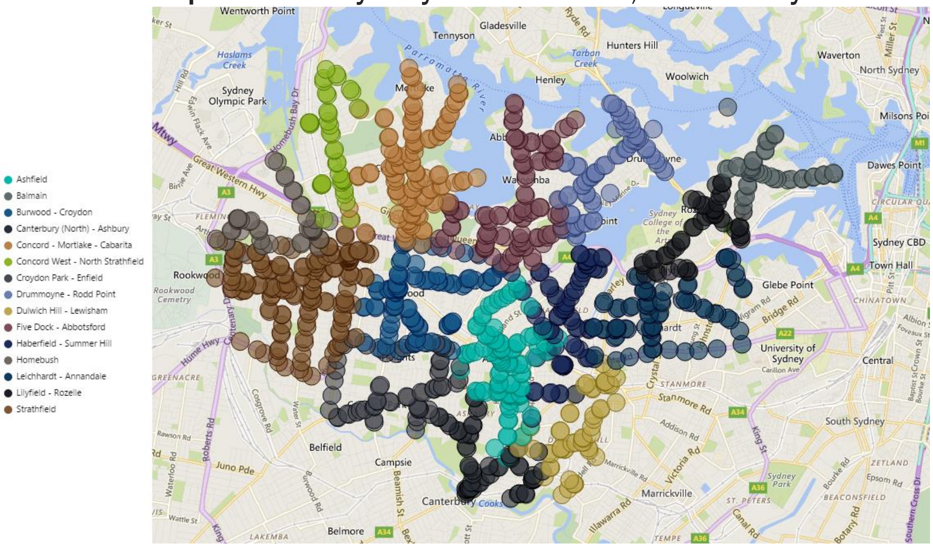
Figure 54: ASGS ABS Structures



Source: (ABS 2016, Diagram 1)

As an aggregation example, Figure 55 shows all the transit stops within the region defined by the ABS as SA4 “Sydney – Inner West”. The stops are colour coded by SA2 and this clearly shows the large probability space created by an SA2 spatial aggregation.

Figure 55: Transit stops in SA4 "Sydney – Inner West", coloured by SA2



Source: (TfNSW 2016b), (ABS 2016)

In practice, each stop will serve the full walking distance, and so the stops will serve persons in surrounding areas. The stops on the south-side of Parramatta Rd in Stanmore will still serve the passengers on the north-side of Parramatta Rd in Leichhardt Annandale SA2 (navy blue in Figure 55). An analysis of travel from the Leichhardt Annandale SA2 would need to include stops on both sides of Parramatta Rd and may need to examine stops up to 1.5km walk from the perimeter.

A similar form of aggregation is by Australian Post Code, but 25 postcodes in Sydney have as few as 12 transit stops. Worse, the analysis would be complicated by 24 postcodes in Sydney having over 240 stops, and three postcodes with over 600 stops.

3.5.4.1 Reference Aggregations

In the history of NSW data analysis, the following spatial aggregations (or similar) have been used repeatedly throughout the history of transport planning.

NSW Greater Metropolitan Region (GMR) including the Sydney Greater Capital City Statistical Area (GCCSA), and the Illawarra and Newcastle Region. Newcastle Region covers the Lake Macquarie – East SA3, the Lake Macquarie – West SA3, the Lower Hunter SA3, the Maitland SA3, the Newcastle SA3 and the Port Stephens SA3. Illawarra Region covers the Dapto – Port Kembla SA3, the Illawarra Catchment Reserve SA3, Kiama – Shellharbour SA3, Shoalhaven SA3, Southern Highlands SA3 and the Wollongong SA3.

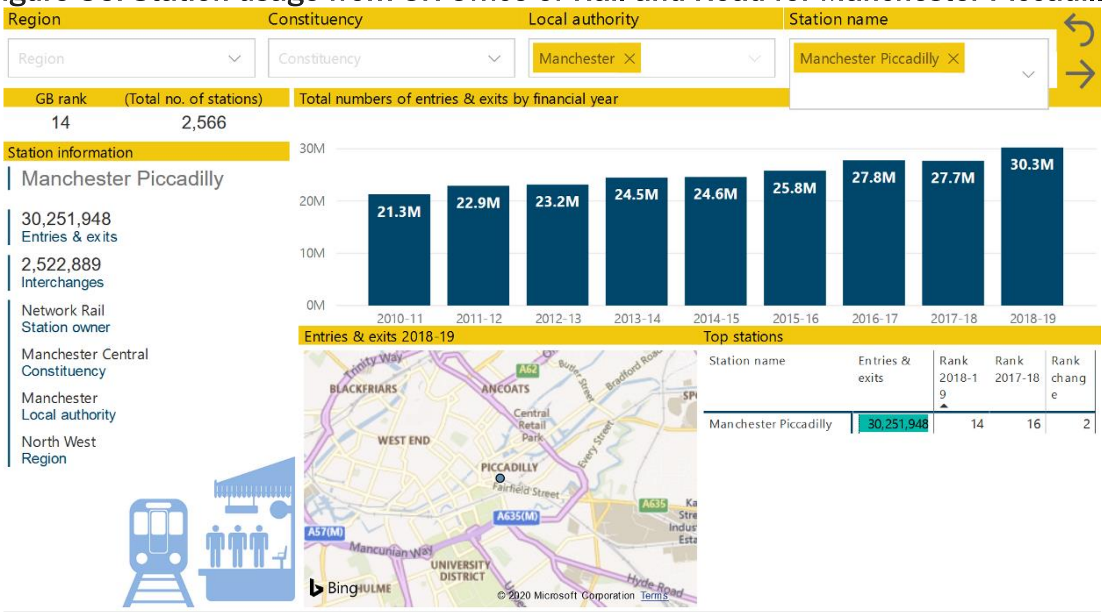
In the NSW GTFS, there are 39,062 mass-transit stops within NSW. Of those, 24,754 are within Sydney and there are 205 train stations.

3.5.5 Aggregation examples

Although this thesis focuses primarily on NSW, there are interesting datasets generated and released in other jurisdictions – especially Singapore which has a TOTOR ETS (Ma et al. 2013; Sun et al. 2012; Zhong et al. 2014; Zhong et al. 2015).

The United Kingdom now releases a dataset for station entries and exits along with detected interchanges for every station in Great Britain aggregated over a year. The duration of this activity dataset is clearly privacy safe and the data is provided both in machine readable files and is accessible from an online dashboard.

Figure 56: Station usage from UK Office of Rail and Road for Manchester Piccadilly



Source: (Steer 2020) https://dataportal.orr.gov.uk/statistics/usage/estimates-of-station-usage

3.6 Summary

Current NSW common law has found that a person's travel history is personal information and must be protected. Planning for privacy protection is an essential early step in the use of empirical datasets such as Opal. Without privacy protections, usage of the data would not be approved by the state, forgoing the opportunity of pursuing the objectives of improvements.

UTS developed a set of guidelines for its researchers to ensure ethical conduct that protects confidential and private information of customers. A key principle in those guidelines was the Need-to-Know information security framework that should be implemented by researchers and service partners when analysing large empirical datasets. Analysts must restrict themselves to utilising only data they require.

One way to implement the Need-to-Know framework is to use derived information in preference to raw data. Derived information can be created through the use of four strategies of privacy protection — deidentification, disassociation, aggregation and elimination. By using these strategies, it is possible to transform datasets that contain biographical information into datasets that only represent activities. Activity datasets describe what occurred on a transport system — in that way they describe the customer response to the transport services delivered. Activity datasets can be privacy-protected and be used safely by transport operators and trusted partners.

Deidentification is a key component of controlling the risk of privacy breaches and is built into the Opal system by the segregation of the registration from the travel history database. It is advised to apply further deidentification by removing all unique card indicators to further control the risk of a privacy breach. During most analysis it is not required to examine complete histories and since the analyst does not Need-to-Know they should remove identifying fields.

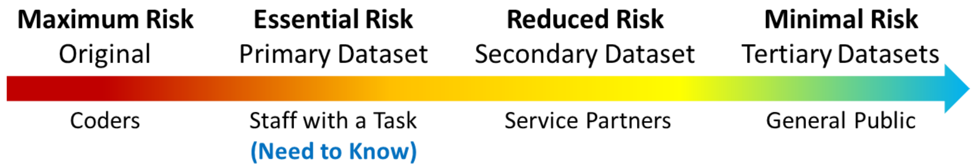
Aggregation is another key method that can be used to control risk such as counting the number of journeys in time, space, or other dimensions such as mode. A positive side-effect is that it becomes easier to undertake useful analyses due to the smaller datasets.

The following chapter — Chapter 4 — describes a method for implementing privacy protection by creating derived datasets using the techniques discussed in this chapter.

CHAPTER 4 – DATA PREPARATION

This chapter describes a method using three stages of transformation upon a TOTOR dataset to control the risk of privacy violations. The method processes the original dataset using the techniques of deidentification, disassociation, aggregation, and elimination to mitigate privacy risks that were described in Chapter 3. This method creates three tiers of datasets for use by the operator, their trusted partners, and finally the wider community.

Figure 57: Using Need-To-Know and Derived Datasets to Control Privacy Risks when creating Empirical Datasets



This method has been created to provide a clear framework to assist TfNSW and overcome the methodological barrier demonstrated during their loss in Waters v TfNSW [2018] NSWCATAD 40. The method described leverages the Need-To-Know framework by creating datasets for decreasing levels of trust from the original dataset.

The method outlined in this thesis, begins by developing comprehensive stable datasets used to describe the transport service delivered and the environment in which they were delivered. These datasets include comprehensive historical records of the transit stops and bus runs. Additional stable datasets should include information on the weather, public- and school-holidays, and special events because all of those factors impact the customer's demand for transport or the customer's response to the services delivered.

The method then constructs a high-resolution activity dataset that uses aggregation and deidentification to control the risk of privacy violation. The high-resolution primary dataset accurately measures passenger response to the public transport service delivered. This high-resolution dataset can be used for many analysis methods as shown in the case studies in Chapter 5 and is suitable to use as a basis for an empirically driven approach. Since the primary dataset has not undergone elimination it will retain information on unique activities and should still be considered privacy sensitive.

Next the method transforms the primary dataset into secondary datasets by introducing more privacy protections. These secondary datasets contain accurate high-resolution data for use by the operator and trusted service partners. These datasets will have been treated with disassociation and/or aggregation in order to remove exact travel histories. These datasets would inform the railway operator about the historical passenger responses to the services offered without using passenger biographies. For example, the primary dataset could be transformed into the number of embarkations per minute for each railway station in the network. Another example would be the number of persons disembarking at each station per minute broken down by their embarkation line, such as X persons disembarked at Town Hall at 08:05 from the North Shore Line.

Finally, the method creates privacy-safe tertiary datasets for release to the general public using all methods of privacy protection. Example datasets include the number of embarkations per month for each light rail stop or ferry wharf. There are many additional datasets that could be released to the general public and some are described in this chapter and the next chapter.

Processing speed is proportional to the number of records. Therefore, even the minimal aggregation used in the primary dataset will allow more analysis by operators.

Implementing this method allows the operators and trusted subcontractors to protect privacy while leveraging Big Data to transition to an empirically driven approach as demonstrated in the upcoming case studies. This method will allow them to measure the services delivered and the customer response to the services delivered.

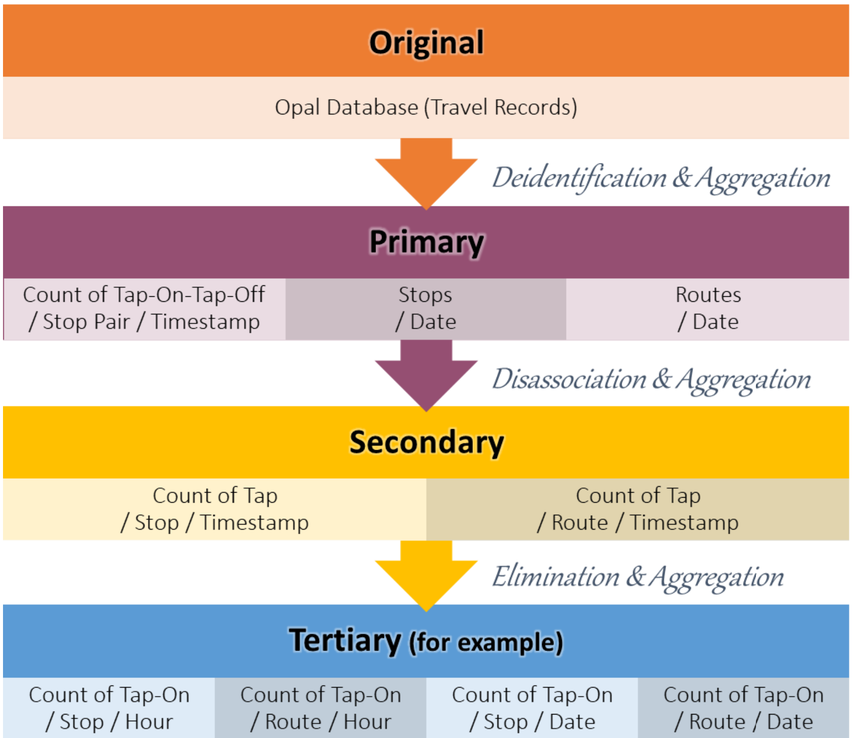
4.1 Derived datasets for privacy protection and empirical analysis

One approach to implementing the Need-To-Know framework is using derived data rather than raw data. Derived data can be created by deidentification, disassociation, aggregation, and elimination. It is possible to generate high-quality, high-resolution datasets derived from the original Opal dataset but using transformations to remove the biographical travel histories from the dataset.

The primary tier of activity datasets would contain actual counts in order to maintain accuracy, so these datasets would be used only by trusted individuals on a Need-To-Know basis. From these accurate high-resolution datasets, further derived datasets could be created through further aggregation and elimination to ensure sufficient privacy protection. These secondary activity datasets could be distributed to associated service partners, such as other agencies and responsible researchers. After further processing, tertiary datasets could be distributed to all service beneficiaries and the general public.

Figure 58 below provides a heuristic to show how the derived datasets can be created in an iterative process moving from completely biographical travel histories to increasingly privacy protected activity datasets.

Figure 58: Possible Structure of derived data generation



The first step in creating risk-controlled datasets is deidentification (disassociation from the identity dimension). That is done by the separation of the PAS and Opal Databases.

Removal of the relatively unique identifiers of card foreign key and hashed card number[25] from the Opal dataset (disassociation) creates a dataset that no longer stores biographic travel histories of customers. This disassociation makes it impossible to unambiguously determine that a combination of journey segments relate to any single card let alone to any specific individual. Simply removing card identifiers from the dataset scrubs out the travel patterns and leaves only activity patterns.

[25] Analysis of Opal datasets indicates the hashed card number may not be perfectly unique.

An important class of primary dataset is the stable repeated data that does not change often or has not ever changed. To save space and facilitate analysis these stable datasets could be extracted and provided separately as reference datasets.

For example, Sydney's Westmead Station[26] opened on the Main Western Line in March 1883 at (150.9876°, -33.8084°) at 25.1 km distance from the zero mark. The station has not moved in 136 years. Repeating its coordinates in multiple records is unnecessary.

4.2 Contextual datasets

To understand the passengers' response to the transport service delivered, it is necessary to consider the environmental and social context in which the response occurred. Patronage will decrease on days when people are on holidays, or when the city is under a pandemic lockdown, no matter how perfectly the transport services were delivered.

The dates of public and school holidays in each state are an important social context for analysing patronage. For example, during the summer school holidays many families travel from the cities to the coast for holidays, especially after December 25th.

The method proposed has been derived from experience and was designed to reduce the effort needed for transport analytics. The approach leverages Left Join with databases.[27]

4.2.1 Holidays & weather

To allow interpretation of the passenger response measured in activity datasets, it is important to create and distribute informational datasets to describe the environmental and social context in which the activities were recorded. The operators should produce master datasets that describe the social and environmental context for each date covered in activity datasets. Suggestion of the data recorded for each date are as follows:

- Date
- NSW Public Holiday Name (if applicable)
- Was it a NSW School Holiday
- Victorian Public Holiday Name (if applicable)
- Was it a Victorian School Holiday, and so on for all states

[26] https://www.nswrail.net/locations/show.php?name=NSW:Westmead

[27] See any reputable text or website on using databases: https://en.wikipedia.org/wiki/Join_(SQL)#Left_outer_join

- Was there track work, and where was the track work
- What was the NSW time zone (was it during daylight savings)
- Sunrise time stamp
- Sunset time stamp.

In addition, there is a need to produce a dataset that records major events, especially those that generate large event crowds (demand), like AFL games, the Vivid Art Festival, the Sydney to Hobart Yacht Race, or indeed any major protest.

For example, IWLR patronage is relatively steady across the week including on the weekends, however the patronage varies with factors such as holidays. Table 25 below shows that the IWLR had four thousand fewer passengers on Good Friday 2016 than on an average Friday for that March. In contrast, the Easter Sunday patronage was two thousand passengers above average and was the highest patronage for a Sunday in that March. The 'Diff to March Mean' and 'Diff to March Max' columns show the difference in patronage for the day relative to the mean/max of that day of the week for that March. Thus, Friday 18 March 2016 had higher patronage levels than the average Friday and less patronage than the maximum for March.

Table 25: IWLR Patronage over the Easter Holidays in 2016

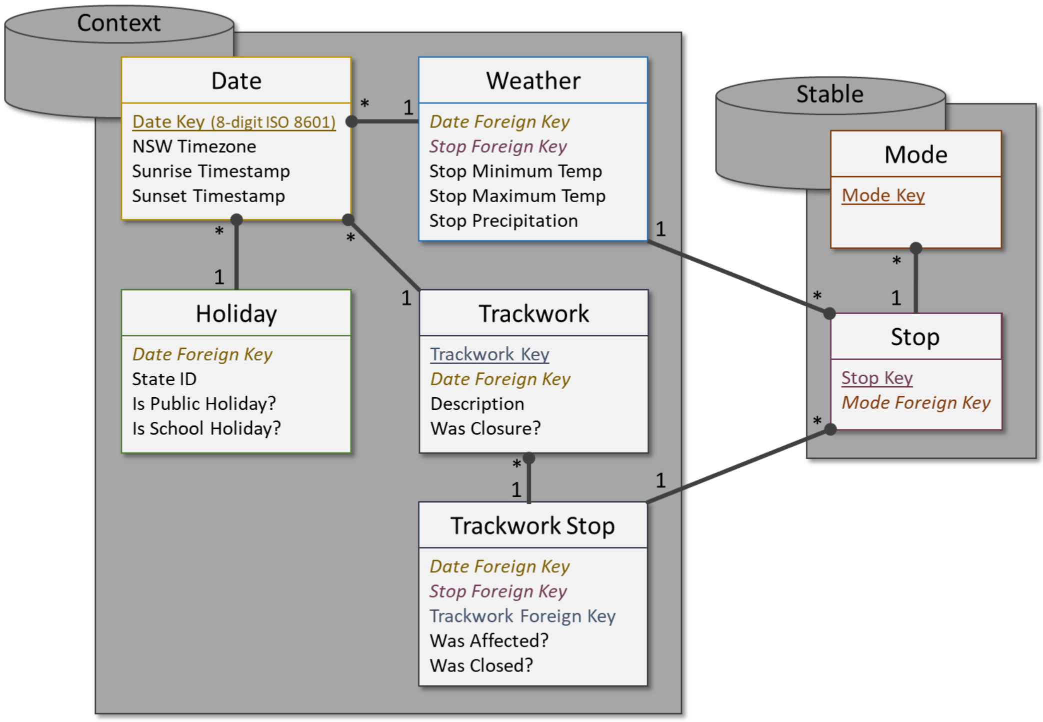
Date	Day Name	Count	Holiday Name	Diff to March Mean	Diff to March Max
2016-03-18	Friday	26,270		1,510	-350
2016-03-19	Saturday	20,300		-730	-1,780
2016-03-20	Sunday	22,820		-1,300	-3,690
2016-03-21	Monday	42,060		1,450	-7,590
2016-03-22	Tuesday	27,280		-550	-3,590
2016-03-23	Wednesday	23,960		-890	-2,210
2016-03-24	Thursday	26,220		1,030	0
2016-03-25	Friday	20,170	Good Friday	-4,590	-6,450
2016-03-26	Saturday	20,760	Easter Saturday	-270	-1,320
2016-03-27	Sunday	26,510	Easter Sunday	2,390	0
2016-03-28	Monday	22,720	Easter Monday	-17,890	-26,930
2016-03-29	Tuesday	30,870		3,040	0
2016-03-30	Wednesday	26,170		1,320	0
2016-03-31	Thursday	24,520		-670	-1,700

Source: TfNSW Light Rail Full Financial Year

Weather also changes people's travel behaviour, therefore an environmental dataset that records minimum and maximum temperatures as well as precipitation for every day at each weather station is needed for comprehensive analytics. A dataset that provides a mapping from transit stop to weather station is needed to enable analysis of passenger response within the actual local weather context.

For example, during FY 2015-16 the recorded adult patronage on the IWLR did not vary substantially in response to precipitation or heat. However, recorded school student patronage dropped substantially with anything more than minor precipitation.

Figure 59: Context dataset showing weather, track work, and holidays.

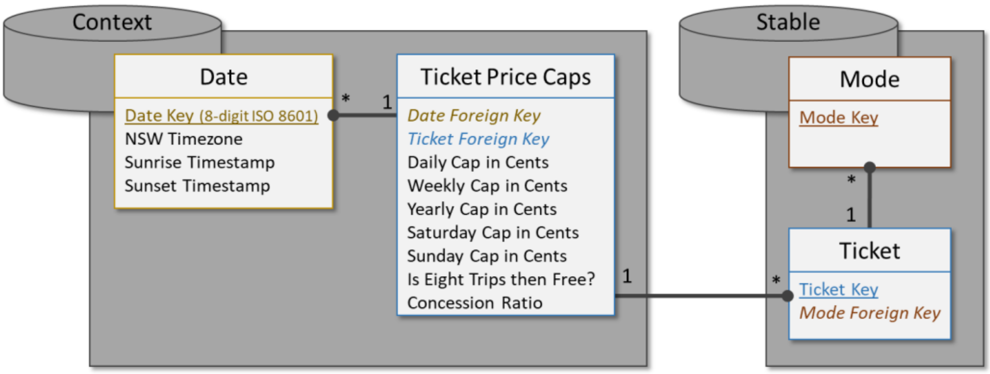


4.2.2 Pricing

Humans alter their behaviour in response to the sticker-price of a transport service. Humans are usually poor at perceiving the full cost of transport and due to cognitive biases (e.g. loss-aversion), they treat the visible costs (e.g. ticket price) as more significant than invisible costs like car maintenance (ATC 2006, p. 20; Neumann & Böckenholt 2014, p. 185; SACTRA 1994, p. 16).

Therefore, it is important to record and describe the pricing history for all customers described in the activity datasets. Since the pricing structure will vary over time a table should be provided to record the pricing structure applicable to all customer classes for each date. At a minimum, the service partners must keep a public record of the pricing structures and how they varied over time, even as descriptive text. For example, transport analysts from 2026 would be very confused by the Monday patronage levels on the IWLR from March 2016, if they were unaware of the fare structures, see Fare Minimisation discussed previously.

Figure 60: Example context database for ticket price caps.



4.2.3 Provide a record for every date and location

There is an important caveat that needs to be remembered when designing analytics tables – computers are only information processing systems and are incapable of reasoning to deduce missing information. Therefore, the designers of stable tables should be explicit about everything in machine readable formats, rather than rely on a rule of inference written in a human language like English.

From the computer's perspective, a lack of data does not represent a negative state – lack of data only represents missing information, uncertainty, or worse ambiguity. Consider the example with the holiday database table described previously. If there is no holiday record for 2019-10-01 (date field) and Queensland (state field), a transport analyst cannot say there were no holidays in Queensland on that date, only that the analyst has no information on whether holidays were occurring or not.

Ambiguities add significant complexity (cost) to an analysis and require assumptions to compensate for the uncertainty. Analysts must construct code to fill in the blanks based on assumptions of the database contents. Assumptions are a significant source of errors, and so every assumption should be considered a risk. To simplify the analytics and prevent incorrect assumptions, it is preferable for the dataset to explicitly record the context for every date, e.g. recording the holiday status for every date and every state.

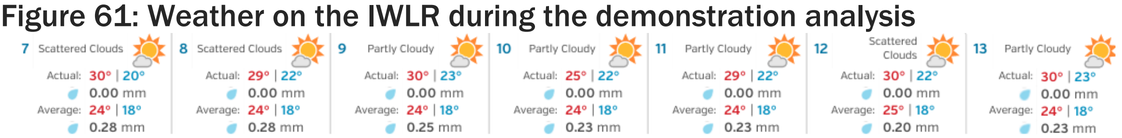
The need for a complete historical context data is especially important for transport analytics. The high quality of the Opal dataset will facilitate long-term trend analysis, but analysts in 2030 will need a clear description of the historical context. Future analysts will not have a complete description of fare structures, public holidays, timetables, or weather, unless one is provided for them. Without an accurate description of the services promised and the environment for the operations, any analysis of passenger response will be inaccurate and lack empirical rigour.

Providing an easy-to-use contextual database increases the likelihood that the analysts will include factors such as holidays and track work in the analysis; see Figure 59 above. Human nature and deadlines mean the more effort required to find contextual information, the less likely the context will be considered.

Provide unambiguous descriptions of the environmental and social context for every date facilitates analysis of any arbitrary date range using the widely available tools for database joins. When joining two datasets, if the computer finds no explicit record it cannot infer meaning and will simply leave a gap in the results table. Missing data forces an analyst to manually address the missing context, uncertainties, or ambiguities.

For example, pick a random date: like Monday the 7 March 2016 — the first date analysed for Figure 5 in the Section 1.1.1. What is the best weather station for Lilyfield? Was that date in the school holidays? What about the Victorian or Queensland holidays?

Figure 61: Weather on the IWLR during the demonstration analysis



Source: • Bureau of Meteorology

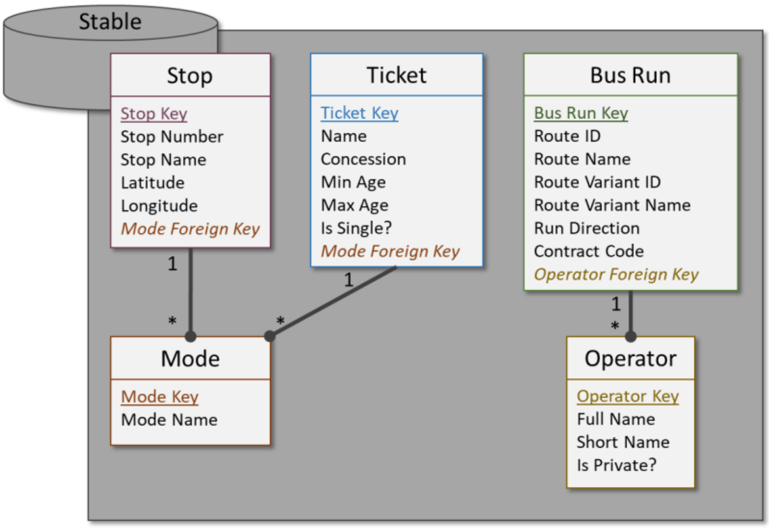
Certain TfNSW practises complicate the use of datasets as they require analysts to be aware of the history of transport locations within NSW. One such practice is the TfNSW policy of reusing what should be unique identifiers to refer to differing entities. For example, TfNSW has reused the same stop number for very different bus stops. Another practice that can cause confusion is TfNSW changing identifiers over time to refer to the same entity. For example, the department has datasets that refer to Wickham station (in Newcastle). After the 2014 removal of the inter-city and suburban rail line from Newcastle, the government changed Wickham's identifier to Newcastle Interchange. These issues create ambiguities and require the creation of datasets to map differing identifiers to a single unique master stop identifier for every date it was in use.

Complete datasets that record the context for every date and location reduces the need for analysts to find data and removes the risks of incorrect assumptions. A comprehensive and unambiguous contextual database improves the likelihood of comprehensive analysis.

4.3 Organisational datasets

To limit dataset size, minimise repetition, and reduce the burden of dataset creation, the proposed method leverages stable datasets that include the mode, stop ticket, operator, and bus route tables for example, as shown in Figure 62 below.

Figure 62: Model of the stable datasets referenced by the primary dataset

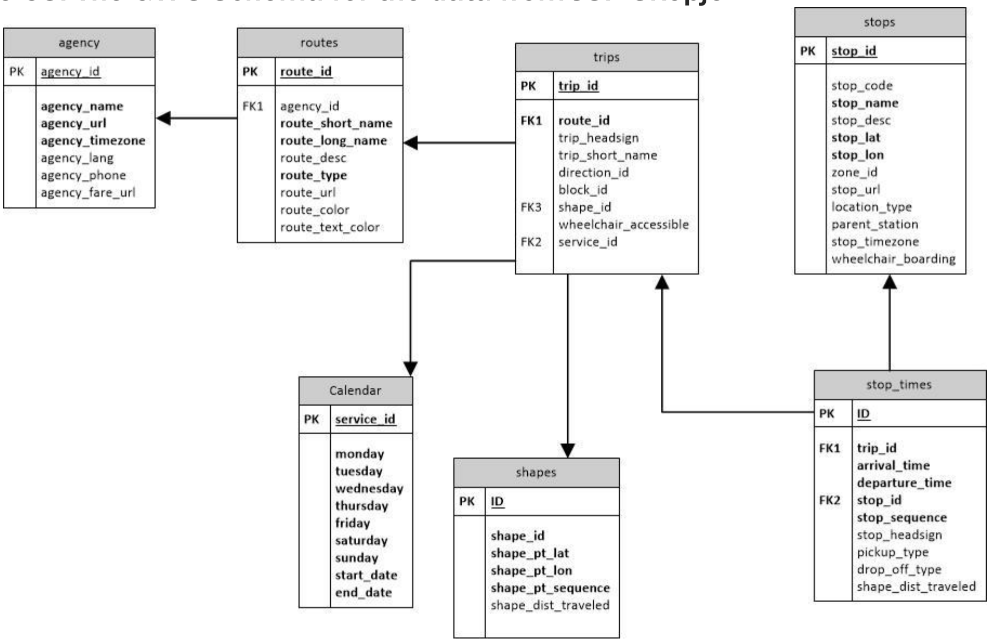


The service quality targeted by an operator is expressed in their timetables, which specify routes and departure-times / vehicle-headways. One type of machine-readable timetable is the General Transit Feed Specification (GTFS) which allow mapping, routing, and mobility services to display stops and services, and to provide trip planning (pathfinding). The GTFS accessed through the NSW Application Programming Interface (API) allows information systems to interpret and display the timetables. Transport analytics requires the use of the timetable order to interpret the services targeted, the services delivered, and the customer response.

4.3.1 Stops

GTFS encodes the timetable using multiple machine-readable text files representing a specific database (Entity Relationship Model). A key database table is the stop descriptions in stops.txt. Each record has the stop identifier, a human readable name, and the stop's latitude and longitude. The stop database table is a perfect example of stable repeated data because the data is not planned to change during the day.

Figure 63: The GTFS Schema for the data from JSP Skopje



Source: (Mishevska et al. 2014, p. 3, Figure 1)

In Figure 58, the stops and routes were identified as separate and important tables for service partners and the public. The NSW GIPA Act under Section 3 (Object of Act) requires that '*access to government information is restricted only when there is an overriding public interest against disclosure*'. Those tables contain no individual's data and refer only to the services targeted. As such these organisational tables should be released to the public.

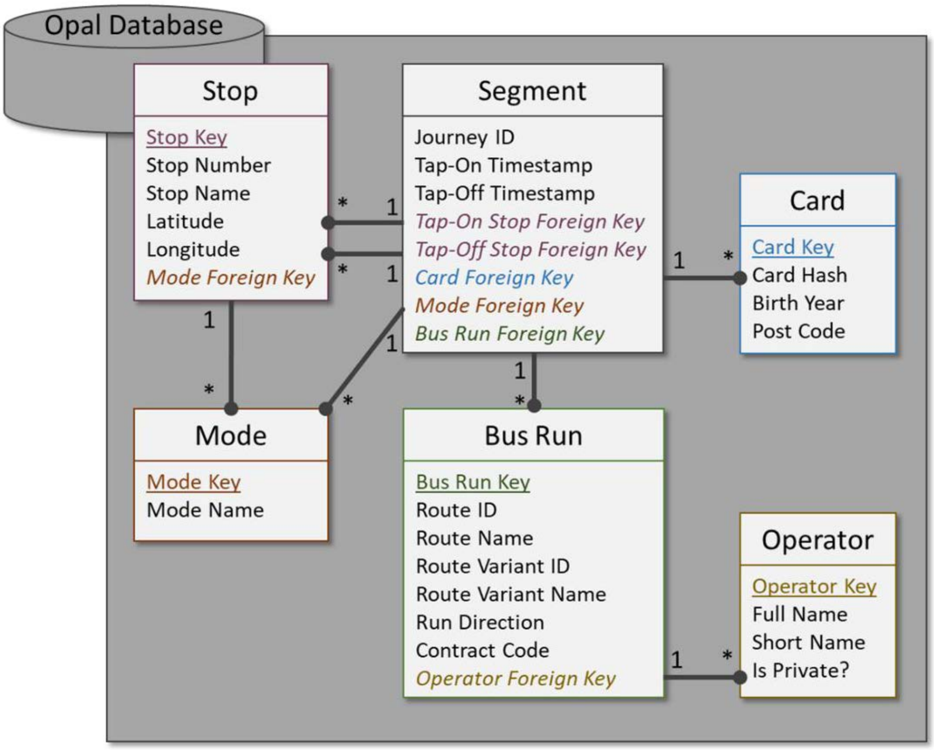
Figure 64 below shows a more complete diagram of the Opal Database. The designers used an Entity Relationship Model and created separate database tables to store the stable repeated data, such as transit stops, operator names, and bus route identities.

NSW changes and reuses identifiers resulting in the stop number and stop name being ambiguous over time. The Opal Database contains a distinct key field for each stop in addition to the stop name and the stop number. To facilitate historical analytics and cross-referencing an additional database table listing all stop keys accessible for all dates would be useful, it would contain just two fields — the stop key and a date.

Additional context for each stop would assist analysts, including the address, postcode, accessibility status, locations of exits, suburbs, and the ABS ASGS areas served.

Another very important context for each stop is its relative position to major economic zones such as the Sydney or Parramatta CBD. Having a unified stable definition of what is in the Sydney CBD would assist transport analysts enormously.

Figure 64: Conceptual structure of Opal Database



Source: (TfNSW 2015)

4.3.2 Routes

Since prices in NSW are based on distance travelled, the Opal system has ticket readers on-board the buses, and the ETS has to know the bus route and all scheduled stops along that route to calculate the ticket price. With each recorded journey segment, Opal stores a foreign key relating to the table describing the bus routes, or it stores -1 if the segment was on a train, tram, or ferry. This information is highly valuable in interpreting the service delivered and customer response because it allows the service partners to analyse the usage of individual routes and indeed individual bus runs.

Since NSW reuses route identifiers, to facilitate analytics it would also be desirable to create a date to run mapping, similar to the stop to date mapping mentioned above.

Table 27 below provides an example of the route datasets that would be useful to analysts. The operator is stable repeated data, so it should be a separate dataset.

Table 26: Extract of an NSW Operator database table

Operator Key	Short Name	Full Name	Is Private?
1	Syd	Sydney Buses	Not Yet
3	BBT	Busways Blacktown	Yes
4	BGT	Busways Gosford	Yes
19	TDV	Transdev	Yes

Table 27: Extract of NSW Routes Table database table

Bus Run Key	Route ID	Variant ID	Variant Name	Run Direction	Contract ID	Operator Foreign Key
2435525	363	363-19	-1	2	2451	12
2525153	53	53-64	-1	3	2452	4
2534971	38	38-58	-1	3	2452	4
2535629	M92	M92-57	-1	1	2442	19
2726670	962	962-26	-1	1	2442	19

4.3.3 Card types

For businesses it is important to segment the customers into groups. This facilitates identification of the factors driving different service qualities sought and thus different responses from each market segment to the delivered transport (Bartels 1988).

For example, NSW half-priced concession tickets are for eligible tertiary students, apprentices, trainees, and job seekers. The analysis below shows that the behaviour of concession card holders differs from the behaviour of adult card holders.

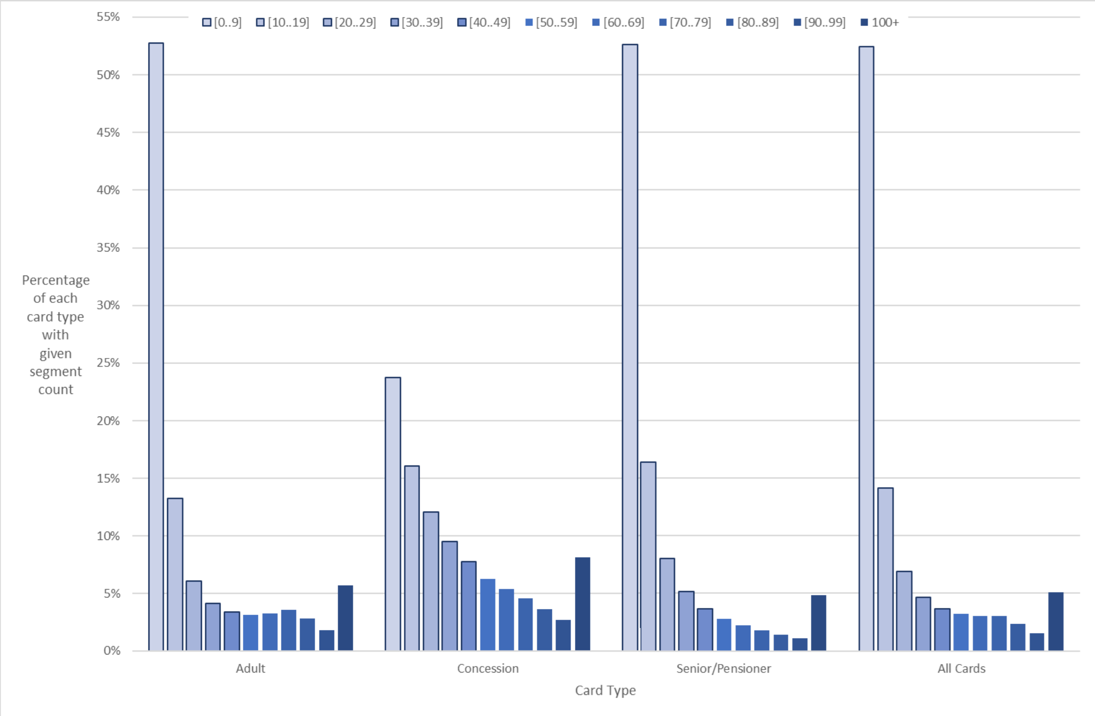
Table 28: Usage statistics on segments per Opal card in six weeks

Number of Segments per Card	Median	Average	Max	Standard deviation
Child/Youth	5	10.65	351	17.43
Free Travel	7	19.95	395	31.57
Adult	8	26.54	765	37.57
Senior/Pensioner	8	23.50	906	37.61
Employee	9	24.98	1,065	35.80
School Student	22	30.85	273	30.11
Concession	28	39.66	730	38.31

Source: (Hounsell 2017a, p. 10, Table 3)

Table 28 above shows the median concession card holder undertook a higher number of journeys than the median adult (Hounsell 2017a). The mean number of trips was broadly similar for Adult and Senior's cards. Adult usage was highly skewed towards zero, with the majority of cards used only 2-4 times over six weeks – Figure 65 below

Figure 65: Journey segments recorded per card type over six weeks



Source: (Hounsell 2017a)

The above patterns might reflect wider global trends such as a reduction in the derived need for private cars in order to access intrinsic needs such as food, social engagement, or love (Rosenfeld, Thomas & Hausen 2019). Alternatively, the behaviour pattern might reflect local changes such as increased congestion, road usage costs, and decreased disposable income. The patterns may represent the self-optimising behaviour of students as they undertake their education at schools and universities. The trend of younger people choosing to drive less has been occurring since 2011:

> 'Younger people driving less: As compared to 2001/02, people aged 30 years and below were less likely to drive in 2011/12. Those aged over 60 were more likely to drive in the recent year. PT used most by younger travellers: Public transport use was highest for those aged 11-30, the years associated with schooling and comparatively low access to vehicles.' (BTS 2013, p. 26)

The behaviour of different market segments identified by different ticketing categories have different behaviours. Further research is warranted into this topic as such analysis represents the shift towards a customer-centric approach based on empirical measurements to understand the customer. Therefore, an additional organisational table should be created to allow the aggregation of travel data by ticket type. Table 29 below provides an indicative table that could be used in the creation of other tables.

Table 29: Indicative Card Type Table

Ticket Key	Is Single	Simplified Type	Concession	Opal Card Type Labels
1	Multiple	Adult	Adult Price	Adult
2	Multiple	Child/Youth	Child Price	Child/Youth
3	Multiple	Concession	Concession	Concession
4	Multiple	Special	Special	Employee or Free Travel
5	Multiple	School Student	School Student	School Student
6	Multiple	Senior/Pensioner	Concession	Senior/Pensioner
7	Single	Adult	Adult Price	Sgl Trip Ferry Adult, Sgl Trip LR Adult, Sgl Trip Rail Adult
8	Single	Child/Youth	Child Price	Sgl Trip Ferry Child/Youth, Sgl Trip LR Child/Youth, Sgl Trip Rail Child/Youth
9	UNKNOWN	UNKNOWN	UNKNOWN	UNKNOWN

4.3.4 Intermodal transfers

In discussions with TfNSW, the topic of intermodal transfers came up repeatedly. Intermodal transfers represent less than 10% of all journeys and were considered relatively unimportant. Over the entire Opal coverage area, that is technically true, however, intermodal transfers are vital at key locations.

Table 30 below shows the major locations for intermodal transfer throughout the Sydney GMA. The Top-10 locations are served by significant bus/ferry corridors, with facilitates and timetables designed to assist interchanging. Passengers only perceive a single transport network, therefore these interchange assisting services designs (targets) encourage intermodal interchange.

For example, it is substantially faster to transfer from bus to train at Bondi Junction to access destinations in the Sydney CBD than to stay on buses that travel slowly in mixed traffic along Oxford St.

Some locations have a high proportion of interchanges in their daily patronage. Almost two in five embarkations at Bondi Junction are transfers from a bus, and over one in seven embarkations at Circular Quay are transfers from ferries. To allow assessment of passenger behaviour it is essential to count the number of intermodal transfers.

Despite its location within the middle of the CBD, even Town Hall Station serves intermodal interchanges for over 10,000 journeys on the day surveyed.

Table 30: Top-10 Locations for Intermodal Transfers on 2016-11-16

Intermodal Location	Number of Segments	Percentage of Continuations
Central Station	42,270	18.1%
Bondi Junction Station	20,690	8.8%
Wynyard Station	15,440	6.6%
Parramatta Station	12,530	5.4%
Circular Quay	11,340	4.9%
Town Hall Station	10,400	4.4%
Blacktown Station	6,780	2.9%
Strathfield Station	6,310	2.7%
Hurstville Station	4,330	1.9%
Burwood Station	4,230	1.8%

Source: • (Hounsell 2017a)

Table 31: Intermodal Transfers at Key Station on Wednesday 2016-11-16

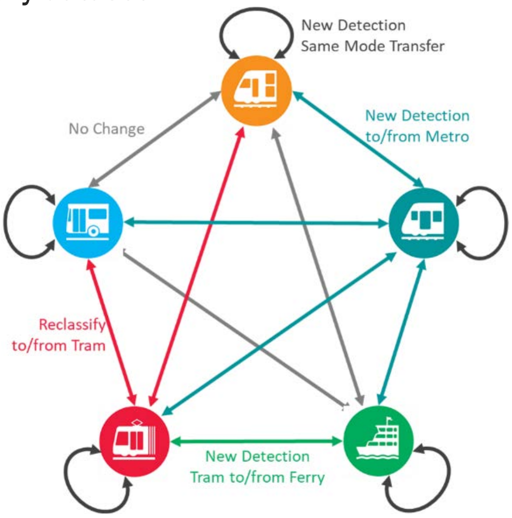
Intermodal Continuation	Circular Quay Station	Parramatta Station	Bondi Junction Station
Bus to rail	400	6,080	10,830
Ferry to rail	4,420	20	
No inter-modal transfer	24,980	28,780	17,220
Grand Totals	31,010	36,320	28,770
Intermodal transfers %	15.5%	16.8%	37.6%

Source: • (Hounsell 2017a)

The Opal system detects if multiple trips across different modes were used to undertake a single journey. The system records trips as a continuation and adds a flag to each record indicating the type of transfer used. However, not all combinations are recorded uniquely, for example, to/from a tram is recorded as an interchange with a ferry. Only flagging continuing embarkations means a simple count of continuations over-counts the number of journeys and under-counts the number of trips.

Intermodal continuation flags do not capture transfers without a mode change. To empirically analyse customer behaviour, it is preferable to record all inter-modal transfers and also to record intra-modal transfers, as shown in Figure 66 below. For example, a person catching an eastward 333 to Bondi Junction and transferring to a southward 400 should be recorded as a transfer.

Figure 66: The Pentagram of Transfers – new intermodal detections desired for the primary dataset



Since the creation of the primary dataset will be a post-processing step, it is possible to add additional processing. The transformation program should detect and classify transfers more accurately than the ETS. Table 32 below represents the mode table for the primary dataset. Zero is reserved as a flag to be used in recording non-transfers.

Table 32: Extract of Mode Table

Mode Key	Mode Name
0	NONE
1	UNKNOWN
2	Bus
3	Ferry
4	Light rail*
5	Train
6	Metro†

* Light rail in some existing data sets is capitalised as Light Rail, which can be a problem for collation.

† Sydney Metro could be described as a separate railway mode in the patronage counts released.

Transfers are relatively common, 36% of cards were used for at least one intermodal continuation during a six-week survey in 2016. Further research should investigate the percentage of cards that undertook an intermodal transfer over a full year in NSW.

4.4 Primary dataset

The method proposed by this thesis centres on the creation of a primary dataset that is rich enough to describe customer response, but which has been designed to protect customer privacy. Deidentification, disassociation, and aggregation are used to create a derived dataset that controls the risk of privacy violation by transforming the original into impersonal activity records. The purpose of the primary table is to create a high accuracy data source that describes travel activities on transport systems without customer identities and biographic travel histories. The primary tables are high accuracy and enable further processing to generate derived datasets for service partners to assess customer behaviour and transport operations while still protecting customer privacy.

The primary dataset is relatively simple and counts the number of cards undertaking a unique Tap-On Tap-Off pair at any given minute – Table 33 below. The dataset will refer back to the context databases. Since the bus run information is readily available, referencing that data adds value at minimal cost. Counting by ticket categories allows analysis of the behaviour of market segments. Recording the standard price of that pair allows assessment of each market segments response to the prices charged.

Table 33: Primary Dataset – Segments Count Table

Field Name	Description
Segment Key	An identifier to reference this combination from other tables
Tap-On Timestamp *	Date and time of the Tap-On for these segments
Tap-Off Timestamp	Date and time of the Tap-Off for these segments
Tap-On Stop Foreign Key †	Reference to the Tap-On location
Tap-Off Stop Foreign Key ‡	Reference to the Tap-Off location
Ticket Foreign Key	Reference to the ticket type for these segments
Bus Run Foreign Key	Reference to the route name table; if applicable
Journey Segment Count	The number of journey segments for this combination
No Intermodal Transfer §	The number of segments which were not transfers
Standard Price	The price of a standard fare between those stops at that time

* A time stamp record contains the date and time as ISO 8601 – 2019-12-29T05:58:52Z. Ideally time should be stored as UTC; however, the Opal system stores Sydney time without time zone information.

† The mode for the Tap-On stop will be the mode for this journey segment. No clean segment should have two different modes.

‡ Time stamps can be NULL (empty) if the Tap-On stop or Tap-Off stop was not known.

§ This count is for the convenience of analysts.

As discussed above, analysing transfers in the network is essential to an empirical analysis of customer behaviour in response to the transport network delivered. For example, in 2016, tram users were the most likely to be multi-modal with 91% using trains and 68% using buses. Multi-modal behaviour is not symmetrical – 81% of ferry users used trains, but only 16% of train users used ferries (Hounsell 2017a).

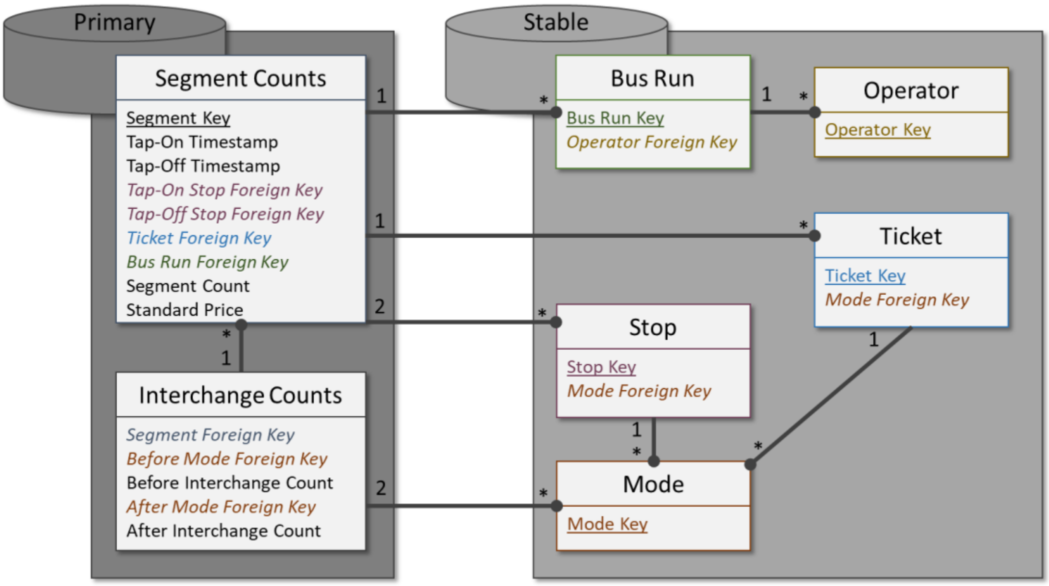
Analysts need to know the mode passengers used on previous and subsequent journey segments. However, the design should conserve storage space and processing time, thus transfers should be separated into an additional table. Table 34 below outlines the structure for the primary interchange / transfer database table.

Table 34: Primary Dataset – Interchange Count Table

Field Name	Description
Segment Foreign Key	The journey segment this interchange count refers to
Before Mode Foreign Key *	Mode of the segments used before this segment; can be NONE
Before Interchange Count	The number of journey segments for this before combination
After Mode Foreign Key *	Mode of the segments used after this segment; can be NONE
After Interchange Count	The number of journey segments for this combination

* If both modes are NONE then both counts are zero. There is no reason to have an entry for that pair.

Figure 67: Model of primary tables and relationships to the stable tables



The primary dataset contains no direct biographic information, only indirect activity records. Their accuracy and resolution could theoretically be used by an individual for undesirable activities — that perception represents a reputational risk.

The transformations for the primary dataset are sufficient to mitigate the main risks of privacy violation. The primary dataset provides a starting point for the analytics programme within the service partners under a Need-To-Know framework.

For example, the primary dataset could be filtered to return only the embarkations at Town Hall station over 2017-2018. That activity dataset could assist with behavioural analysis for the development of cyber-physical systems to support management of crowds at this crucial station in Sydney — as in 1.3.5 Minute by Minute. That activity dataset would allow estimation of passenger demand for arriving and departing platforms with reference to holidays, weather, and major events like the Vivid festival.

A high accuracy primary dataset allows easy correlation with other high-value datasets. For example, the CoA census includes the Journey to Work (JTW) question. The JTW can be assessed by the ABS ASGS, transport mode, occupation, income and industry. Thus, the ability to correlate the primary dataset with the census could allow a priceless analysis of customer behaviour. The next census is due at the start of August 2021.

4.5 Secondary Datasets

Retaining accuracy in the primary dataset, allows the derivation of accurate secondary datasets from the primary dataset. These secondary datasets will provide significant and useful information for trusted service partners. Maintaining accuracy in the primary dataset allows future development of novel datasets, methods, and applications.

The second dataset is the aggregation of the primary table into counts of Taps (either On or Off), as well as aggregating the associated counts of interchanges. This step breaks apart the stop-pairs by disassociating the Tap-Off from the Tap-On. Repeated disassociation and aggregation will further control the risk of privacy violation.

The structure of the secondary database table is listed in Table 35 below. Note that the tap type has been added to differentiate between Tap-Ons and Tap-Offs. Also note that the two tuples of (time stamp, stop reference) have been replaced by a single tuple.

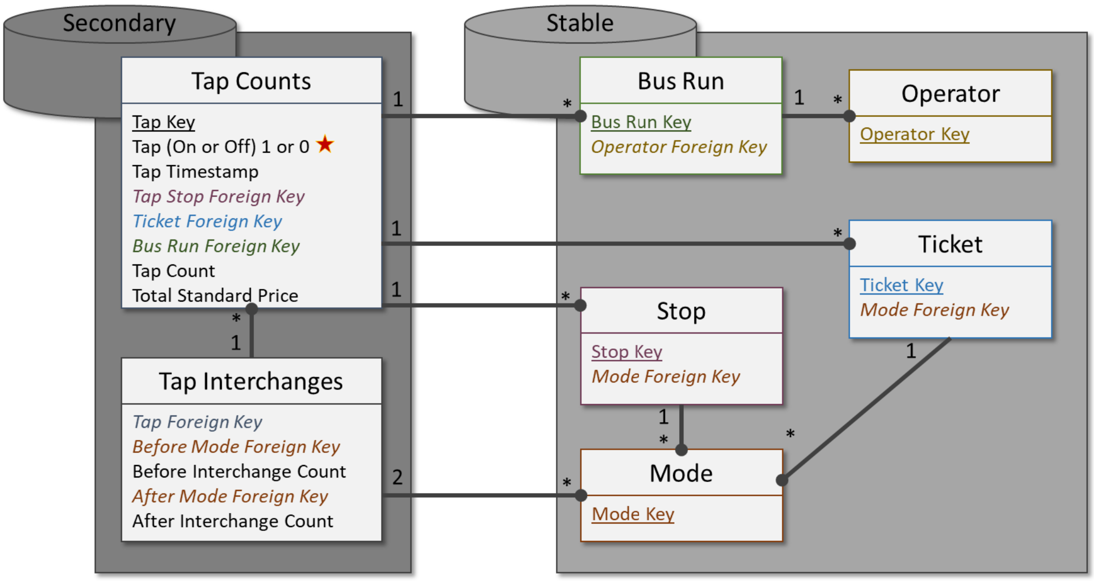
In NSW, the ticket price varies based upon the embarkation-disembarkation pair, therefore it is no longer possible to record the standard price, but it is possible to calculate the total standard price that theoretically could have been collected from the passengers counted.

Table 35: Secondary Dataset – Tap Count Table

Field Name	Description
Tap Key	An identifier to reference this combination from other tables
Tap Type	On or Off; possible encoded as 1 or 0 respectively
Tap Timestamp	Date and time of these Taps
Tap Stop Foreign Key	Reference to the location for these Taps
Ticket Foreign Key	Reference to the ticket type for these Taps
Bus Run Foreign Key	Reference to the route name table; if applicable
Tap Count	The number of Taps for this combination
No Intermodal Transfer	The number of Taps which were not transfers
Total Standard Price	The total price for standard fares for these Taps

As can be seen by comparing Figure 68 below (structure of the secondary dataset) and Figure 67 above (structure of the primary dataset), the stability of the operational datasets allows both derived datasets to refer to the same stable tables.

Figure 68: Conceptual diagram of secondary database tables



The primary interchange dataset is associated to individual segments, whereas the secondary interchange dataset is associated to a particular combination (tuple) of tap-type, stop, and time stamp. Therefore, although the structure is similar, a secondary interchange dataset will need to be constructed and treated as a distinct dataset.

Table 36: Secondary Dataset – Interchange Count Table

Field Name	Description
Tap Foreign Key	The tap (secondary dataset) this interchange count refers to
Before Mode Foreign Key	Mode of the segments used before this segment; can be NONE
Before Interchange Count	The number of journey segments for this before combination
After Mode Foreign Key	Mode of the segments used after this segment; can be NONE
After Interchange Count	The number of journey segments for this combination

4.6 Tertiary Datasets

Primary data sets are highly detailed and suitably accurate to support transport analytics. High resolution and high accuracy are required to provide a thorough analysis of customer behaviour and transport operations. Secondary datasets are still high-resolution but use further disassociation to increase the risk mitigation, and those datasets could be used by additional service partners and responsible researchers.

A tertiary dataset is a dataset focussing on a narrow topic and is generated by further processing using aggregation and elimination to ensure they are privacy safe.

Tertiary datasets should only be created from the accurate primary dataset to prevent errors that would result from calculating the tertiary dataset from a tertiary dataset. For example, if every hour 109 persons embarked at a particular stop, a tertiary dataset might record 100 embarkations at that stop. If those counts were then aggregated to a daily count it would sum to 2,400. However, the actual daily total was 2,616 embarkations, therefore 216 embarkations would not be counted because of an artefact caused by using a less accurate tertiary dataset as a source.

4.6.1 Rules of Thumb

As stated previously, the public has a Right-to-Know (GIPA) which cannot be ignored, but must be reasonably balanced against the need to protect individual privacy.

Tertiary datasets use aggregation, disassociation, and elimination (the fourth technique of privacy protection) to mitigate any risks to privacy. Tertiary datasets are already distributed to the general public by several service partners.

The first rule of elimination is *round down the passenger count to a multiple of 10*. This rule ensures that the analyst knows the patronage was always equal to or greater than the stated passenger count, while preventing privacy leakage through analysis detecting individual passenger changes. This is especially important at minor stations, for example, Bell in the Blue Mountains.

The rule of elimination is not absolute because the risk to privacy is proportional to the temporal and spatial resolution of the dataset. A high-resolution minute by minute activity dataset presents a greater risk of privacy infringement through biographic re-association than a low-resolution monthly or yearly dataset. For example, a tertiary database indicating there was 3 passengers in 2017 at Bell does not present a risk to any person's privacy, however datasets recording only 1 person in one hour of one day might risk an individual's privacy.

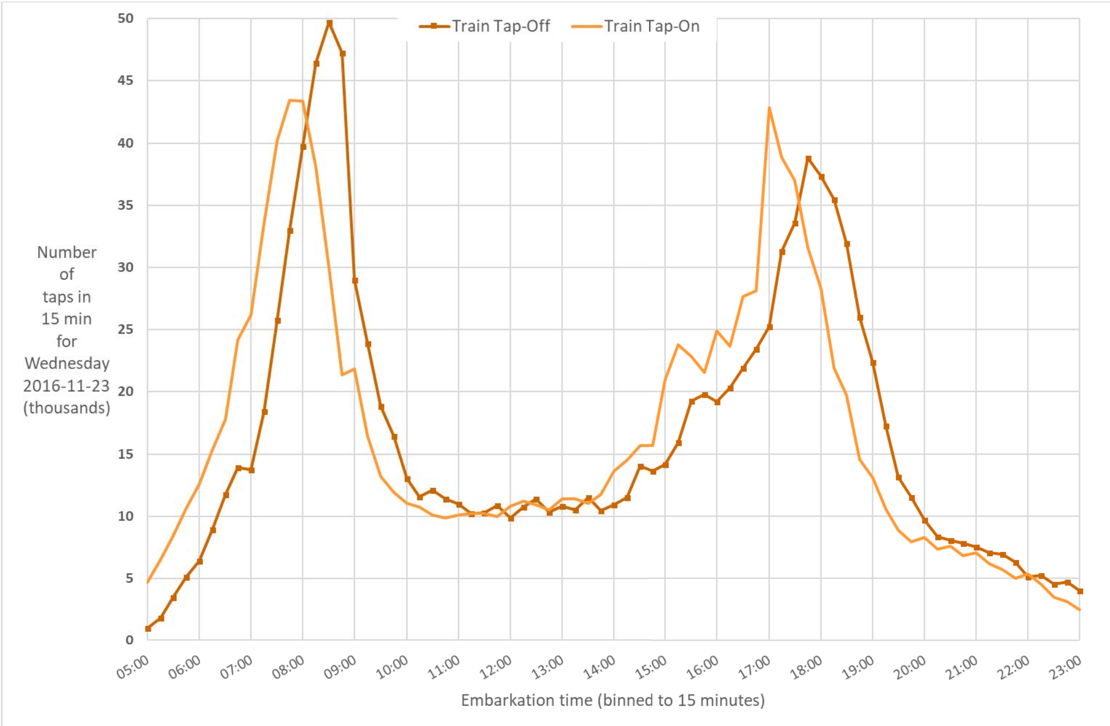
As a second rule of thumb, eliminating any record with less than 20 in the patronage count will mitigate most risk. However, this is only a guidance, not a requirement, as low-resolution datasets (yearly / LGA) cannot risk passenger privacy.

4.6.2 High resolution in only one dimension

A dataset that is high resolution in only one dimension (location, time) is of minimal risk to privacy. Such datasets include yearly embarkation counts for all locations, or minute-by-minute embarkation counts for the entire coverage area. Care needs to be taken to ensure that the low-resolution does not make a detailed record of rare services.

As an example, TfNSW released a synthetic dataset called 'Opal Tap On and Tap Off Release 2' which aggregated embarkations and disembarkations per mode throughout the entire Opal coverage area in 15-minute bins.

Figure 69: Number of taps in 15-minutes for trains on 2016-11-23



Source: • (TfNSW 2017a)

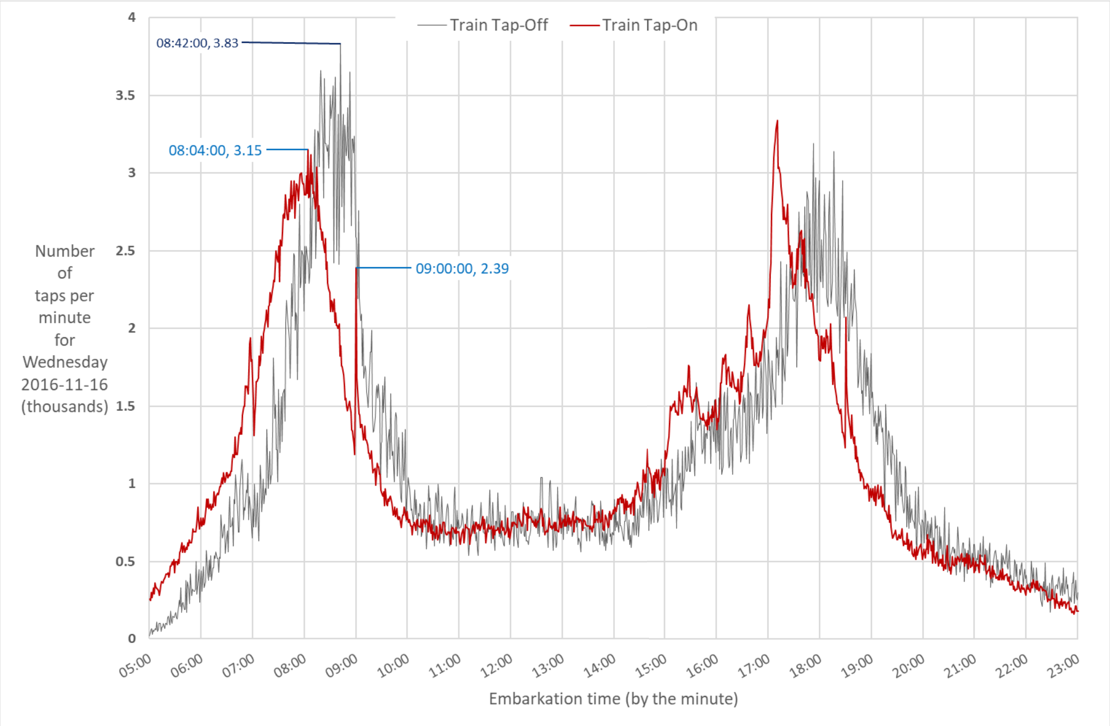
Figure 69 above is the train Tap-Ons and Tap-Offs from that dataset between 05:00 and 23:00. In this figure it is clear that the majority of embarkations are before 08:00, while the majority of disembarkations appear arranged to be before 09:00. Also, the figure shows, that evening peak embarkations surge in the 15 minutes starting at 17:00. The figure shows a burst of passengers starting at 09:00 when the off-peak discount begins.

Since the probability space for an individual tap in the dataset used above is the entire NSW GMA and beyond, there is no risk to personal privacy from releasing the actual data. As such, it would be an acceptable risk to release the actual Tap-On and Tap-Off counts per minute for each mode, whenever there is more than one service in operation. A minute-by-minute dataset aggregated for the entire GMA would allow analysis of overall passenger behaviour but not risk anyone's privacy.

There is a caveat here, if there is only one ferry, bus, or train operating at a particular minute of a particular day, then any disembarkations can be associated back to the timetable to compute the number of disembarkations at a particular wharf. Therefore, it might be worth eliminating data from the low service periods.

Figure 70 shows the number of Tap-On and Tap-Offs for all trains by minute on one Wednesday surveyed. The higher resolution of this plot allows it to more accurately demonstrate passenger behaviour. For example, peak pricing is charged based on Tap-On time starting at 07:00 and ending at 09:00. Figure 70 shows that there is a substantial number of persons who time-shift their travel patterns to access those lower prices.

Figure 70: Number of taps in a minute for trains on Wednesday 2016-11-16



On that Wednesday over 2,000 passengers delayed their journeys until the station clocks displayed 09:00 and the 30 percent discount would apply. This behaviour could also explain the spike at 18:30. People finishing work could explain the bursts around 16:30, 17:00 and 17:30. The increase at 15:00 is explained by the end of the school day.

The advantage of a minute-by-minute dataset to researchers, is that the user of the dataset can further aggregate the data to the temporal resolution needed, be it 20 minutes or 5 minutes. These examples demonstrate that with care and a large enough aggregation, high-value, high-resolution datasets can be created that also protect passenger privacy and be suitable for public release.

4.7 Future Research

Consultation with Queensland, Victoria, Western Australia, and New Zealand could develop a common Australasian dataset design. Once the structure of these datasets is defined, it would be possible for multiple transport analysts to develop code to process all datasets. Afterwards, the agencies and researchers can share processing code to reduce the cost of creating new analysis methods.

It would be possible to distribute a context table for the LOS with respect to hours-of-service for each stop per date. As well as a table for the LOS with respect to vehicle-headway for each hour per stop per date. LOS for on-time running and headway adherence could also be generated by operators keen to empirically measure and improve their customer service operations (Kittelson et al. 2003). Operators should also report actual runtimes (Hounsell 2018a).

Some additional tertiary datasets are described in the case study section as they are used. No thesis could ever exhaustively list all desirable datasets because access to rich datasets would allow researchers to invent new datasets to empirically measure the service delivery and actual passenger response. As such, it is important for the service partners to freely release datasets that are known to be useful, thus allowing researchers to develop new analytical techniques to assist the service partners in the future.

The following research proposals, and countless others, could be undertaken using a dataset that aggregates the number of embarkations for each transport mode per day.

Since such aggregations would cover many services in many locations, the activity dataset produced would contain no biographical information and is thus privacy safe. These datasets would also use elimination techniques for further privacy protection.

Figure 71: Tertiary Daily Count and Interchange Tables

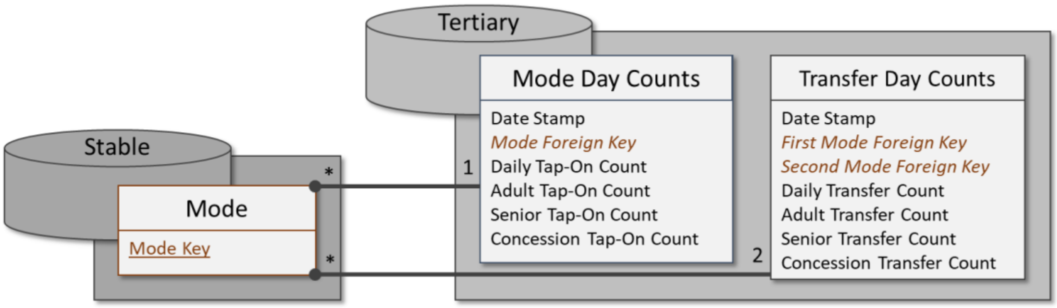


Table 37: Tertiary Daily Count Tables

Field Name	Description
Date stamp	Date of these embarkations
Mode Foreign Key	Reference to the mode of these embarkations
Daily Tap-On Count	The number of embarkations for this mode on this date
Adult Tap-On Count	Embarkations by Adult cards for this mode on this date
Senior Tap-On Count	Embarkations by Senior cards for this mode on this date
Concession Tap-On Count	Embarkations by Concession cards for this mode on this date

Table 38: Tertiary Daily Interchange Count Tables

Field Name	Description
Date stamp	Date of these transfers
First Mode Foreign Key	Reference to the mode transferring from
Second Mode Foreign Key	Reference to the mode transferring to
Daily Transfer Count	The number of transfers
Adult Tap-On Count	The number of transfers by Adult cards
Senior Tap-On Count	The number of transfers by Senior cards
Concession Tap-On Count	The number of transfers by Concession cards

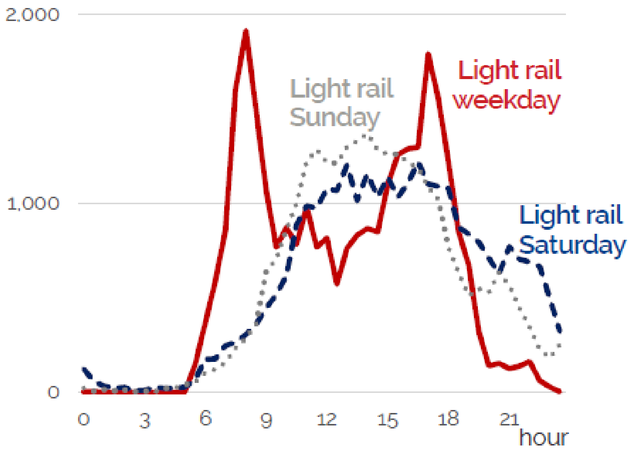
4.7.1 Sydney Ferries, Metro, and Trams are Different

Although weekend patronage on NSW trains and buses is only half of the weekday patronage, this is not the case for Sydney's ferries or light rail. As described previously, the IWLR shows patronage on the weekends at a level consistent with weekday patronage, despite the service operating at only 15-minute vehicle headways.

In contrast, Sydney's ferries are very popular on weekends, especially with unlimited travel available on Sundays for $2.50. Figure 72 and Figure 73 are from the latest pricing regulators report into public transport fares. Unfortunately, the average weekend patronage numbers for Sydney's trains has never been released.

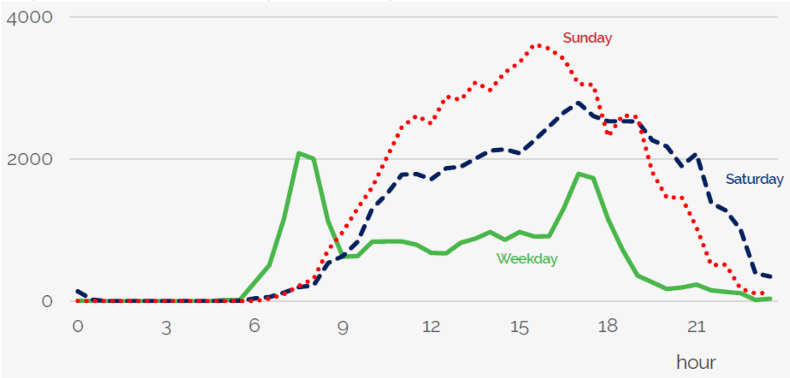
The Tertiary Daily Count and Interchange Tables could be used to research the factors, such as frequency, span, and connectivity, influencing the desirability for transport on weekends.

Figure 72: Light Rail Trips (in a typical day)



Source: (IPART 2019, p. 19) includes Newcastle Light Rail

Figure 73: Ferry Rail Trips (in a typical day)



Source: (IPART 2019, p. 19)

4.7.1.1 Sydney Metro

Passengers will be ready to travel at random times. If there is a significant wait until the next scheduled service, they may choose to do something productive rather than wait at the station. However, from the customers' perspective they are just killing time while they are waiting for the next usable service. As such, *no matter where they wait, the average passenger wait-time experienced will be half of the vehicle-headway while the maximum passenger wait time experienced will be one full vehicle-headway*.

That is partly why humans have been observed to respond more positively to high-frequency public transport (Kittelson et al. 2003, pp. 3-29; Walker 2012). Presented with a high-frequency service, passengers arrive in a random distribution as observed in 5.2 – Turn Up and Go Trains. The research suggests passengers do not check the timetable before preparing to catch more frequent public transport services. The data indicates that Bondi Junction passengers find spending the maximum (10 min) wait-time at the station tolerable, suggesting a wait time of less-than ten minutes is an acceptable cost to most people.

Sydney Metro began operations on 26 May 2019, and every weekend since it has operated as an all-stops service with a 10-minute vehicle headway (LOS B) for most of Saturday and Sunday. Thus, the maximum passenger wait time experienced was 10 minutes and the average was 5 minutes. Table 39 below shows the average Saturday patronage on the Sydney Metro in July was 54% of the average workday patronage, and the Sunday patronage was almost 60% of the average workday patronage.

In contrast, for many stations on the Sydney Trains network, the off-peak vehicle-headway is 15 minutes (LOS C). This higher level of weekend patronage on a new high frequency service, suggests that Sydney residents would respond with mode shift or additional trips by public transport, if the vehicle-headway on Sydney Trains was reduced to 10 minutes (LOS B) for off-peak services. Further research is warranted on this critical aspect of customer travel preferences.

Table 39: Average Sydney Metro embarkations by day in June-July 2019

Month	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
	Average Passengers per Day of the Week from Monthly Totals						
June 2019	58,240	63,930	65,420	70,080	68,700	34,490	39,640
July 2019	64,030	66,410	66,920	66,340	66,140	36,170	39,680
		Average TWT Workday				% of TWT	
June 2019				66,480		51.9%	59.6%
July 2019				66,570		54.3%	59.6%

Source: • TPA Data Request with Matt O'Sullivan at Fairfax

During the Stage III presentation for this thesis the expert examiners — A/Prof Garry Glazebrook and Paul Van Den Bos — hypothesised that the increase in monthly patronage on Sydney's trains and buses is in part a significant increase in the off-peak patronage, to confirm that, further research could be conducted with a daily patronage tertiary dataset.

4.7.2 Other Tertiary Datasets

There are many tertiary datasets that can be generated from secondary datasets. Examples previously listed include counting the number of Taps (On and Off) at each stop for every whole day. The daily counts must be generated from the secondary dataset because that was not treated with elimination and would create accurate counts.

Another highly desirable tertiary dataset would leverage the route tables maintained in the Opal Database to provide information on the importance of particular bus routes. As seen above, using the route information in the secondary dataset could provide significant insight into transit usage throughout the Opal coverage area. This insight can be followed by further research into the specifics of bus patronage on key corridors, as shown by Hounsell (2018b); and Murdoch (2017).

Table 40 below was calculated from daily embarkation counts per bus route — a tertiary dataset mentioned previously. Since some bus routes have just one service per day, care would be needed to maintain privacy in this dataset. However, since rare services function primarily as a social transport system, it may be reasonable to exclude them from the functionally different mass transport routes. Some buses are so important they operate more frequently than the city's railways, such as the 343 through Waterloo.

Table 40: Number of embarkations per day for Top 10 bus route Nov 2016

Rank (Mean)	Route	Minimum	Maximum	Mean	95% CI	Standard Deviation	Samples
1	400	15,050	16,180	15,430	750.0	649.5	3
2	343	13,750	14,330	13,987	351.4	304.4	3
3	389	13,140	13,470	13,307	190.6	165.0	3
4	440	11,860	12,830	12,427	583.4	505.2	3
5	380	11,060	12,600	11,997	949.5	822.3	3
6	M52	11,230	11,370	11,280	90.2	78.1	3
7	333	10,270	11,740	11,057	855.0	740.4	3
8	M20	10,220	10,500	10,340	166.5	144.2	3
9	T80	9,460	10,180	9,720	461.3	399.5	3
10	M30	8,170	8,680	8,457	301.2	260.8	3

Source: (Hounsell 2018b, p. 2, Table 1)

Sydney's bus network uses common stops and master routes to target higher frequencies and interchanges in key bus corridors, such as Parramatta Road. Not all passengers embarking at a particular bus stop will continue travelling to the end of the bus corridor. For example, the 343 goes between Waterloo and St Leonards via the CBD (where most passengers disembark). Operators could benefit from a dataset of monthly counts for stop pairs like Queensland's; see 5.3.1 – Mobility datasets.

4.8 Summary

That Opal datasets contains biographical travel histories which causes a risk of privacy violation, and that is a significant methodological barrier preventing their wider use. The contribution of this thesis to the field of transport analytics, is leveraging the Need-

To-Know framework to develop a method of using privacy protecting transformations to create useful derived activity datasets that no longer contain travel histories.

First create a high accuracy primary activity dataset recording the count of Taps between stop pairs. This dataset must be carefully controlled but allows the creation of additional datasets without requiring any reference to the original dataset with its travel histories. As such, the transport operators can begin the transition to an empirically driven approach using this primary dataset as the basis for analysing the transport services delivered and the customer's response to the services that were delivered.

Thereafter privacy protected secondary datasets, such as tap counts per stop, can be created for distribution and use by a wider range of service partners. Secondary datasets would still be high resolution and completely accurate. Finally, more limited tertiary datasets that are specifically designed to be safe for distribution would be created for the wider community to fulfil the public's Right-to-Know (GIPA).

This method controls risk by creating a powerful primary dataset that eliminates the need to access the original dataset, as shown in Figure 57 above. Once the risk is controlled, further empirical analysis of the activity datasets will not violate a person's privacy by accessing their travel histories.

The next chapter describes several case studies that have used the primary or secondary datasets described in this chapter to allow empirical analysis of the transport system.

Part II

Case Studies

CHAPTER 5 – CASE STUDIES

This thesis has outlined the historical and academic context for this research, including further research into facilitating the transition to an empirically-driven customer-centric approach to transport operations. As described in the Introduction this thesis focuses on questions arising from analysis of the first axis of transition:

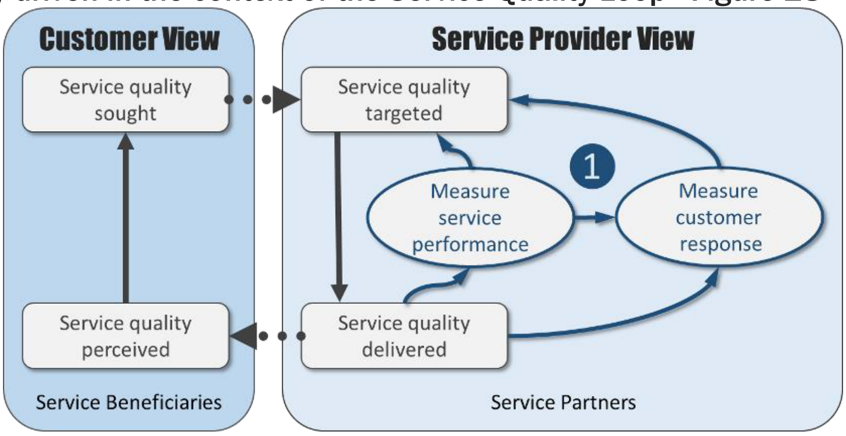
❶ Moving from model driven to empirically-driven:

i) How is the transport network actually being delivered?

ii) How are passengers responding to the delivered transport network?

These questions can be shown to influence operations in the Service Quality Loop as shown in the following figure.

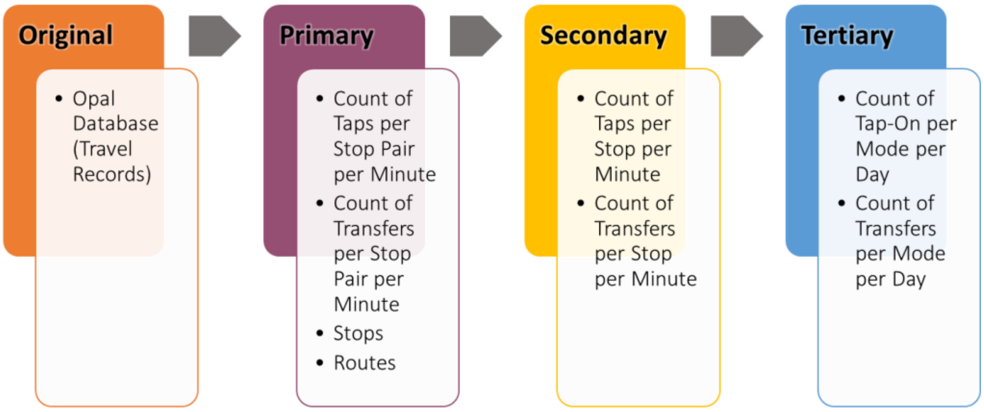
Empirically-driven in the context of the Service Quality Loop - Figure 28



Ensuring privacy protection prevents the uncontrolled access to travel-histories in a TOTOR dataset. Thus, privacy protection concerns limit the use of Opal datasets as a source of measurement for empirically-driven operation and those concerns will always remain an ongoing methodological factor in analytics.

To overcome these obstacles, this method uses three stages of transformation to control the risk of privacy violations. The original dataset is processed using the techniques of deidentification, disassociation, aggregation, and elimination described in Chapter 3 to mitigate privacy risks. The proposed method creates three tiers of datasets for use by different service partners. Since each subsequent transformation increases privacy protection, the tertiary datasets can be made available to the general public.

Figure 74: Framework of dataset transformation



Consequently, operators can use the high accuracy and high-resolution primary and secondary activity datasets as one means to measure the service performance and customer response in the context of the Service Quality Loop, as shown above.

> 'What gets measured gets done!' (Gummesson 2008, p. 319)

Transport customers respond to the services delivered based on their personal values. A vital future step in service delivery is aligning operator with customer values and their service quality sought. However, the first step is for the operator to seek evidence on how customers respond to the services promised and delivered.

The following section outlines several case studies that leverage these datasets to perform analysis of passenger behaviour or transport operations. These case studies are examples that use primary and secondary datasets to measure service delivery or passenger response. The case studies discuss techniques that build on each other towards the development of techniques to determine the human throughput of intersections to assist the empirical optimisation of intersections for customers.

The following case-studies demonstrate the use of the analytics techniques developed during the research programme and apply those to all modes of transport. It has been suggested that the case-studies should refer only to the IWLR. However, the thesis focusing only on trams introduces the risks of practitioners discounting the techniques as tools for analysis of trains, buses, ferries, etc. The risk of mode bias and mode blindness in practitioners can be controlled by covering all modes in the case studies.

5.1 Well above forecast

The single most important reason for favouring empirical analysis over models is because models are based upon an organisation's assumptions and interpretations of observations and so are easily distorted by incorrect assumptions or Misknowns. Distorted models can create situations where long-term planning is based upon fallacies.

5.1.1 Long Term Transport Master Plan

As shown in Figure 4 above, the modelling significantly underestimated the patronage on the IWLR, (GHD 2010; Hounsell 2018c). In just eight years after its release, the figures and assumptions central to the NSW LTTMP have been disproved (TfNSW 2012). Here is the rundown of what TfNSW modelling was expecting in 2031.

Figure 75: Projected patronage from 2012 NSW LTTMP

Our transport network is coming under greater pressure	
While around 11 percent of all trips taken in Sydney each weekday are by public transport, 23.9 percent of Sydneysiders use public transport to get to work.	Based on current trends, the proportion of commuters using public transport to get to work in Sydney is expected to increase by 3% by 2031.
Commuter trips currently account for around 16 percent of all weekday trips.	Based on current trends, there would be a 34 percent increase in commuter trips by 2031.
CityRail carried 303.5 million passengers in 2011-12, providing 2,781 daily services on weekdays and 1,943 daily services on weekends.	The number of trips made by rail is expected to increase by 26 percent by 2031, based on current trends.
Approximately 850 bus routes operated in the GMA, servicing 223.4 million passengers in 2010-11. 529 bus routes operated in rural and regional areas, carrying 5.7 million passengers in 2010-11.	By 2031, based on current trends, the number of weekday trips made by bus is expected to increase by 29 percent.
Around 68 percent of all weekday trips by Sydneysiders are made by car.	Under a 'do nothing' scenario, most travel would continue to be by motor vehicle, with roughly the same percentage of trips still made by car in 2031.

Source: (TfNSW 2012, p. 25)

Figure 76 below shows the historical data used in the LTTMP recorded by BITRE with the forecast patronage shown for 2031. Overlaid is the patronage recorded from the Opal system. NSW exceeded the patronage forecast for 2031 in just five years. The LTTMP forecast for buses show a different accuracy for the model's forecasts – for buses the forecasts were exceeded in four years, as shown in Figure 77 below.

Figure 76: Number of NSW railways journeys in comparison to the NSW LTTMP

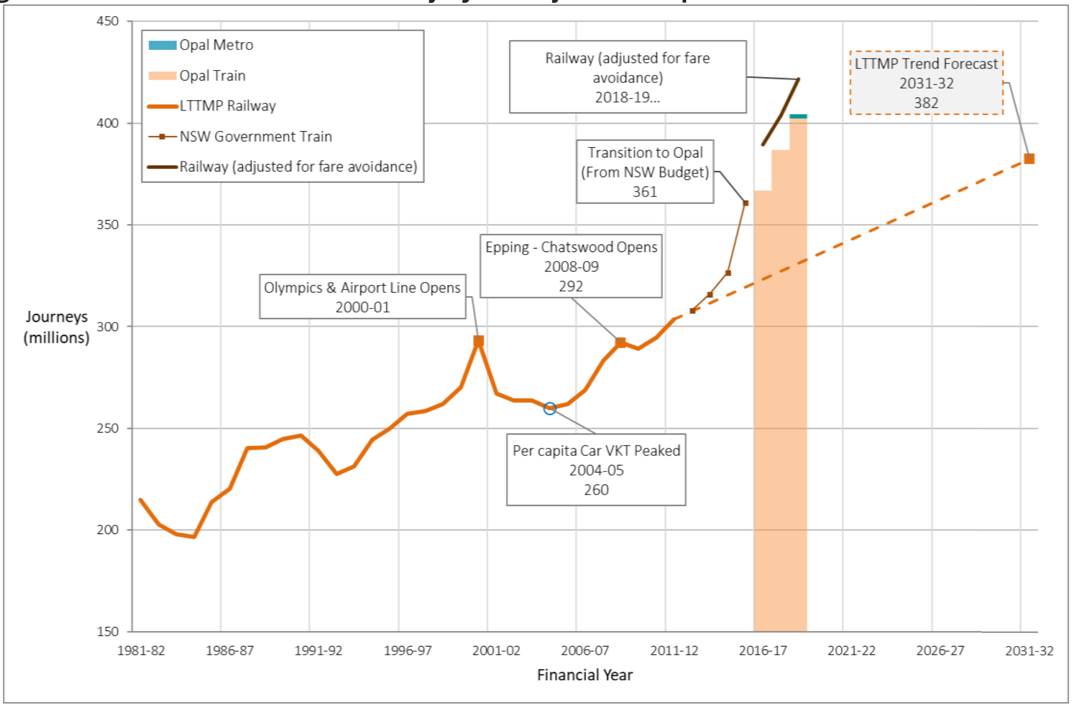
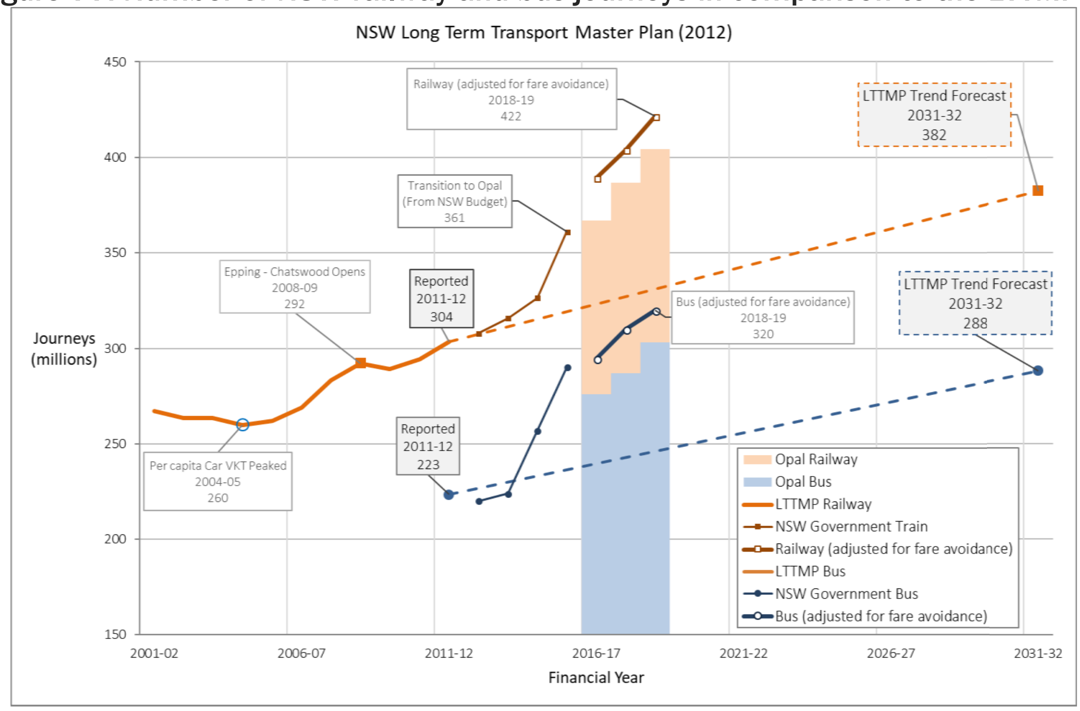


Figure 77: Number of NSW railway and bus journeys in comparison to the LTTMP



Sources for both figures: (TfNSW 2019b), (BITRE 2016, p. Table T 5.a, (BITRE 2018, Table T 5.5a), (TfNSW 2014b), (NSW-Treasury 2017), (TfNSW 2012, p. 25)

5.1.2 Westconnex

One of the most likely reasons for incorrect forecast is the assumption that 'most travel in Sydney would continue to be by motor vehicle, with *roughly the same percentage of trips still made by car* in 2031' (emphasis added) (TfNSW 2012).

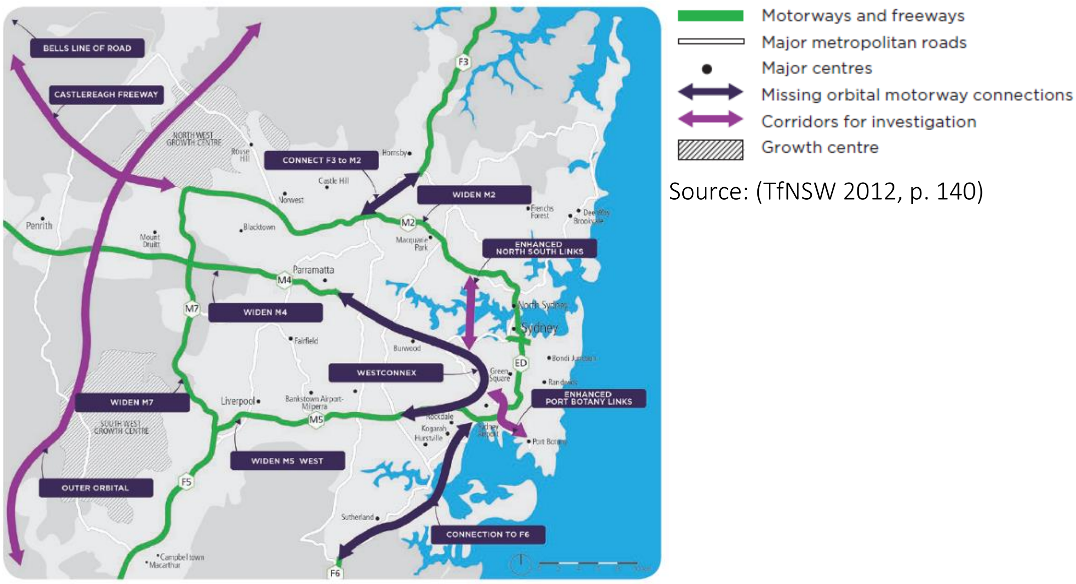
NSW modelling has been relying on the assumption that increasing congestion and travel time will not impact the mode choice of individuals and thus the mode share of private cars.

> 'Sydney's population is forecast to grow by 1.3 million over the next 20 years from 4.3 million to 5.6 million people' (SMPO 2013, p. 13).

These figures show RMS and TfNSW expected a population growth of 30%. They projected the number of road trips for that period would grow from 12.1 million to 15.8 million representing a 30% growth rate (SMPO 2013, p. 13). Throughout the LTTMP, Westconnex Business Cases, and Westconnex Environmental Impact Statement (EIS) it was assumed that road trips would grow exactly in proportion to population (NSW-RMS, INSW & TfNSW 2012; RMS-WDA 2014; SMPO 2013; TfNSW 2012).

For persons from outside of Sydney, Figure 78 below shows the scope of the proposed motorways that are planned through the middle of the city, and the surrounding farmland and national parks. The following figure is included for convenience and does not show all of the proposed motorways. Note that in the following figure the darker grey background represents either farmland or national park.

Figure 78: The Future of Sydney's Motorway Network



Source: (TfNSW 2012, p. 140)

The assumption that increased travel times would not result in any mode shift for any journeys leads to some interesting outcomes. Page 111, 113 and 135 of (RMS-WDA, Jacobs & AECOM 2014) shows the RMS and TfNSW assumes that the Sydney Road Network Average Speed will have dropped from 31 km/h in the AM Peak of 2012 to

20km/h in 2031 due to the Car VKT rising by 35% and the Commercial VKT rising by 30-45% leading to unrealistic travel time forecasts.

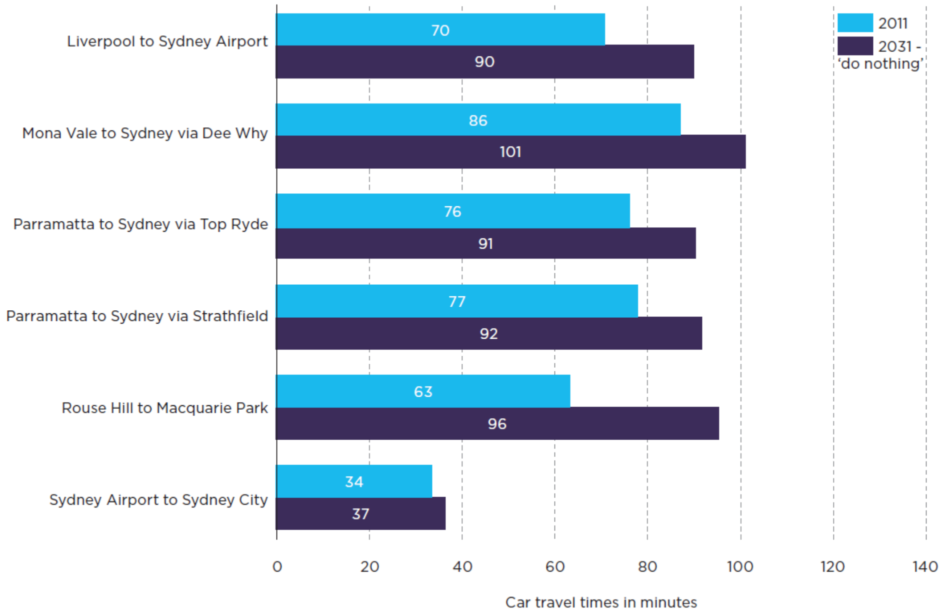
Table 41: How much travel time would WestConnex save road users?

	2021 Without Westconnex	2021 With Westconnex	Trains to Wynyard
Penrith CBD via City West	114 min	92 min	59-62 min
Penrith to Parramatta	~50 min	~48 min	31 min
Parramatta CBD via City West	64 min	44 min	31 min

Adapted from: (Adapted from NSW-RMS 2012)

Table 41 outlines the expected travel times by the RMS for journeys with and without the Westconnex motorway. The 114-minute travel time from Penrith, is symptomatic of the assumption of stable mode share. The majority of the motorway network is modelled with a volume capacity ratio over 120% — the M4 peaks at a V/C ratio of 133%, the M2 at 150% and the M5 at 174%. When these unrealistic numbers are fed into the model, they give rise to significant Travel Time Savings (TTS) that are overinflated. Those inflated TTS are used to justify significant benefits in the business case, like Greenhouse Gas (GHG) reductions.

Figure 79: Peak travel times along strategic corridors for cars (in minutes), 2011 and 2031 'do nothing' scenario



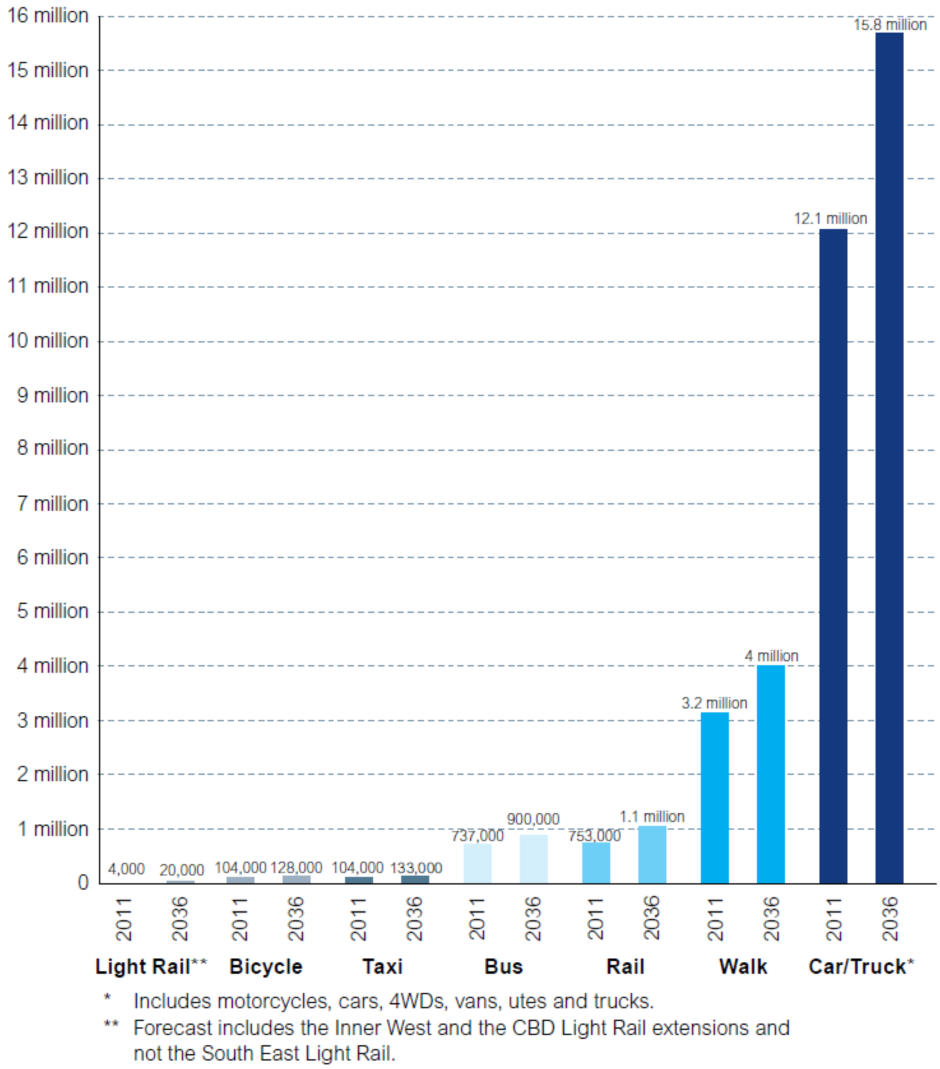
Source: (TfNSW 2012, p. 83, Figure 4.7)

5.1.3 Modelling the transit

The Westconnex modelling also covered the expected patronage on the region's rail services enabling assessment of the model's ability to forecast future scenarios.

The models used in the LTTMP and the Westconnex EIS projected rail patronage of 0.75 million in 2011 rising 46% to 1.1 million on an average workday in 2036. In reality the November 2016 average workday railway patronage was 1.2 million. The light rail patronage in 2016 was 0.029 million per day, exceeding the projected 0.020 million passengers by 46%. Similarly, buses in November 2016 were over 0.94 million journeys on an average workday exceeding the 2036 forecast of 0.90 million.

Figure 80: Number of Sydney trips by main mode for an average workday



Source: (SMPO 2013, p. 13)

The Mogridge Conjecture[28] states that the travel speed on the variable speed network (road) will achieve an equilibrium proportionate to the speed on the fixed speed network (segregated rail) as persons shift from the slower mode to the faster mode to reduce

[28] Investigations using Google maps indicate parallel roads to the rail lines had a travel speed of 80% of the rail line; however, dedicated research should be undertaken to assess the impact of the Mogridge conjecture on Sydney.

travel time (SACTRA 1994; WSP 2018; Zeibots 2007; Zeibots & Petocz 2005). The National Guidelines for Transport Systems Management strongly encourage modelling that considers travel time as a determining factor in mode choice (ATC 2006).

As such, in a comprehensive model used to plan the future of Sydney's transport and to assess the viability of an $18 billion motorway (520,833 times yearly minimum wage), it is reasonable to expect that modelled mode share would consider travel time.

The Vehicle Kilometres Travelled (VKT) per capita data suggests that the current passenger response to the network is significantly different when compared to 2006, (BITRE 2019, Table T4.2 and T1.5c). The LTTMP and Westconnex were based on a census with unusually high levels of private car use that have since declined. The census contained a distorted measurement of the response to the transport network. Therefore, the models were distorted and thus their predictions were inaccurate. Change in behaviour and incorrect measurements can easily lead models to incorrect conclusions, which is an important reason for preferring an empirically driven approach over a model driven approach.

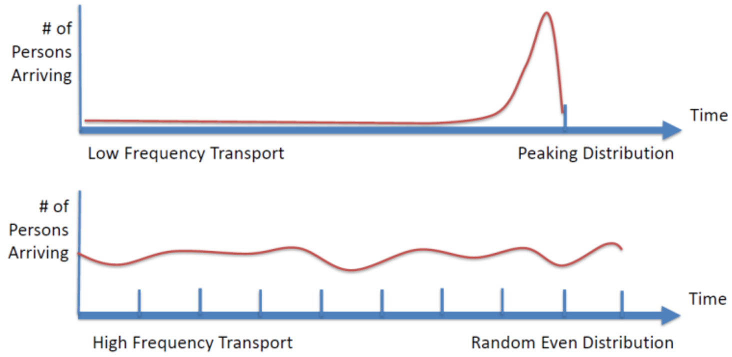
5.2 Turn Up and Go Trains

Service partners can use the minute-by-minute resolution of the secondary dataset to analyse passenger behaviour like station entry patterns. People dislike waiting and will minimise the amount of time waiting on a platform for a service (ATC 2006).

> 'Wait time: reflects the valuation of waiting time, both for the first public transport vehicle of the trip and for all other public transport vehicles used in the journey plus a 5-minute penalty for each boarding.' Sydney Strategic Travel Model – Stage 2 Model Estimation – Technical Report (TfNSW 2001a, 2001b)

People will consult a timetable for low-frequency services and arrange to arrive at the stop shortly before the vehicle. In contrast, for high-frequency services such as the Paris Metro, it has been empirically observed that people arrive at random, creating a steady stream of arrivals (Walker 2012, Chapter 7 - Frequency is Freedom). The evidence used for the "Transit Capacity and Quality of Service Manual" indicates that when the vehicle-headway is 10 minutes or greater, passengers will consult a timetable (Kittelson et al. 2003, pp. 3-29). These arrival patterns are demonstrated in Figure 81, which was created for the UTS subject 48370 "Road and Transport Engineering".

Figure 81: Passenger distribution on low- and high- frequency transit



Source: Hounsell for UTS 48370

TfNSW refers to services with a 15-minute vehicle-headway as ‘Turn Up and Go’ (TfNSW 2013b) The difference in expectation on timetable usage between TfNSW and other operators may indicate a different behavioural pattern in NSW, inaccuracies in the previous research, or that it is not an accurate description of passenger behaviour. As such, further research is suggested using the secondary dataset.

The secondary dataset can be analysed to assess arrival behaviour. For example, Figure 82 shows Tap-Ons at Lithgow terminal on a Wednesday in November 2016 between 09:00 and 09:25 as well between 11:00 and 11:25. There are almost no taps between 09:25 and 11:00 because the train from Lithgow to Central has a two-hour vehicle-headway showing people timed their arrival to match the timetabled departure time.[29]

In contrast, Figure 83 shows the number of embarkations at Bondi Junction terminal during the same period on the same day, where services had a maximum vehicle-headway of 10-minutes. Note that Figure 83 explicitly excludes all intermodal transfers as interchanges from buses would distort the analysis because those passengers do not control their arrival time. Figure 83 shows that passengers arrived in what appears to be a random distribution. In the three hours sampled, only twice did less than three persons arrive per minute at Bondi Junction terminal (remember, this is excluding transfers). These figures show that the secondary dataset provides a significant level of detail to allow further analysis of passenger behaviour while also protecting individual privacy.

[29] An Opal Tap-On can be reversed, without record, if the person Taps-Off at the same station in under 15 minutes.

Figure 82: Embarkations from Lithgow terminal per minute on Wed. 2016-11-16

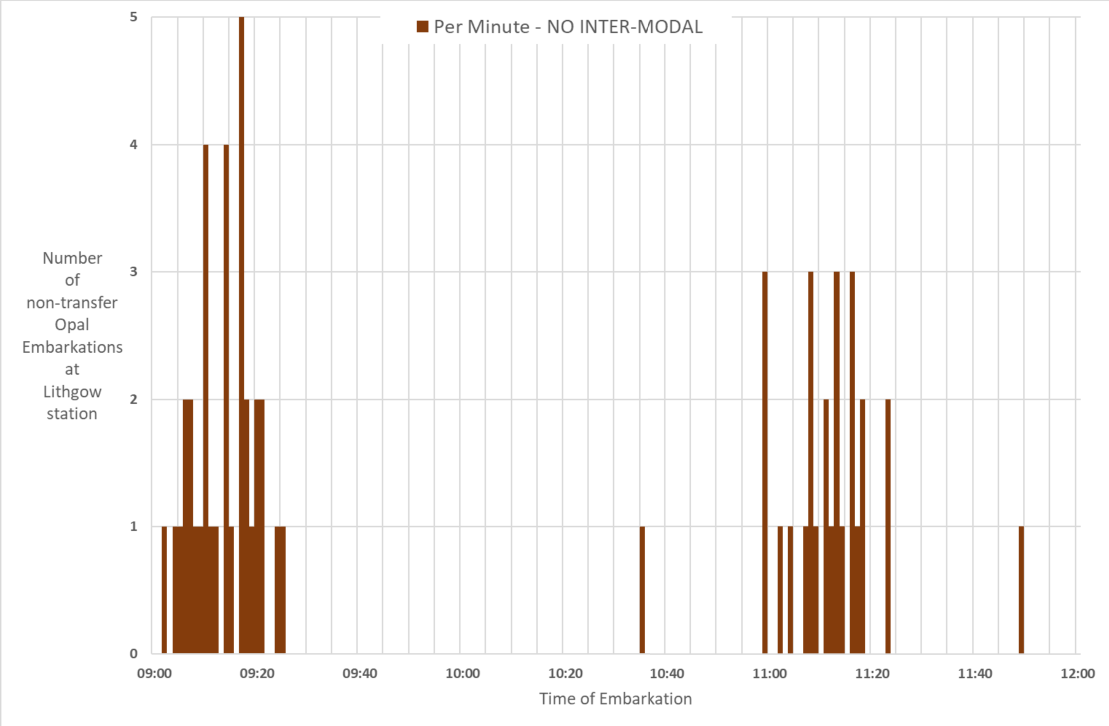
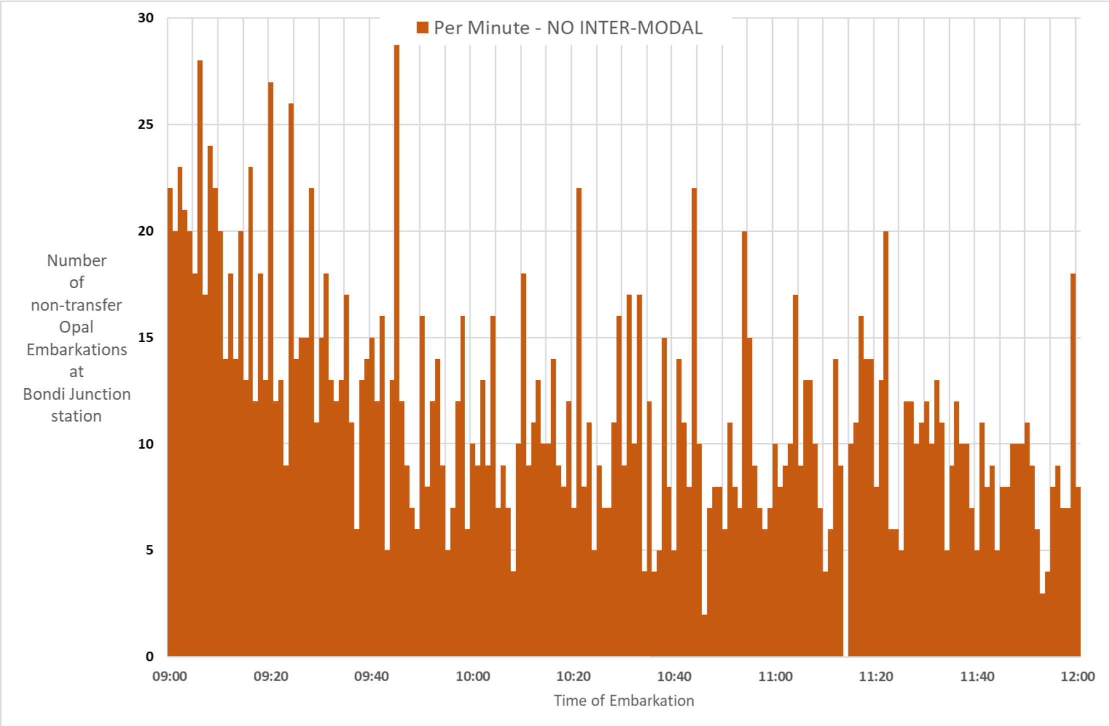


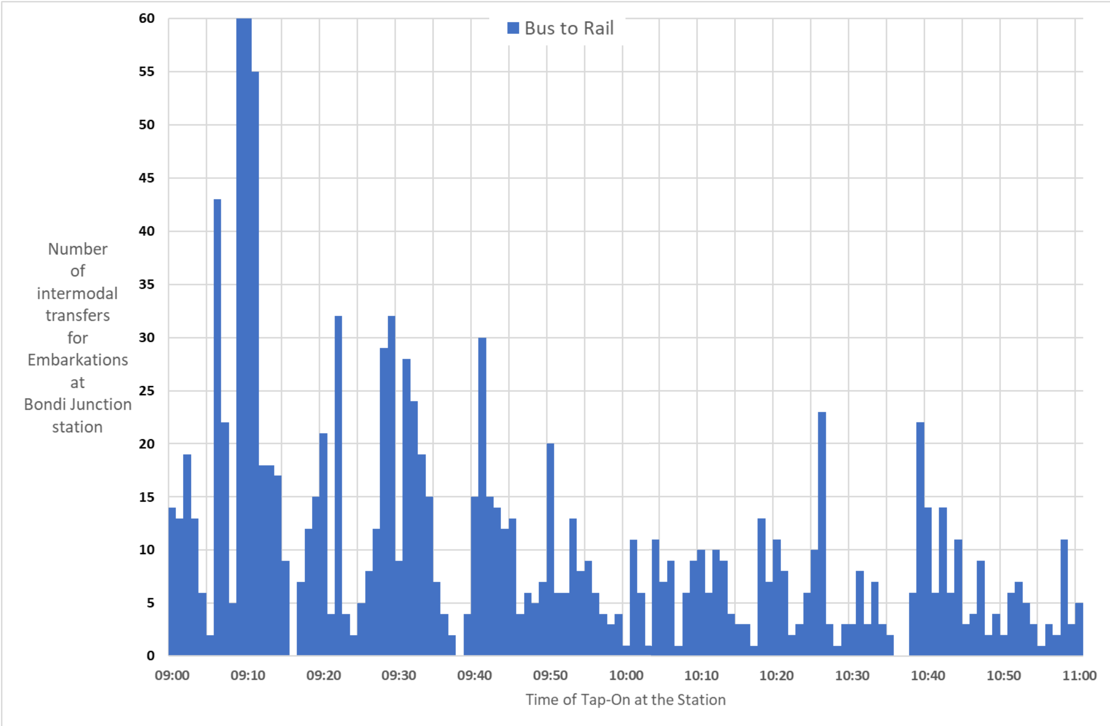
Figure 83: Embarkations at Bondi Junction terminal per minute on 2016-11-16



5.2.1 Happy to interchange

From the secondary interchange table, it is possible to demonstrate the passenger transfers in response to the transport network delivered as shown in Figure 84 below. Figure 84 shows that even between the peaks there was a steady stream of people transferring from buses to trains at Bondi Junction terminal. This arrival distribution shows surges that are consistent with a significant number of passengers transferring from each bus. Further research is suggested to understand the interchange behaviours of passengers between the Bondi Rd buses and other services.

Figure 84: Number of Intermodal transfers from bus to rail at Bondi Junction Terminal per minute on Wednesday 2016-11-16

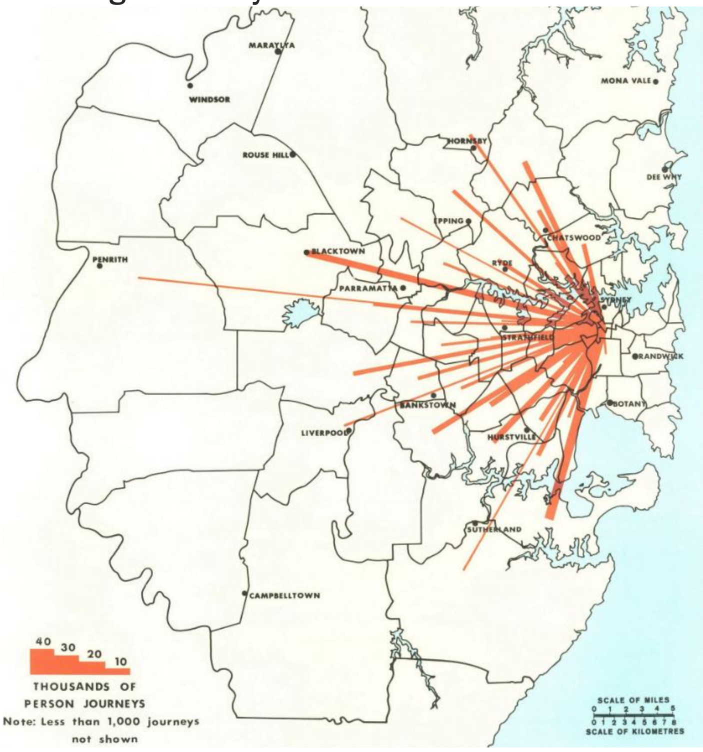


5.3 Passengers travel here, there, and everywhere

The first paradigm shift an operator undergoes is a transition from a technology centric approach to a mobility centric approach. This shift acknowledges that users approach transport as a derived need to satisfy their desire to move between locations. The mobility paradigm is a staple of transport planning and modelling. This paradigm came under significant criticism for the assumption that travel was tied to locations rather than human desires, such as participating in employment (SACTRA 1994).

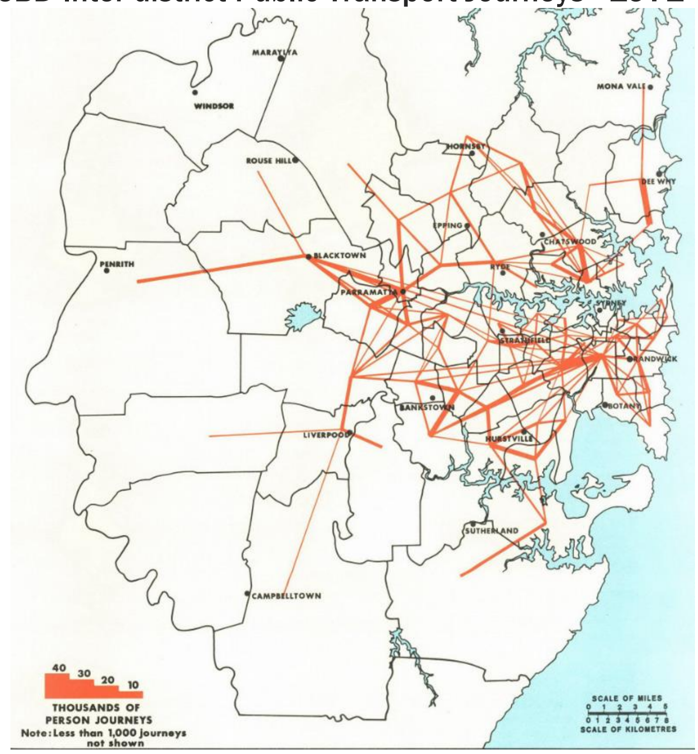
As a historical example of the mobility paradigm, Figure 85 and Figure 86 below are excerpts from the Sydney Area Transport Study (SATS) completed in 1974.

Figure 85: Train Passenger Journeys between CBD and Internal Districts - 1971



Source: (NSW 1974a, p. 181, Figure 6.19)

Figure 86: Non-CBD Inter-district Public Transport Journeys - 1971



Source: (NSW 1974a, p. 181, Figure 6.24)

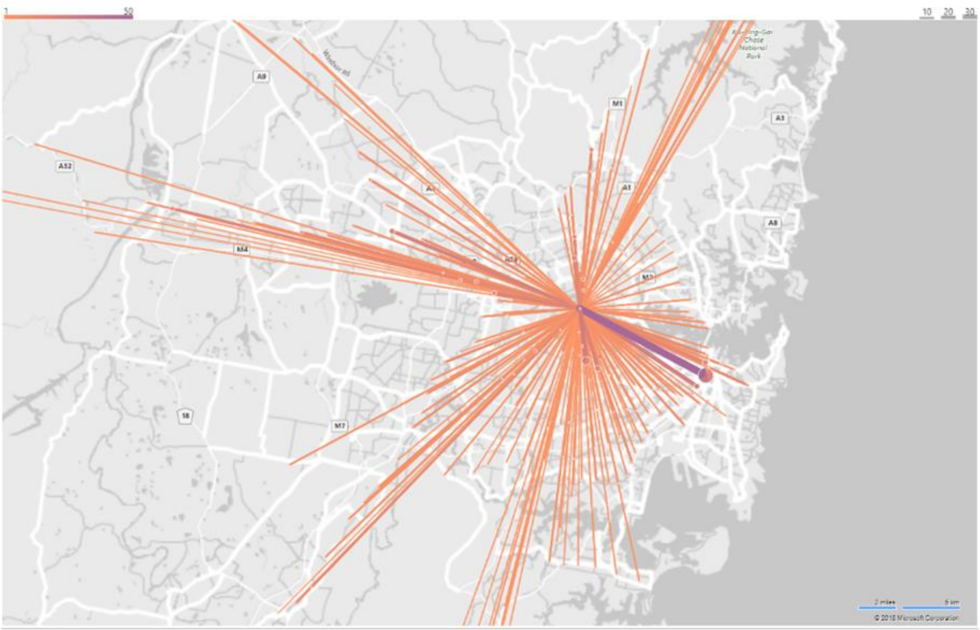
This study undertook a significant manual survey of the travel behaviour, land-use plans, and network operations at the start of the 1970s. These figures show total travel between key regions and the CBD, and then between non-CBD regions.

Since SATS required manual surveys, it lacked the resolution possible from a TOTOR dataset. That is why the results of the older analysis stand in contrast to Figure 87, which was constructed from a TOTOR dataset. This figure shows the number of passengers travelling to Rhodes station and disembarking in the hour after 8 AM.

The richer dataset allows even this simple visualisation to reveal useful patterns. For example, Rhodes serves as a peak-hour destination for people travelling from almost every station within the Sydney Trains network and even from some stations in the inter-city network. The simple visualisation shows that Rhodes is an attractive activity location for employment where businesses can access the entire city's labour market.

These figures show that the high-resolution primary activity dataset with its travel pair counts, can provide a description of actual movements in response to the current transport network delivered. The analysis of this passenger response can then be used to vet existing assumptions, and to identify and eliminate Misknowns that may have crept into the operators institutional understanding.

Figure 87:Embarkation station for 08:00-09:00 disembarkations at Rhodes



Source: (Hounsell 2017b)

5.3.1 Mobility datasets

Each city has a single transport network that emerges from the combination of all services delivered to customers. In Australia, all jurisdictions publish their metropolitan public transport timetables as machine-readable GTFS feeds. These timetables define the public transport service that is targeted by the transport partners (Curtis 2016). It is not possible to know actual transport demand, because that is dependent on the fluctuating manifestation of the customer's needs. In addition, the demand will alter in response to transport availability, cost and congestion (temporal cost) (ATC 2006). However, it is possible to measure the response to the network emerging from the services targeted and the services delivered. Previously, the measurement of response was undertaken through sampling surveys, including the Australian Census and the Household Travel Survey (BTS 2013; TfNSW 2001a).

In jurisdictions with a monopoly TOTOR ETS, operators are able to take a census of travel patterns because every journey segment's embarkation-stop and disembarkation-stop are timestamped and recorded. With a TOTOR ETS, unlike a TOR ETS, there is no need to estimate the disembarkation stop. A TOTOR census documents the actual response to the transport network and allows for new and innovative transport analytics approaches (Legara et al. 2014; Sun et al. 2012; Zhong et al. 2015).

> 'As demonstrated by researchers like (Munizaga & Palma 2012) a significant amount of the literature on processing smart card based electronic ticketing records either focuses on estimating a passengers disembarkation location or contains an outline of the method used by the research for estimation. The passengers travel duration can be estimated using methods such augmenting the TOR dataset with the mass transport timetable and a trip planner. The estimated disembarkation location and travel duration may be within an acceptable margin of error when assessed against other survey methods, but they are still fundamentally inaccurate' (Hounsell 2017c).

As an example of empirically measuring passenger responses and mobility, Queensland releases embarkation-disembarkation pairs for each month. Datasets for February 2016 until April 2019 have been released. The dataset covers the key periods of weekday morning-peak, inter-peak, afternoon-peak, and evening as well as weekends. The fields in the dataset are listed below in Table 42.

SEQ has a large number of shared busways as well as a large number of bus routes that use shared stopping patterns to create master routes. As such, some stop pairs may be recorded under several different bus routes. Caveats such as those are why transport analytics requires domain knowledge — that is a knowledge of the specific configuration of the transport system and its elements and how these affect customer travel behaviour — and not just data-science tools.

Table 42: Fields in the Queensland O-D Dataset

Field	Values	Description
operator	Queensland Rail, Gold Coast Light Rail, etc	The transport service operator
month		The month of the aggregation
route	Rail, GCLR, or Route Numbers	Bus route number, or service type
direction	Busway, GCLR, Rail Clockwise, Counterclockwise East, West, North, South Inbound, Outbound	The direction of this route number
time	Weekday [12:00-8:30), [08:30-15:00), [15:00-19:00), [19:00-24:00) Weekend	The period of the day that was aggregated
ticket_type	go card, Paper	The type of tickets used for this pair
origin_stop		The embarkation-stop
destination_stop		The disembarkation-stop
quantity		The number of journey segments

Queensland may be at risk of breaching their privacy budget due to an apparent lack of an elimination pass. The dataset had 370,000 records referring to stop-pairs in the month with less than 10 passengers.

Table 43 and Table 44 below were created by aggregating the original embarkation-disembarkation stop-pair dataset to disassociate from the bus route, direction, and ticket type dimensions and to reduce the time dimension to weekday or weekend. The GCLR is highlighted in yellow.

SEQ has a distinct weekday and weekend travel pattern with substantially lower passenger volumes on the weekend. The average weekend day is estimated to have approximately 40.5% of the average workday patronage.

During weekdays, the GCLR accounts for only 6.1% of the SEQ patronage, but on the weekend, it accounts for 11.9% and has eight of the Top-10 stop pairs. This contrast is derived from an exceptionally rich embarkation-disembarkation pair activity dataset that accurately describes passenger responses.

Perhaps one of the reasons the GCLR maintains its popularity on weekends is that from early morning to mid evening it provides an all-stops service with a vehicle-headway of 10 minutes (LOS B). From 5am to midnight, it delivers a vehicle-headway of 15 minutes or less (LOS C or better). Queensland Rail's (QR) service frequencies at most stations are usually lower than the GCLR during off-peak periods.

Table 43: Top-10 Public Transport Origin-Destination Trips for Queensland's TransLink November 2018 on Workdays

operator	origin	destination	weekday	origin name	destination name
QR	600014	600018	25,334	Fortitude Valley station	Central station
QR	600389	600018	24,851	Ferny Grove station	Central station
QR	600018	600389	22,838	Central station	Ferny Grove station
GCLR	600801	600811	21,678	Broadbeach South station	Cavill Avenue station
GCLR	600811	600801	21,578	Cavill Avenue station	Broadbeach South station
QR	600407	600018	21,030	Northgate station	Central station
QR	600018	600014	20,678	Central station	Fortitude Valley station
QR	600018	600407	19,916	Central station	Northgate station
QR	600395	600018	18,907	Eagle Junction station	Central station
QR	600018	600395	18,408	Central station	Eagle Junction station

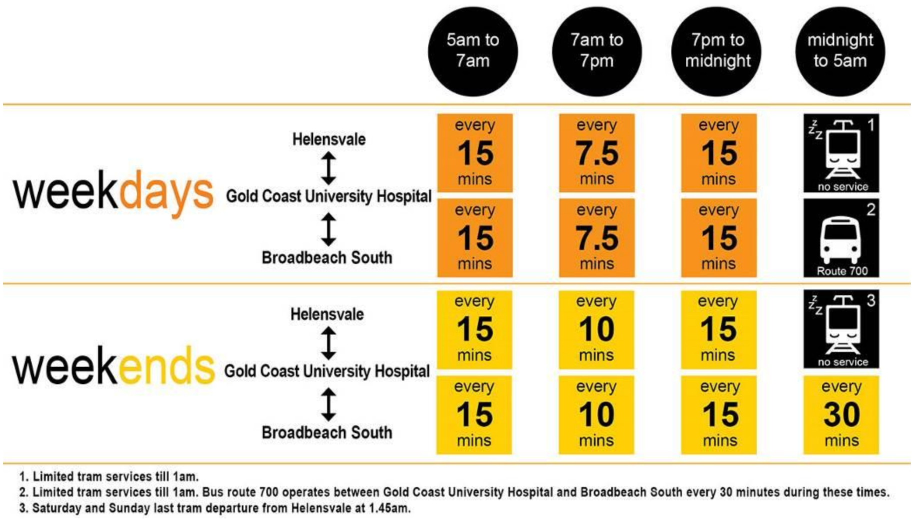
Source: (QLD-TMR 2019)

Table 44: Top-10 Public Transport Origin-Destination Trips for Queensland's TransLink November 2018 on Weekends

operator	origin	destination	weekend	origin name	destination name
GCLR	600811	600801	8,044	Cavill Avenue station	Broadbeach South station
GCLR	600801	600811	7,716	Broadbeach South station	Cavill Avenue station
GCLR	600810	-1	6,569	Cavill Avenue station	*No Tap Off*
GCLR	600811	600836	4,691	Cavill Avenue station	Helensvale station
GCLR	600836	600811	4,246	Helensvale station	Cavill Avenue station
GCLR	600800	-1	3,981	Broadbeach South station	*No Tap Off*
QR	600014	600018	3,514	Fortitude Valley station	Central station
GCLR	600802	-1	3,362	Broadbeach North station	*No Tap Off*
Brisbane Transport	19052	919	3,361	South Bank busway station	Queen Street station
GCLR	600811	600823	3,208	Cavill Avenue station	Southport station

Source: • (QLD-TMR 2019)

Figure 88: G:Link - Riding the G - Frequency



Source: GCLR Operator - https://ridetheg.com.au/%EF%BF%BCriding-the-g/time-table/

The above example from Queensland shows how a monthly aggregation of embarkation-disembarkation pairs could provide a high value activity dataset for transport analysts within the service partners and general community.[30]

5.3.2 Passengers travel not just in the peak

The peak period receives the greatest focus from service partners because the high demand for travel reaches the capacity limits on the network.[31] The high volume-to-capacity ratios cause crowding, which causes operational issues and potentially customer dissatisfaction. That is why the NSW train load surveys are all based on measurements during the peak period, see 1.3.2 – Urgent Need.

Surveys and manual measurements are expensive because they require significant labour to gather the dataset, and then specialised labour to interpret the samples. Due to cost, service partners often choose to forgo analysis of the off-peak and inter-peak periods when those services are not reaching vehicle capacity limits.

However, the TOTOR ETS can generate a primary dataset that also covers off-peak and inter-peak periods. Primary datasets enable service partners to analyse operations and passenger behaviour during all periods of the day, not just the peak.

As an example, Figure 89 below shows maps of travel patterns for customers embarking or disembarking on the Bankstown Line in Sydney (Bankstown –Marrickville). This segment of the Sydney Trains network will soon be closed for conversion to extend the Sydney Metro from Rouse Hill through Sydenham to Bankstown.

> ‘The ‘do nothing’ alternative would involve maintaining existing operations along the T3 Bankstown Line and not completing the Sydenham to Bankstown component of Sydney Metro City & Southwest. The ‘do nothing’ alternative would involve the T3 Bankstown Line continuing to operate as part of the Sydney Trains network, and Sydney Metro operating between Rouse Hill and Sydenham, rather than to Bankstown. … Implementing the ‘do nothing’ alternative would have the following issues:’ … ‘• there would be approximately 27,000 fewer trips on Sydney Metro in the one-hour AM peak, which would impact the effectiveness and viability of Sydney Metro between Sydenham and Rouse Hill’ (TfNSW 2017c).

[30] However, I would suggest additional steps be taken to mitigate the risk of privacy violation where a similar dataset be created in other jurisdictions. One approach to achieving privacy protection without resorting to elimination is through disassociation. A stop-pair dataset could convey the results as rankings and thus disassociate from the actual patronage counts. Then analysis of passenger behaviour would be able to use this dataset to show relative importance in combination with a dataset on actual counts for the major pairs.

[31] Capacity of the network is sometimes limited by the fiscal constraints of the state

The metro operator has stated that their choice of technology requires a complete rebuild of every station on the line. During this conversion, train services will not operate on the line. In the past, manual surveys would be undertaken to assess peak hour patterns and patronage, then from the results a bus-substitution strategy for the major locations would be devised. However, given the rich datasets available from the Opal census, it is now possible to undertake a deeper analysis of passenger behaviour.

Since the primary dataset is a census of activity patterns, it is possible to generate the first plot in Figure 89 that shows all stop pairs connecting to the Bankstown line over the entire day. In this case, passengers using this line connect to almost every other station within the Sydney Trains network, with many journeys requiring transfers.

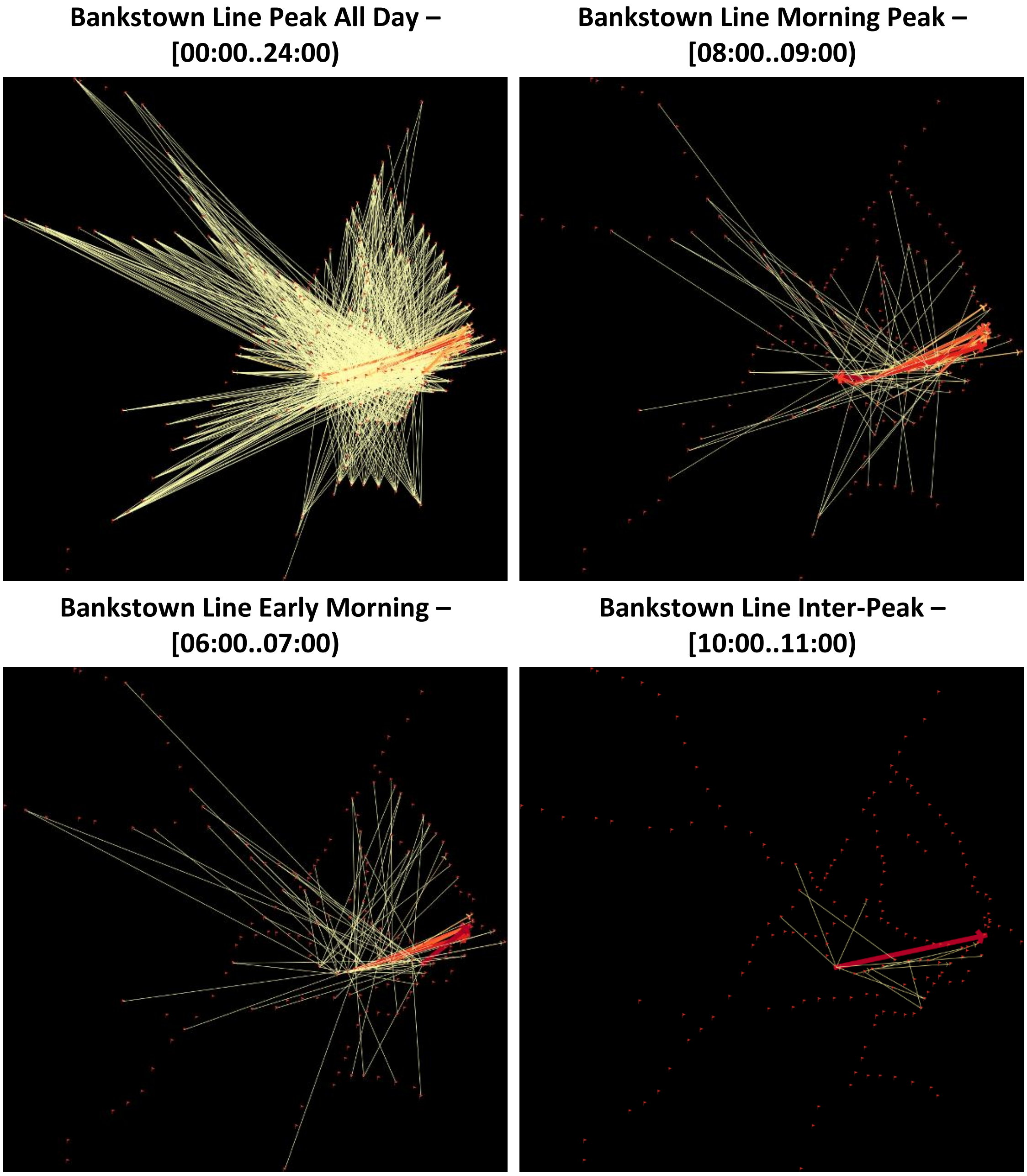
The second plot is an indicative map of the travel pairs used by passengers during the morning peak hour for embarkations and disembarkations on the Bankstown Line. This map shows that there is a wide range of travel behaviours, with the Sydney and Bankstown CBDs being major destinations. However, passengers are currently using the line to connect to stations all over the GMA.

The third plot shows passengers travelling in at 6 AM connect to widely dispersed stations, but they connect to different stations compared to the 8 AM passengers.

The fourth plot shows that patronage and behaviour changes substantially in the inter-peak. At 10 AM the majority are heading to the CBD with just a few heading off to other destinations such as Parramatta or Olympic Park. In the inter-peak passenger wait times increase substantially (vehicle-headway $\geq$ 15 min – LOS C), in comparison road congestion decreases and road speeds increase.

Using the census of behaviour in the primary activity dataset, it is clear that special care in the bus-substitution design must be taken to ensure there is easy connection to centres to the west, north, and south of the Bankstown line and not just to the Sydney CBD.

Figure 89: Stop Pairs on the Bankstown Line on a Wednesday in 2017



Source: • (Hounsell 2017a)

Mobility is a central measurement of the macro responses of passengers to the transport network promised and delivered. The high-resolution primary activity dataset would facilitate a number of measurements of mobility. Tertiary datasets that aggregate mobility to spatial areas or across longer time frames should also be created. Some Australian jurisdictions are already releasing them to the public. Datasets from SEQ indicate that passengers respond favourably to services with higher frequencies. Datasets from NSW demonstrate passenger desire to use the public transport network to access all areas of the city even if interchanging is required.

These secondary and tertiary mobility datasets could help transport service partners reduce their Blind-Spots through empirical analysis — especially of weekend, off-peak and travel behaviour to minor locations. Therefore, it can be concluded that mobility datasets could be used to improve services to customers by measuring the passenger's empirical response to the network promised and the services delivered.

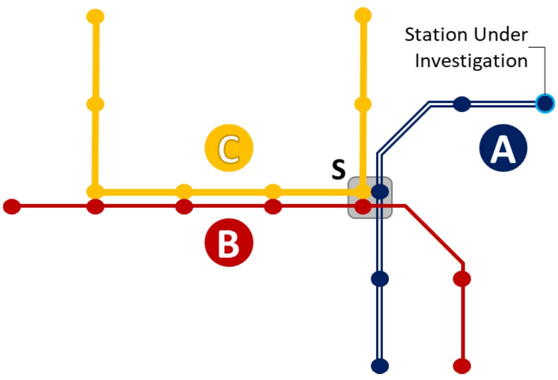
5.4 Identifying Transfers from Travel Pairs

It is possible to detect interchanges and transfers using the primary dataset, or indeed any tertiary dataset with stop-pairs such as the SEQ stop-pair dataset. The technique is quite simple and was defined while attempting to categorise the number of passengers disembarking at Rhodes in the morning peak hour into northbound and southbound travellers (Glazebrook et al. 2017). Thorpe (2017) undertook an analysis that showed many rail passengers on the Eastern Suburbs Line transferred (Thorpe 2017, p. 76).

Figure 90 below provides an example network to show the technique. In this example, the analysis will determine the number of transfers that occur at station S. Step one is categorisation of all stations that have services passing through the station under investigation — call this group A. Thereafter, all other stations can be categorised as group B, or for further accuracy, broken down into individual lines.

The next step is categorising each record in the stop-pair dataset with an embarkation category and disembarkation category. Any record with a category pair of A-A, B-B, or C-C is not a transfer while all other category combinations are. As shown in Figure 90 below, this framework can be expanded to detect different types of transfer.

Figure 90: Example of detecting transfers from Stop Pairs



Embark	Disembark	Transfer	Transfer to A	Transfer at S
A	A	No	No	No
A	B	Yes	Yes	Yes
A	C	Yes	Yes	Yes
B	A	Yes	Yes	Yes
B	B	No	No	No
B	C	Yes	No	Maybe
C	A	Yes	Yes	Yes
C	B	No	No	Maybe
C	C	Yes	No	No

The categorisation starts with a stop pair dataset that contains three fields (columns) the embarkation-stop identifier, the disembarkation-stop identifier, and the count of journey segments undertaken between those stops. Additional fields for the date and the period of travel may be desirable, depending on the analysis needed.

The next step is to create a table where the first field contains all the stop identifiers and the second field contains their categories. Then left joins matching the embarkation stop identifier will determine the embarkation category, and then a second left join matching the disembarkation stop identifier will determine the disembarkation category.

This analysis can be undertaken with simple tools, such as Microsoft Excel using Power-Query. The pseudo code-for that sequence of operation follows:

```
CREATE TABLE segments ( on_stop number, off_stop
number, count number ); LOAD segments;
CREATE TABLE categories ( stop number, category
character ); LOAD categories;

interim = SELECT segments.on_stop, segments.off_stop,
segments.count, categories.category FROM segments LEFT
JOIN categories ON segments.on_stop = categories.stop;

final = SELECT interim.on_stop, interim.off_stop,
interim.count, interim.category, categories.category
FROM interim LEFT JOIN categories ON interim.off_stop =
categories.stop;
```

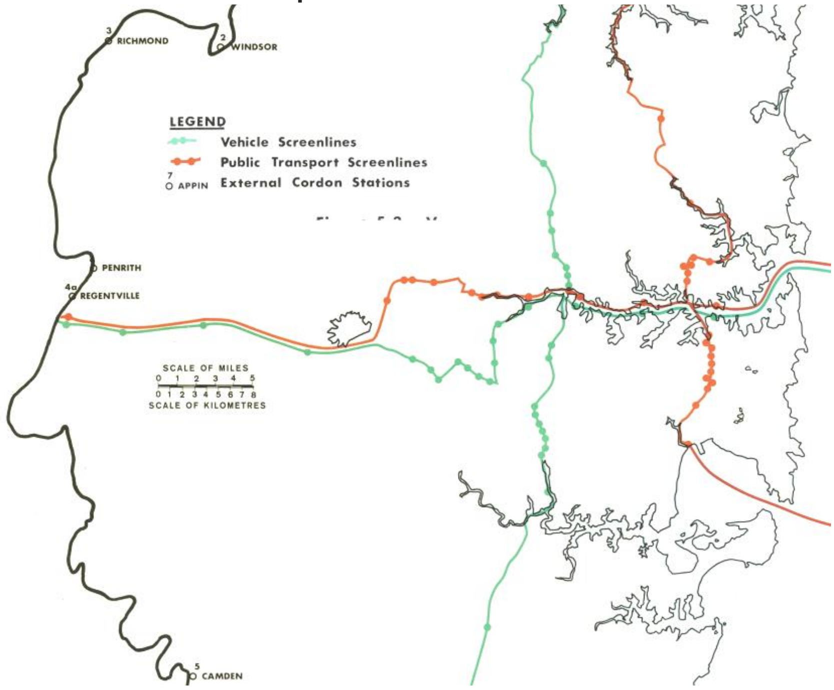
5.5 Public Transport Screenlines

As well as Flow Maps, there is another venerable and elegant technique for assessing passenger and motorist behaviour – screenline counts. It is possible to extend the categorisation techniques to allow for the easy creation of screenlines during transport analysis using the primary dataset. Techniques for doing so are described below.

5.5.1 What is a Screenline?

> ‘When testing for induced traffic growth within a city system using traffic volume counts, boundaries called screenlines, or traffic cordons, are often used to gauge general changes to movement patterns. A screenline is a conceptual line drawn through a section of the urban transport system that attempts to capture all traffic movements between the same broad set of origins and destinations. Where several different routes could be taken for the same trip, a screenline draws a boundary across all of these, thereby accounting for all the traffic moving in and out of the areas on either side’ (Zeibots 2007).

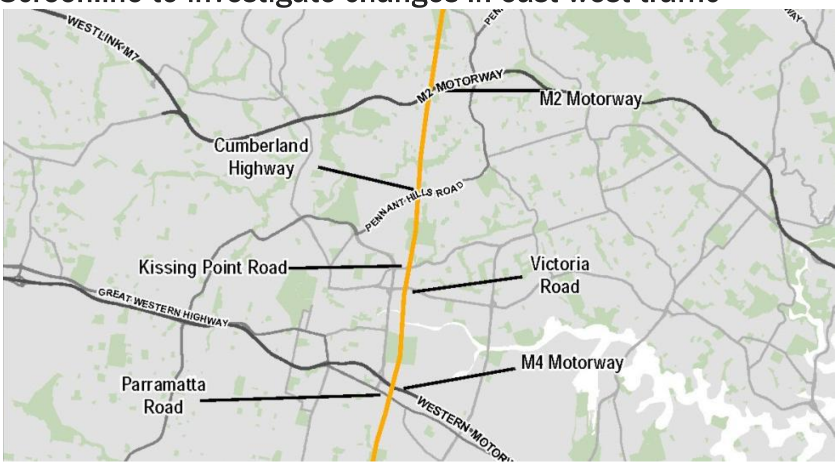
Figure 91: Vehicle & Public Transport Screenlines and External Cordon Survey



Source: (NSW 1974a, Figure 5.3) Study Area, 1971

Screenlines are a major component of the NSW approach to assessing the impacts of major transport projects such as the Westconnex motorway (RMS-WDA, Jacobs & AECOM 2014, p. 144, Chapter 7). Table 45, Table 46, and Figure 92 below outline the assessments developed for RMS as the proponents for the Westconnex motorway. Westconnex was developed by the agency in 2011-2012 and started with a widening of the M4 East-West Motorway from Parramatta to Homebush.

Figure 92: Screenline to investigate changes in east-west traffic



Source: (RMS-WDA, Jacobs & AECOM 2014, p. 144, Figure 7-9)

> ‘A screenline is an imaginary line that splits a study area into a number of parts, usually along physical barriers such as rivers, railway lines or roads. … Ideally screenlines will measure traffic by all modes that cross it, thus allowing estimates of increases in traffic and mode switching.’ (WSP 2018, p. 9)

The assessment used an East-West screenline running North-South, parallel but east of James Ruse Drive in Camelia as shown in Figure 92. The analysis showed the number of vehicles crossing East-West or West-East over the screenline when using key arterial roads and highways. Table 45 and Table 46 below show that after the proposed widening, the number of vehicles using the M4 to cross the screenline would decrease compared to a scenario without the M4 widening.

The EIS analysis assumed that mode-share would remain largely unchanged despite increases in travel time resulting from increased congestion. As such, the model assumed private-car demand would grow linearly with population. In addition, the modelling used 2006 as a base year, however, the evidence by 2014 was that the rate of private-car use during the 2006 census was unusually high and has since decreased. These are important considerations when examining the tables below.

Table 45: Screenline volumes (number of weekday vehicles) in 2021

Road in 2021	Without M4 Widening ('do minimum')	M4 Widening	Difference
M4 Motorway	179,620	114,890	-64,730
Parramatta Road	43,990	59,370	15,380
M2 Motorway	118,050	123,940	5,890
Victoria Road	60,440	70,250	9,810

Table 46: Screenline volumes (number of weekday vehicles) in 2031

Road in 2031	Without WestConnex (Future 'do minimum')	Full WestConnex	Difference
M4 Motorway	194,180	168,760	-25,420
Parramatta Road	52,030	62,490	10,460
M2 Motorway	140,430	140,840	410
Victoria Road	68,250	75,770	7,520

Source: (RMS-WDA, Jacobs & AECOM 2014, p. 144, Table 7-2) and (RMS-WDA, Jacobs & AECOM 2014, p. 145, Table 7-3), from WestConnex Road Traffic Model, 2014 developed by Jacobs SKM using 2011 STM.

The model estimated significant shifts of vehicles to the parallel arterials of Parramatta and Victoria Roads. The EIS indicates changes in route choice would occur due to the introduction of a toll on the M4, increasing the perceived cost of that route compared to the un-tolled Parramatta and Victoria Roads. The EIS estimated a reduction in traffic in the corridor because of the reintroduction of tolls (Hounsell 2014).

Planning for the transport network at that time was supposed to be conducted by TfNSW who would consider all modes as an integrated network. Thus, the ideal screenline analysis would estimate all passenger movements across the line. However, the Westconnex analysis did not estimate patronage changes from the planned project on the parallel railway lines or the planned North-West Metro. In fact, the EIS notes that no estimates on the impact of tolls considered modal split and movement to parallel railway services. The Westconnex was announced in 2012 and work began in 2014, with planning for the project completed by the end of 2019.

> 'The rail passenger demand forecasts used in the development of Sydney's Rail Future (TfNSW 2012) have recognised WestConnex however the detailed impact of, for example, the tolling strategy on the widened section of the M4 Motorway on modal split and therefore passenger demand for rail services would require further detailed analysis' (RMS-WDA, Jacobs & AECOM 2014, p. 187).

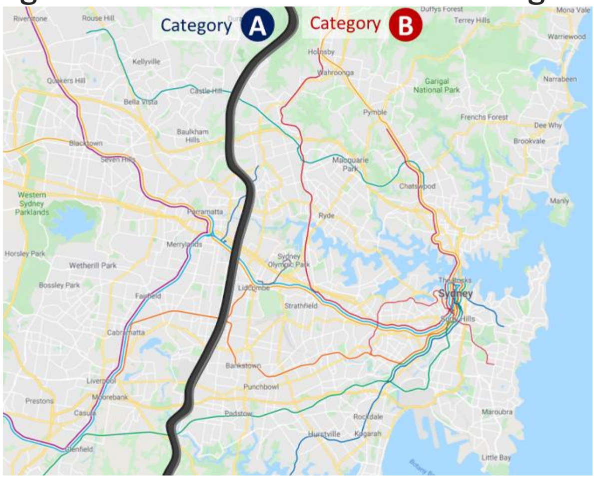
The richness of the primary dataset allows categorisation to provide an elegant mechanism to measure public transport patronage across a screenline. That mechanism should be used to assess the volume of people crossing a screenline, not just vehicles. Measuring the actual volume of people is essential when evaluating transport projects in order to find the most benefits that could be derived from limited investment. An investment analysis requires the measurement of both vehicles and rail passengers.

5.5.2 Screenlines in TOTOR Analytics

At its simplest, a screenline divides an area into two categories – A and B. In the case of the Westconnex M4 Widening, those categories were East (A) and West (B) as shown in Figure 93 below. As shown in Figure 91 from SATS, an analysis is likely to have multiple screenlines. In Sydney, a natural screenline exists for North -South travel due to the harbour and its limited number of bridges and ferries.

The embarkation and disembarkation pair technique identifies trips that cross the screenline using a variant of the technique used to identify interchanges. The technique requires an additional two field table (stop key and category) for all stops under consideration. Then the mobility dataset is filtered to the time period and two left joins are used to categorise the embarkation and disembarkation stops. Thereafter all the records that describe an A-B or B-A category journey segment have crossed the screenline. The counts are summed to calculate the persons crossing during each period.

Figure 93: Using categories to create screenlines to investigate changes in transit



This screenline technique can be extended further to multiple screenlines as well as adapted to intersection analysis. Figure 94 below shows how multiple categories allows an embarkation and disembarkation pair analysis to determine if a journey segment crosses multiple screenlines.

Figure 94: Using multiple public transport screenlines

Embark	Disembark	Cross Screenline A:B	Cross Screenline B:C
A	A	No	No
A	B	Yes	No
A	C	Yes	Yes
B	A	Yes	No
B	B	No	No
B	C	No	Yes
C	A	Yes	Yes
C	B	No	Yes
C	C	No	No

Note that with a large number of categories it may be faster and simpler to run the test iteratively using set logic. For example, testing the A-B crossing as A versus B∪C and then testing the B-C crossing as C versus B∪A.

In addition, screenlines can be used to assess public transport alone. For example, a natural screenline exists on the Oxford St – Bondi Rd buses at the Bondi Junction Interchange (BJI) (Hounsell 2018b). Further research is warranted to determine how many inbound bus passengers choose to transfer to the train versus how many cross the BJI screenline and where they choose to disembark — is the 333 really two bus routes?

By restricting the analysis to services using a given network-link, the screenline technique can be used to determine demand for that network-link. In Figure 94 above, the B:C screenline for train journey segments serves as a proxy for measuring the demand on the Miranda-Caringbah railway link, and the A:B screenline serves as a proxy for the Sutherland-Kirrawee link. In this way, the A:B screenline could also be used to measure the demand from the Cronulla branch to access the rest of the city.

5.5.3 Sydney CBD Screenline

This technique is very flexible and can be adapted for use in different analyses. It takes advantage of the numbering convention used by TfNSW to categorise transit stops that are within the Sydney CBD. The categorisation allows for counting the number of journey segments per hour per stop broken down into either "To the CBD" or "To Elsewhere" categories. Breakdowns for every stop that count the number of embarkations heading to the CBD or not, would be a useful tertiary dataset. It describes macro behaviour because it does not describe actual stop pairs. The following python code was one way the categorisation was implemented:

```
# if a stop begins in 2000 it is in the CBD or the edge
of Pyrmont or the middle of the harbour

stops_not_in_cbd = { 200097, 2000111, 2000144, 20009,
2000262, 2000263, 200088, 2000265, 2000264, 2000266,
20007, 200092, 200091, 200015, 2000278, 2000410 }

def is_in_sydney_cbd( stop_number_text ):
    postcode = int(stop_number_text[0:4])
    if postcode != 2000:
        return False
    return postcode not in stops_not_in_cbd
```

The results below are from ongoing research into the IWLR for 8 AM on the day surveyed. The number of Tap-Offs outside the Sydney CBD was over 18,000, and only 8,000 of those came from inside the Sydney CBD. This is consistent with the majority of persons using the IWLR disembarking at the Pyrmont Bay stop and walking across the Pyrmont Bay Bridge. The 8,000 are from a significant number of passengers that use the IWLR to travel from Central to work in Pyrmont and Star City.

tap	mode	location	card	hour	all segments	intermodal	Sydney CBD Tap On
off	Light rail	Not Sydney CBD	Adult	8	18,210	5,670	8,020
off	Light rail	Sydney CBD	Adult	8	8,650	880	3,280

Research was undertaken into the feasibility of applying this technique throughout the coverage area and the above code was used to assess the number of journey segments travelling to the Sydney CBD or not. Table 47 below is an indication of the results possible and shows the number of embarkations outside the Sydney CBD during that day in November 2016, along with a count of the number of disembarkations in the CBD. Initial results indicated more research is warranted.

Table 47: Number of embarkations not in CBD with disembarkations in the CBD

Tap	Mode	Location	Hour	Segment Count	Intermodal Continuation	Sydney CBD Tap-Off	Sydney CBD Tap-Off as %
on	Light rail	Not Sydney CBD	8	25,930	1,650	8,050	31.0%
on	Ferry	Not Sydney CBD	8	45,570	3,430	39,850	87.4%
on	Bus	Not Sydney CBD	8	1,723,970	93,810	354,030	20.5%
on	Train	Not Sydney CBD	8	1,933,450	198,620	1,054,620	54.5%

A CBD-Or-Not embarkation-disembarkation-pair dataset is a perfect example of a tertiary dataset that could be released to allow service partners and the community to understand passenger responses to the transport network available. In addition, further categorisations/screenlines could be applied to understand travel to and from major centres within NSW, such as the Parramatta CBD and Macquarie Park.

5.6 Estimating Passenger Load Over Intersections Screenlines

Since it is possible to use categorisation to determine the number of journey segments crossing a screenline, it is also possible to determine the number of passengers using a network-link (road or rail). Consequently, it is possible to compute the number of passengers crossing through a signalised intersection in an hour.

Ignoring the issues of the economic, environmental, health, and many other impacts of private cars and focusing solely on the number of persons using each network link for mobility, it becomes a simple mathematical question as to whether or not public transport deserves priority over private cars.

Respected City Planner Brent Toderian states ‘The truth about a city's aspirations isn't found in its vision. It's found in its budget’.[32]

In order to choose which projects to spend taxpayer’s money upon, service partners must evaluate alternatives and rank them. The prioritisation will be based on their values-framework. As such, budget expenditure is an empirical demonstration of the service partners’ values-framework. Similarly, the relative allocation of citizen’s other resources — land, air, water, quiet and time — by service partners will also demonstrate their values-framework. Thus, if the number of private vehicles were more important to the service partners than the number of people travelling, this would manifest in the majority of funding being spent on improving private vehicle travel instead of improving public or active transport.

The RMS prioritises travel time savings in the overall system by optimising the values of traffic engineering practice: vehicle throughput, vehicle wait times, vehicle flow rates, and vehicle queue lengths.

The following case studies describe methods to analyse the road and tram network to determine the human throughput, human wait times, human flow rates, and human queue lengths. Measuring humans and not vehicles is a key component for service partners to implement an integrated transport approach.

There will always be one, and only one, top priority in resource allocation. Intersection signalling that forces trams to wait full cycles for the right of way are manifestations of the prioritisation of time and land for private car movements. Conversely, an intersection that pre-empts car phases for tram phases has prioritised public transport over private vehicles.

The results of the IWLR optimisation research programme were presented to TfNSW in 2018. The research demonstrated that the signalised intersections on the IWLR require the trams to wait for a prescheduled optional phase in an unchanging cycle, resulting in significant delays and runtime variability for the IWLR services.

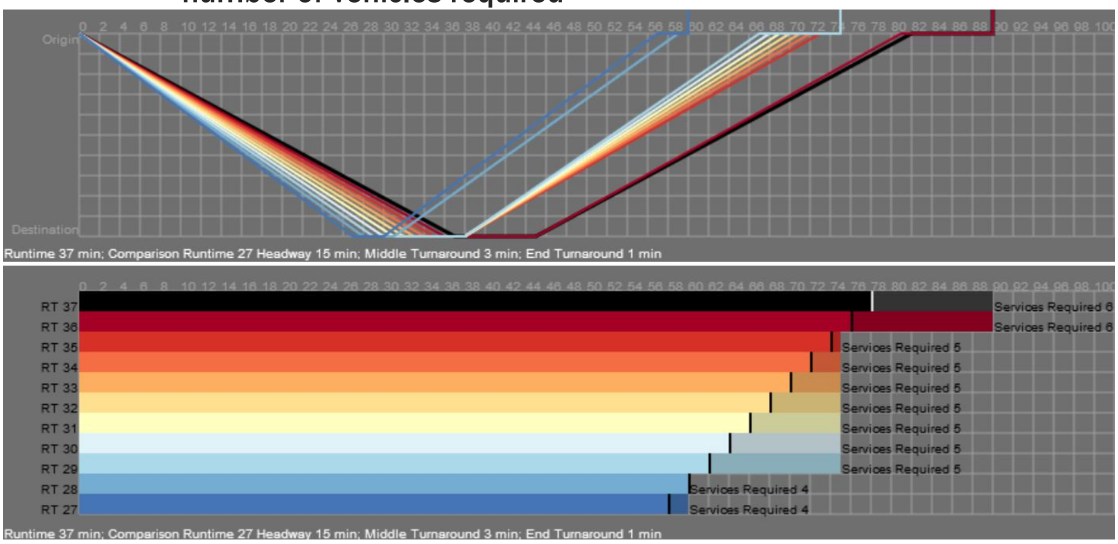
[32] https://twitter.com/BrentToderian/status/704148379182366720

5.6.1 Why Person Throughput Matters

As part of the IWLR optimisation research programme undertaken at UTS, several capstone students applied the analysis methods developed here, with the goal of stabilising and then reducing runtimes on the IWLR.

Figure 95 below is a Marey chart that shows end-to-end runtimes for the number-of-vehicles required to operate IWLR services at 15-minute vehicle-headways. The chart allocated half of the timetable filler to the turnaround-time at either end of the run. The second half of the chart shows the waste time added to the timetable as the vehicles wait for the next scheduled departure. Notice that reducing the running time by two minutes, from 37 to 35, would allow the operator to save an entire vehicle plus crew, allowing that to be reinvested into increased services. This insight drove the investigation into reducing the time the IWLR waits at traffic lights.

Figure 95: Marey chart for the IWLR with different runtimes and the resulting number of vehicles required



The ability to empirically measure the number of people travelling through a signalised intersection on public transport represents a significant technological advance and allows for improvements to road engineering. Previously, the approach was to measure only the number of vehicles through an intersection per hour, focusing on traffic delay, queue lengths and vehicle flow-rates (Roess, McShane & Prassas 2011, Chapter 20). Since the previous approach optimised vehicle throughput, it often resulted in the human throughput of the intersection becoming suboptimal because the larger numbers of vehicles were favoured over larger numbers of people in high capacity vehicles.

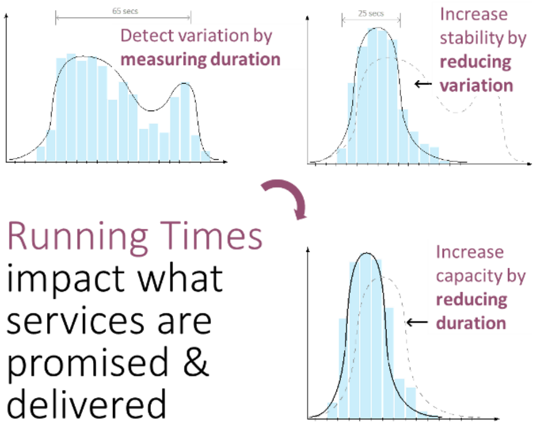
An initial site analysis was undertaken by walking the 300m from TfNSW, to examine the IWLR on-street performance. It was immediately obvious that the signalised intersections had prioritised private vehicles over the IWLR. When a tram arrived at the signal it would have to wait up to a full cycle before it was given a clear signal.

That was especially true for the intersection between Darling Dr, Hay St, and the light rail corridor (see Figure 101 below). Darling Dr was observed to be a minor road with minimal road traffic volumes, however the service partner had prioritised the majority of the signal time for small numbers of road vehicles.

The operator reported that after many years of problems, the trams are still only given one chance to clear an intersection. If a tram could not proceed the first time because a car was blocking the intersection, the signal programming never gives the tram another chance to proceed. This periodically results in stuck trams where trams are left waiting indefinitely unless the operator gets the Transport Management Centre to manually override the programming for that intersection.

Smalley (2017) undertook a capstone project that investigated delays at intersections and proposed changes to on-street signals to stabilise runtimes on the IWLR.

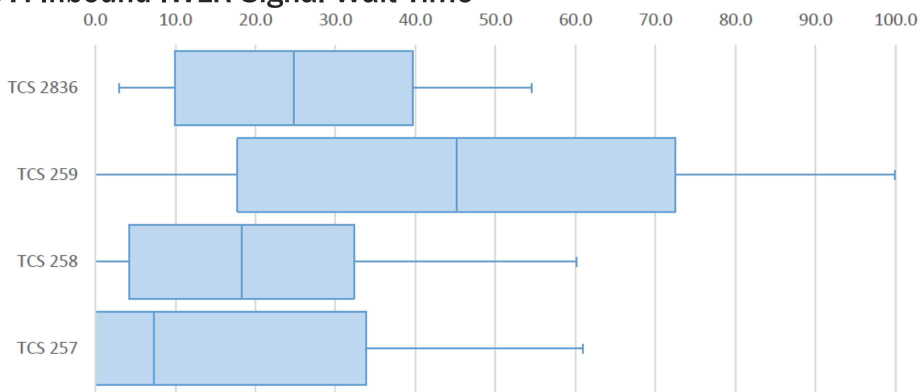
Figure 96: Measure, Stabilise and Reduce



Original Source: Dr Michelle Zeibots

To measure the variability of delays incurred at intersections by IWLR operations, Smalley undertook a survey of the delays at intersections. Figure 97 below shows that delays are not stable, that is because the occurrence of a delay is functionally random, the duration of the delay is not controlled destabilising the running times and timetable.

Figure 97: Inbound IWLR Signal Wait Time

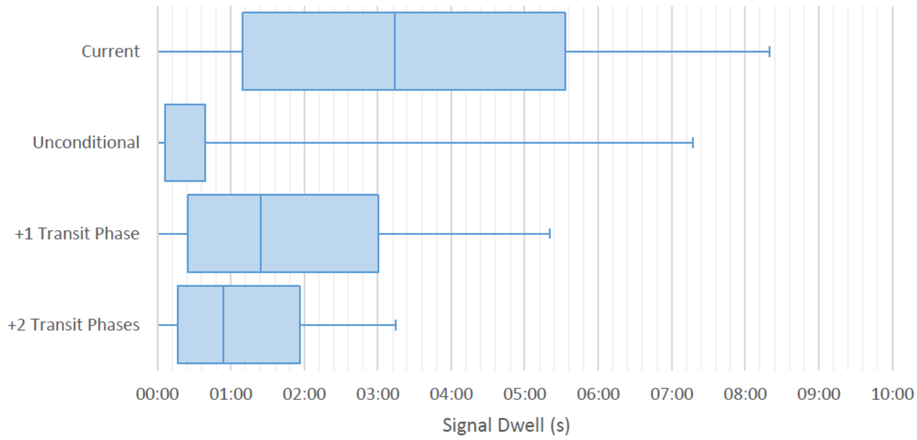


Source: (Smalley 2017, p. 46, Figure 4.11)

#2836 - Darling / Hay. #259 - Sussex–George / Hay. #258 - Pitt / Hay. #257 - Castlereagh / Pitt

Smalley proposed and modelled four options: do nothing, unconditional tram priority, two optional tram phases per signal (+1), or three additional optional tram phases (+2). The results of his modelling with SCATS Access 2017 are shown in Figure 98 below.

Figure 98: Proposed IWLR Wait Consistency



Source: (Smalley 2017, p. 76, Figure 5.2)

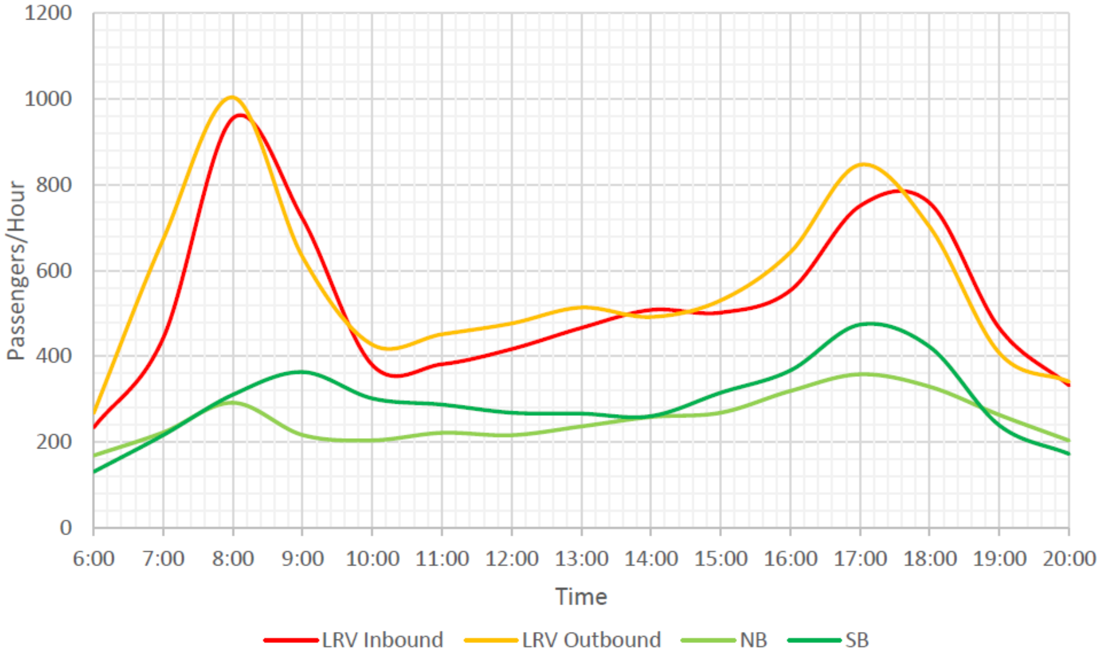
As shown in Figure 98 above only 25% of trams would wait more than 2 minutes for traffic lights if there were three optional tram phases (+2) at each traffic signal. That stands in stark contrast to the current situation, where 25% wait more than 5.5 minutes at traffic lights, and 50% of trams wait at least 3 minutes.

Reliably saving two minutes off the runtime would save one vehicle from the timetable. Therefore, Smalley's analysis shows that adding two optional transit phases into the signal cycles would be enough to save an entire vehicle. Subsequent research has shown there are many other additional savings that could be achieved on the IWLR, to ensure the runtime was further reduced thereby improving services and customer experience.

5.6.2 How to Measure Person Throughput

It was thought the RMS might be swayed by an argument comparing the number of people crossing an intersection in trams versus the number crossing in private vehicles. Using the technique discussed in Section 5.5.2, Smalley (2017) estimated the number of persons on the IWLR, as shown in Figure 99 below. His results clearly show that the number of passengers in the light rail shown in yellow and red crossing the screenline at the signalised intersection is higher than the number of persons in private cars shown in green. This occurs despite the majority of passengers travelling inbound on the IWLR disembarked at Pyrmont Bay before the on-street section containing the intersections.

Figure 99: Darling/Hay Intersection - TCS 2836 - Hourly Passenger Throughput



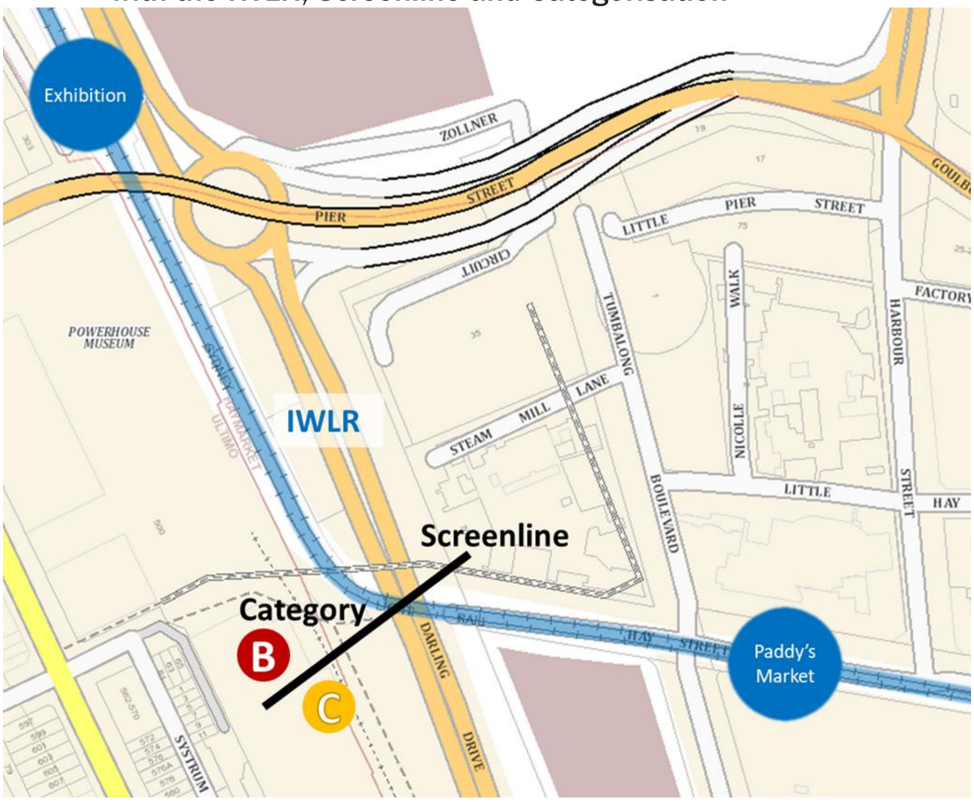
Source: (Smalley 2017, p. 63, Figure 4.19)

The IWLR only has 23 stops. At the Darling Dr – Hay St intersection there are three stops before the intersection and twenty beyond. Figure 100 below shows the stops while proceeding inbound from the terminus at Dulwich Hill to the balloon loop at Central Station. The Darling Dr – Hay St intersection is located between the stops at Exhibition Centre and Paddy's Market. Thus, all the stops before the screenline are given one category (yellow) and all the stops afterward are given another category (pink). Any journey segment from a yellow stop to a pink stop was a person travelling outbound through the signalised intersection, and any segment from pink to yellow was a person travelling inbound after which it is a matter of counting the segments.

Figure 100: IWLR stops in sequence broken into categories for Darling St screenline

Dulwich Hill	Dulwich Grove	Arlington	Waratah Mills	Lewisham West	Taverners Hill	Marion	Hawthorne	*More→ Inbound*
← More Outbound	Leichhardt North	Lilyfield	Rozelle Bay	Jubilee Park	Glebe	Wentworth Park	Fish Market	*More→ Inbound*
← More Outbound	John Street Square	The Star	Pyrmont Bay	Convention Centre	Exhibition Centre	Paddy's Markets	Capitol Square	Central Station

Figure 101: Layout of the Darling St and Hay St Signalised Intersection overlaid with the IWLR, Screenline and Categorisation



Map Source: (NSW 2019a)

This technique can be generalised to use the primary dataset to determine the public transport passenger volumes across any signalised intersection in the city. Then a customer centric integrated transport department could then make a decision to optimise intersections based on the volume of people moving through the intersection and not the volume of vehicles.

With intersections optimised, the runtime for buses throughout the city would be stabilised. Therefore, unnecessary filler in the bus timetables could be removed and the buses would spend more time providing mobility and access to passengers and less time waiting at traffic lights for minor roads that only move small numbers of people by car.

Of course, optimisation of vehicles must consider all customers and beneficiaries of the transport network, including pedestrians, cyclists, logistics and local businesses, as pointed out by Matt Faber in private correspondence.

> ‘By applying a Movement and Place approach “a customer-centric integrated transport department” would not necessarily want to optimise the volume of people moved through an intersection, as this approach would fail to account for the importance – at certain times and in certain places – of delivery and service vehicles being advantaged. Also, providing less intersection priority for “minor roads that move small numbers of people by car” could have the unintended consequence of reducing accessibility for pedestrians needing to cross the main road’ Matt Faber (2020).

5.7 Passenger Volume Distribution

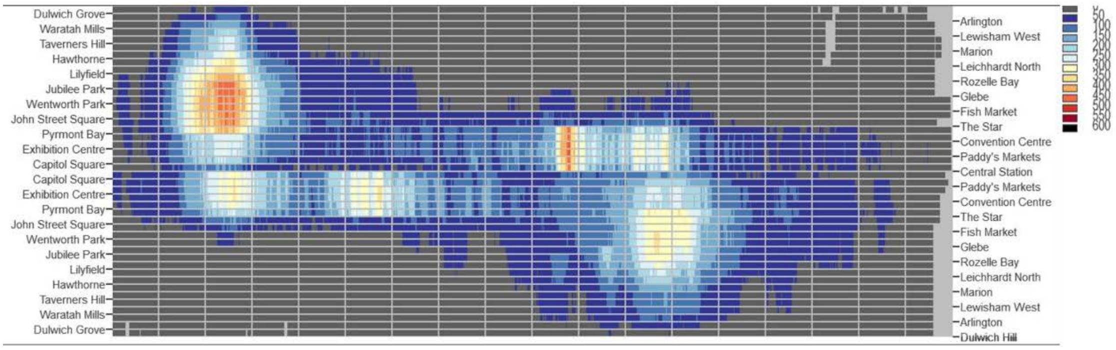
One of the most impactful case studies was the following interactive passenger volume distribution analysis tool that was developed for the opening of the UTS TRC. The visualisation has been shown in interactive form to many transport professionals and the clear rendering of demand per link has allowed them to interpret demand on the IWLR.

The IWLR volume distribution plots were generated using an Opal dataset, however, similar plots could be generated using the primary stop-pair dataset. The visualisation shown in Figure 102 below is from a purpose built web-script that dynamically renders the volumes as an interactive tool for analysing the IWLR.

The horizontal axis represents the time of the day from 06:00 until midnight. The vertical axis shows the demand for the light rail inbound from Dulwich Hill to Central Station in the top half, and demand for outbound journeys to Dulwich Hill in the bottom half. Figure 102 is comprised of boxes that represent the demand (colour) for a single screenline between two stops (vertical) and each box is one-minute wide (horizontal) and coloured in steps of 50 using a table developed by ColorBrewer.

Through the use of colour, this plot is able to convey the relative demand for each link of the IWLR over an entire day. Notice in the middle of the plot a red line between The Star and Central Station. That red line is at 15:47 and represents a surge in passengers. The surge corresponds to the finish of a show at the 2000 seat Lyric Theatre located above The Star tram stop.

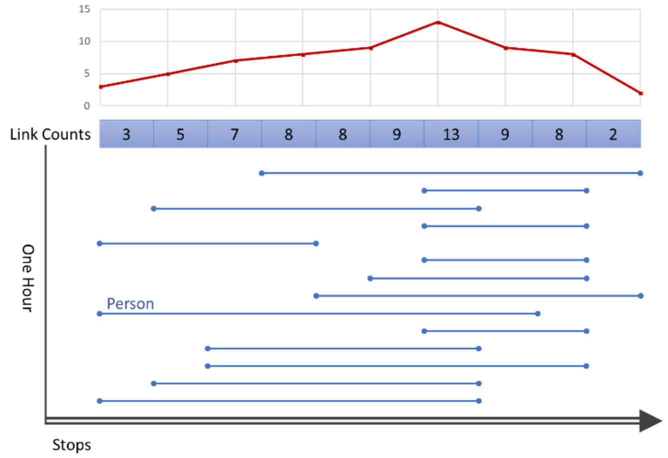
Figure 102: Passenger Load visualisation for the IWLR on Wednesday 2016-03-09



Source: (Hounsell 2017b)

The inputs for this plot are not conceptually difficult to calculate. For a given service operating in a given direction (inbound / outbound) there will be a certain sequence of stops and between each stop is a network link (screenline). When a passenger Taps-On at a certain embarkation-stop, they are expressing demand to travel along a sequence of links to their disembarkation-stop. For each minute between their embarkation and disembarkation time, the algorithm increments the demand counter for all of the links that the person wishes to traverse. This is shown conceptually in Figure 103 below.

Figure 103: Conceptualisation of how to calculate the passenger volume distribution



Source: (Adapted from Hounsell 2018d)

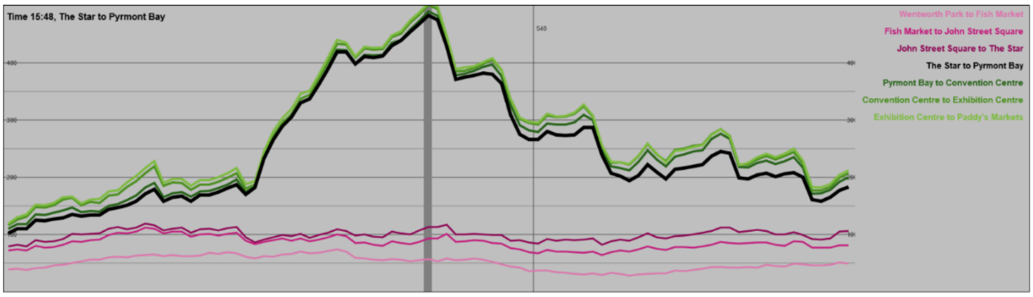
During the demonstrations of this visualisation, a significant criticism was that it did not show the actual loadings on vehicles. That criticism although technically true represents a product-centric assessment of the technique and not a customer-centric assessment. From a customer's perspective, after tapping on, they are in the system and waiting for their mobility demand to be satisfied until they arrive at their disembarkation stop. That is why this visualisation shows both the waiting time and the in-vehicle time as they are experienced by the customer.

The above visualisation does not provide colourisation relative to the actual capacity supplied. A further enhancement would be to colourise each minute relative to the capacity available at that time. Since services are unreliable with full vehicles, that would allow operators to see overloading in relation to the services offered. [33]

The visualisation proved useful as it allowed empirical identification of event crowds on the IWLR during weekday afternoons. After TfNSW were shown this visualisation of the Wednesday peak, they acknowledged there was empirical evidence of high demand between The Star and Central Station. When shown the spike in demand correlated with completion of a show at the Lyric Theatre, they quickly understood the crowding issue. This visualisation resolved many months of dispute over the existence of the crowds. After negotiation with the operator, the partners agreed to an increase in the frequency on Wednesday's afternoons to meet the matinee demand.

The partner's understanding was assisted by additional visualisations that allowed the dataset to be sliced differently and analysed from different perspectives. The interactive visualisation allowed the user to move the mouse around the heat map in Figure 103, and the script would then dynamically update other plots such as Figure 104 below.

Figure 104: Load visualisation for the IWLR on Wednesday 2016-03-09



Source: (Hounsell 2017b)

Figure 104 shows the passenger demand to travel (vertical) from The Star inbound as a thick black line over time (horizontal). The three-preceding network-links are shown as pink lines, and the three subsequent network-links are shown as green lines. This figure shows that there is steady but modest demand for the preceding network-links over the afternoon. It shows there is minimal additional demand from the subsequent network-links. At 15:29 there is a significant surge in the demand for travel inbound from The

33 A frequent service that is not overloaded would optimally service the customers demand with minimal wait time and a fast journey and thus quickly remove them from the load calculation.

Star with network demand peaking at 15:48. This pattern matches what would be expected for an event crowd looking to leave a large capacity theatre (2,000 seats).

This example shows the potential of passenger demand visualisations as an empirical tool for quickly assessing passenger response to the transport network delivered. It also demonstrates the potential for the system to show under serviced demand, which represents opportunities for transport operators to provide additional benefits to their customers in exchange for additional revenue.

5.8 Bus Stop Durations

During discussions with TfNSW, a desire to assess the impacts of the bus stop removal program was identified. A method to estimate the travel time impact of bus stop removals was needed. RMS has been running a programme for many years to remove bus stops to improve private vehicle flows and bus runtimes. The process has often moved or removed bus stops that would be highly valuable to bus customers, especially those stops near corners to allow more space for turning vehicles even though this increases walking distances, decreases transfers, and reduces the catchment.

Murdoch (2017) undertook this as a capstone topic with TfNSW identifying the 373 in Randwick as a particular candidate to be assessed. The following figure shows the route of the 373 in red and the previous bus stops as circles, labelled A, B and D (for Delete).[34] The RMS planned to delete bus stop D with its large walking catchment and short distance to the hospital and light-rail, while retaining bus stops A and B. Stop D shortened one of the forced left turn lanes from Perouse Rd and often resulted in movement conflicts between cars and buses.

When assisting Murdoch (2017), I developed an analytics technique to determine the travel time impact of removing stop B. The technique leverages the domain knowledge that Sydney buses do not stop at a bus-stop if there is no one boarding or alighting.

Each bus run in the Opal dataset is uniquely identified. The technique categorises two types of inbound bus runs in the dataset:

[34] The bus route is forced to go through two right-angled turns and three signalised intersections rather than continuing on the straight path of the tramway it replaced between Belmore Road and Perouse Road.

- IN_B_A_WITH_D –the bus stopped at B, then D, then A
- IN_B_A_NO_D –the bus stops only at B then A, without stopping at D

For each of the categorised bus runs the travel time between the last tap at stop B and the first tap at stop A was calculated. Then the two categories were compared to determine whether there were significant savings.

Figure 105: Route 373 Section Map - Randwick

Row Labels	Grand Total
IN_B_A_NO_D	533
IN_B_A_WITH_D	479
Grand Total	1012

Source: (Murdoch 2017, Figure 14)

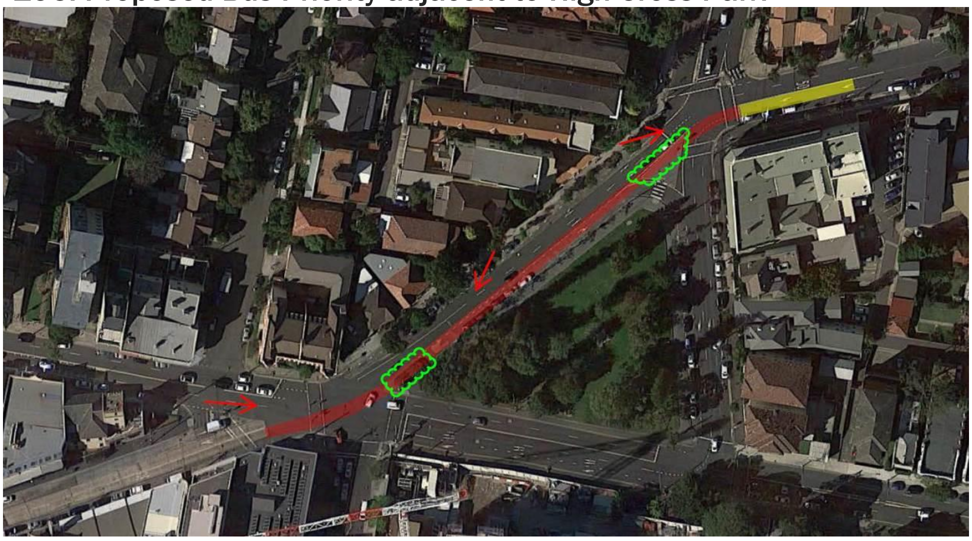
The analysis found reductions in the average runtime between the two stops. However, as Figure 107 below shows, the travel time between stop A and B during the very early morning (when the restaurants are closed and the traffic is light), is consistent and steady with the majority of times being only two minutes despite multiple 90° turns and the traffic lights. In contrast, during the evening when the restaurant strip is buzzing with high numbers of pedestrians, the travel times are more unstable and often double the two minutes required.

The clear failure was the morning peak, where the time taken to travel the 500 m displacement varies significantly. Quite a number of trips took over six minutes to travel the short distance, creating an average displacement-speed of 12 km/h. Analysis indicated that the buses got stuck in traffic and have to wait significantly at all three signalised intersections as the car traffic has priority. Considering those results, it was suggested that Murdoch investigate a contraflow bus lane on the wide Belmore Road next to High Cross Park. He concluded that would be practical and efficient.

This research developed a technique generally applicable for determining travel times and identifying areas for improvement. It was leveraged in a study of the 333 and could allow identification of slow points that prevent optimal operation of the bus network within NSW (Hounsell 2018b).

As can be seen from this capstone, the Measure, Stabilise and Reduce (Figure 96 above) approach is an elegant and powerful framework that operators could adopt to improve bus runtimes and travel outcomes for passengers.

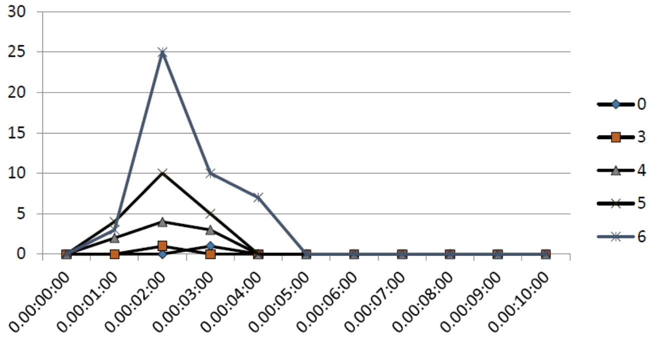
Figure 106: Proposed Bus Priority adjacent to High Cross Park



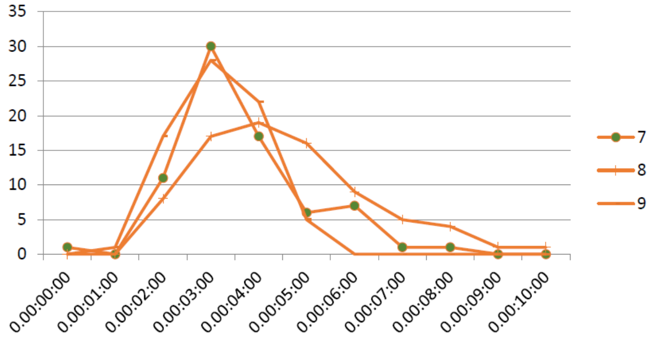
Source: (Murdoch 2017, Figure 19)

Figure 107: Count of time taken to travel between A & B for a 373 by hour

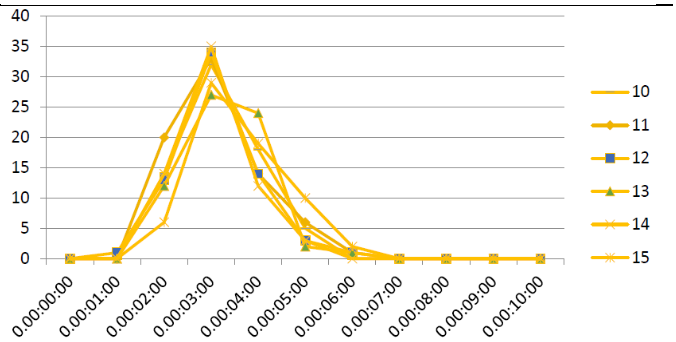
Early Morning



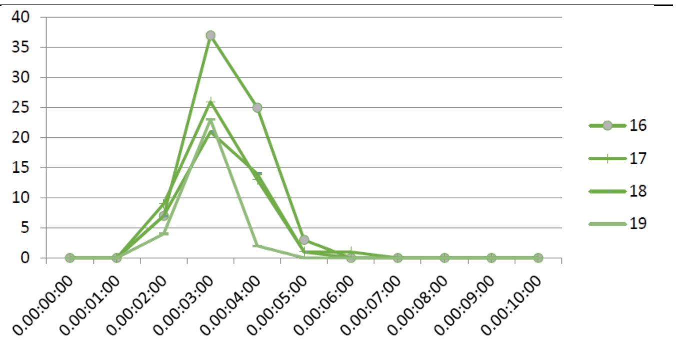
Morning Peak



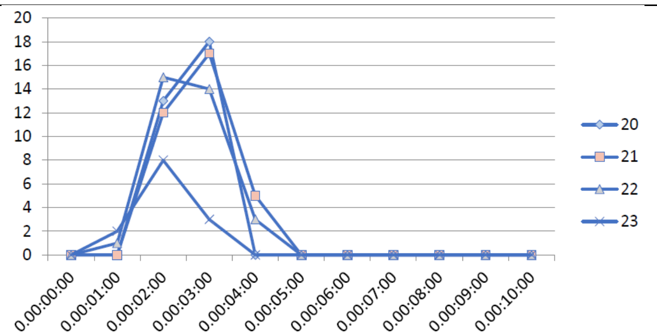
Daytime



Afternoon Peak



Late Evening



Source: (Murdoch 2017, p. 71`, Figures 18.6.1 to 18.6.5)

5.9 Summation

These case studies demonstrate that it is possible to use the primary and secondary datasets to enable an operator moving from a model driven approach to an empirically-driven approach to assess how well the targeted network is actually being delivered. It also enables assessment of how passengers have responded to the services delivered showing the benefits of moving to an empirically driven approach where services can be designed more effectively to meet customer needs and expectations.

TOTOR datasets present a rich high value data source for an empirical approach, but the biographical travel histories contained within create a methodological barrier preventing operators from using them directly. The multistage transformation method proposed here would create privacy protected activity datasets that could be used by operators and their partners to analyse the services delivered and customer response to the services promised and delivered.

There are many analytics techniques available for operators to use for a continuous improvement approach, enabled by privacy protected primary datasets. The analytics techniques include the highly flexible categorisation and screenline techniques with the Measure, Stabilise and Reduce framework as an elegant and effective starting point for a service improvement programme.

END

Part III

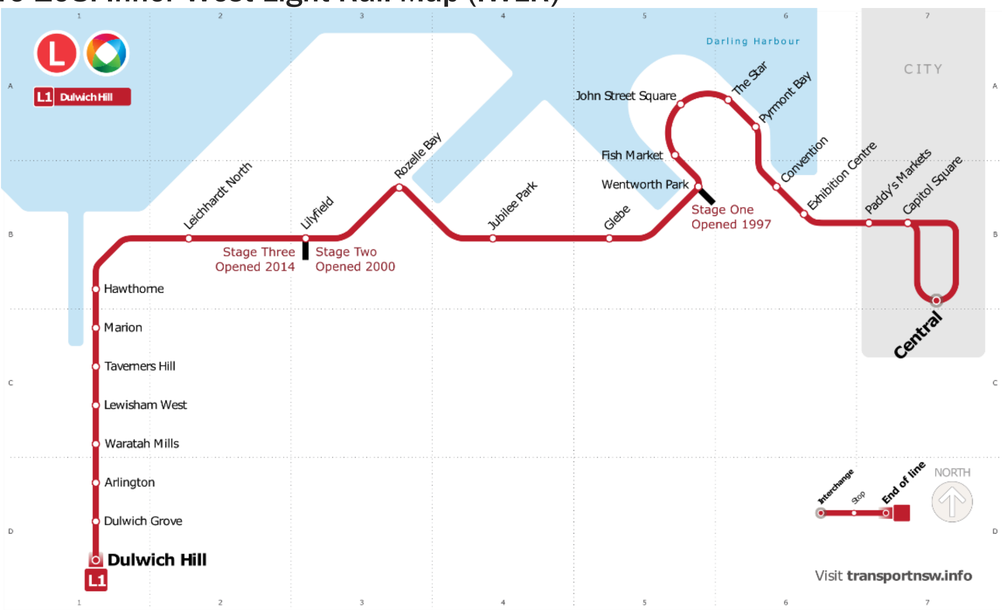
Appendixes

Appendix A – Context Maps

The following maps are provided for your convenience and cover the IWLR, Sydney 333 Bus Route, the Sydney Trains Network (2016 & 2019), and NSW Opal Coverage.

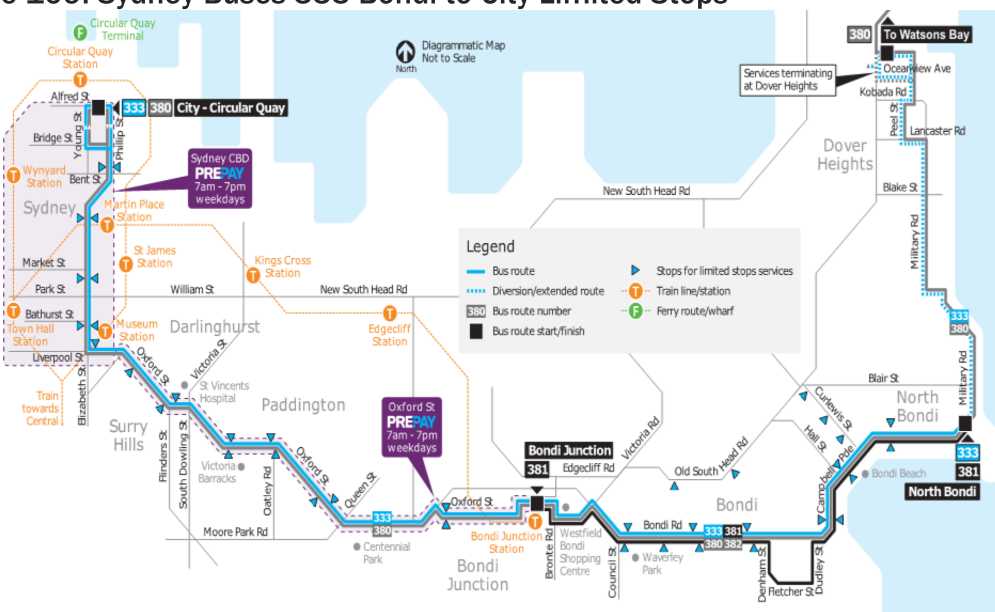
In this thesis Sydney will always mean the Sydney in NSW, Australia. Assume that all locations are in the NSW GMA unless otherwise specified.

Figure 108: Inner West Light Rail Map (IWLR)



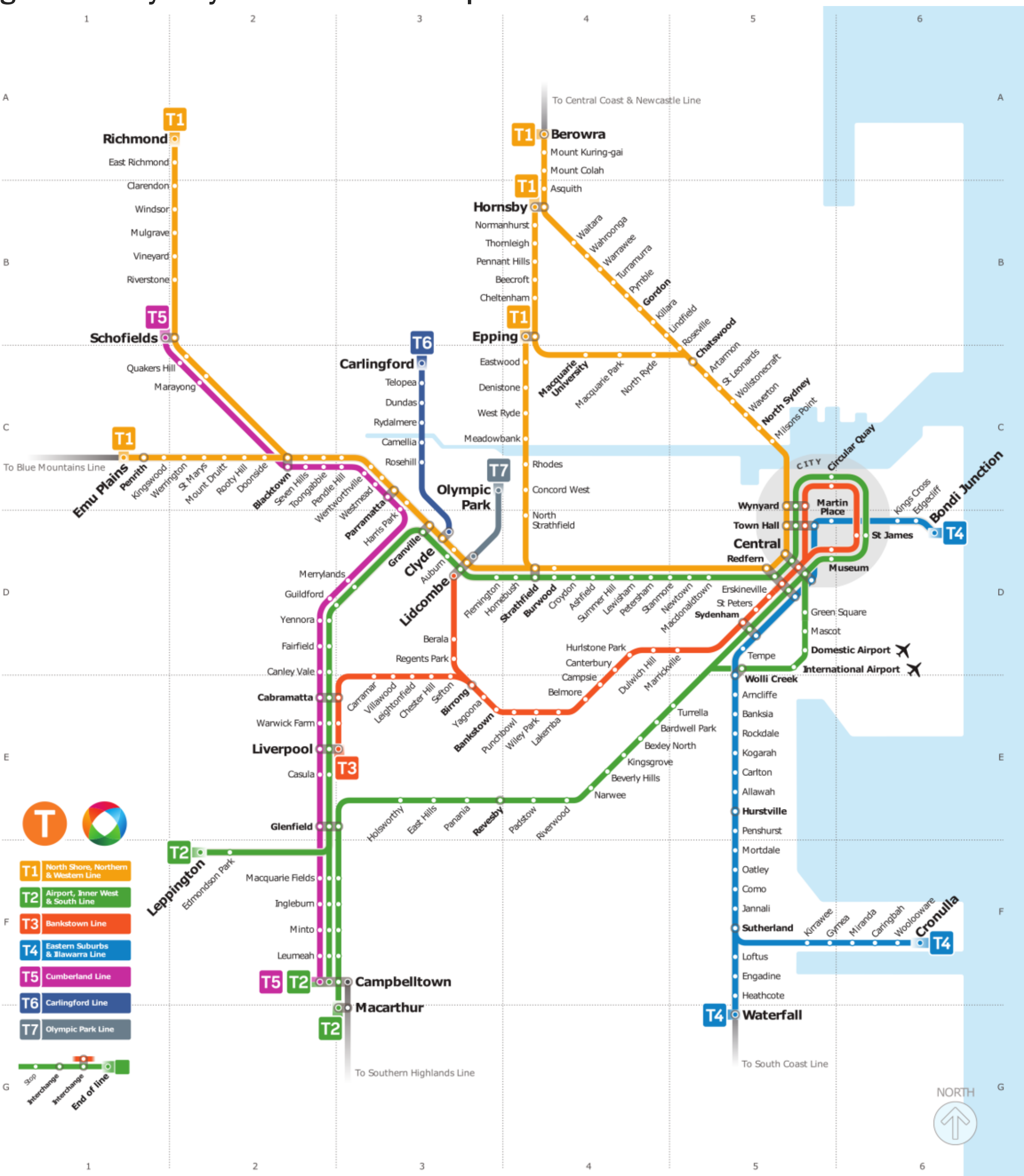
Source: (Hounsell 2018b)

Figure 109: Sydney Buses 333 Bondi to City Limited Stops



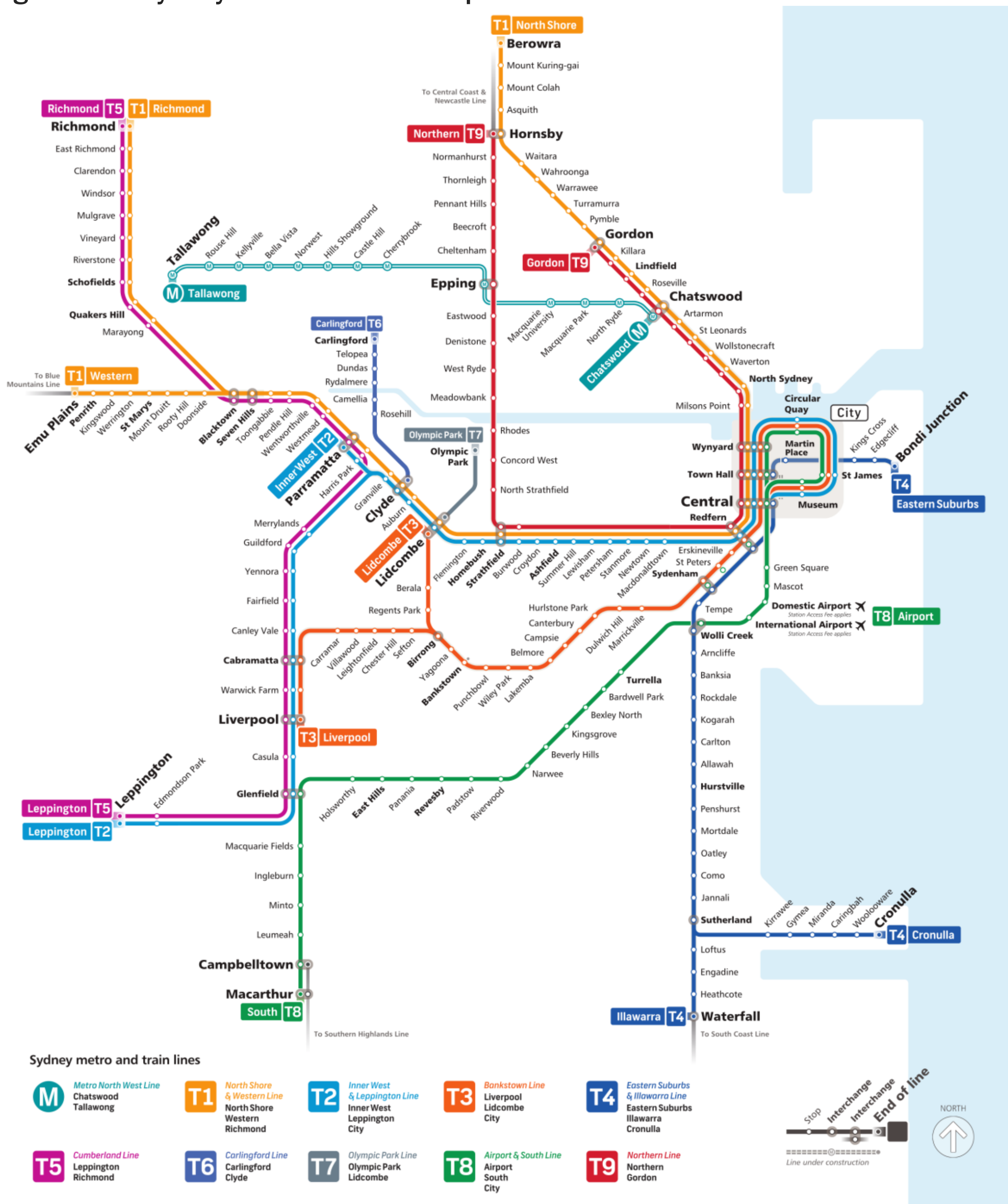
Source: (Hounsell 2018b)

Figure 110: Sydney Trains Network Map from 2016-10-05



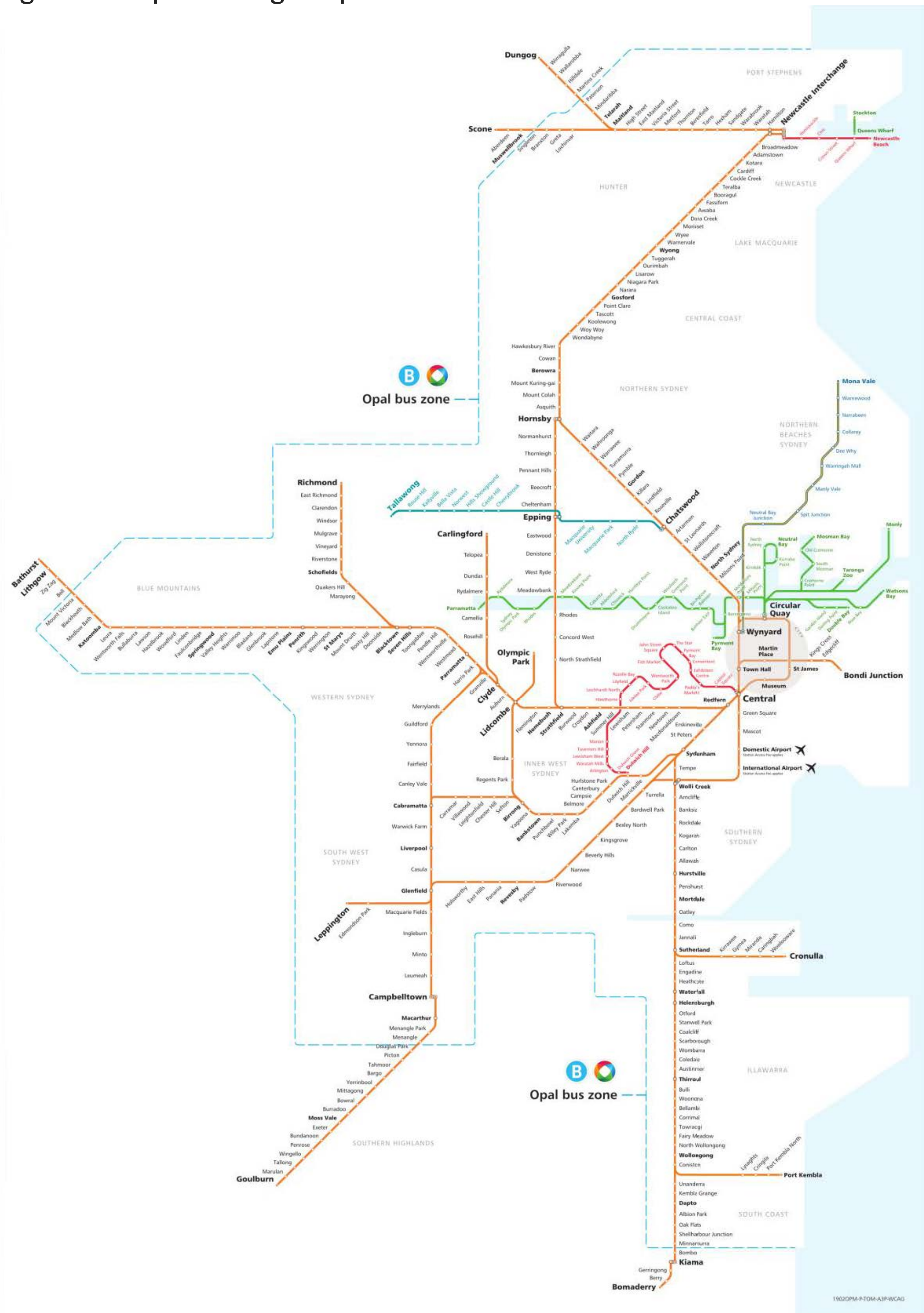
Source: (TfNSW 2016a)

Figure 111: Sydney Trains Network Map from 2019-01-17



Source: (TfNSW 2016a)

Figure 112: Opal Coverage Map from 2019-02-25



Source: (TfNSW 2016a)

Appendix B – Terminology

B.1 Entities

CoA	Commonwealth of Australia
ABS	Australian Bureau of Statistics - www.abs.gov.au - ABS House, 45 Benjamin Way, Belconnen ACT Australia, 2617
BITRE	Bureau of Infrastructure, Transport and Regional Economics - www.bitre.gov.au - 111 Alinga Street, Canberra ACT Australia, 2600
ACT	Australian Capital Territory
NSW	New South Wales
BTS	Bureau of Transport Statistics (now part of TPA)
Opal	New South Wales Greater Metropolitan Area public-transport / mass-transit Electronic Ticketing System for TfNSW - www.opal.com.au -
RMS	Roads and Maritime Services - www.rms.nsw.gov.au - 20-44 Ennis Rd, Milsons Point NSW Australia, 2061
Sydney Trains	Sydney Trains - www.transport.nsw.gov.au/sydneytrains - Customer Relations Unit, Sydney Trains, PO Box 533, Burwood NSW Australia, 1805
TfNSW	Transport for New South Wales - www.transport.nsw.gov.au - 18 Lee St, Chippendale NSW Australia, 2008
TPA	Transport Performance and Analytics - a "Centre of Excellence" at TfNSW - www.transport.nsw.gov.au/data-and-research
UTS	University of Technology Sydney - www.uts.edu.au - Building 1, 15 Broadway, Ultimo NSW Australia, 2007
FEIT	Faculty of Engineering and Information Technology - feit.uts.edu.au - Building 11, 81 Broadway, Ultimo NSW Australia, 2007
ISF	Institute for Sustainable Futures - isf.uts.edu.au - Level 10, Building 10, 235 Jones St, Ultimo NSW Australia, 2007
TRC	Transport Research Centre - trc.uts.edu.au - Building 10, 235 Jones St, Ultimo NSW Australia, 2007
ISO	International Organization for Standardization (Organisation internationale de normalisation)
CEN	European Committee for Standardization (Comité Européen de Normalisation)
	Other
OECD	Organisation for Economic Co-operation and Development (L'Organisation de coopération et de développement économiques) - www.oecd.org - 2, rue André Pascal 75775 Paris Cedex 16, France
Transdev	Transdev Sydney (Light Rail) - transdevsydney.com.au - 220 Pyrmont St, Pyrmont, NSW, Australia 2009 Transdev NSW (Bus) - www.transdevnsw.com.au - 127 Link Road, Bankstown Airport, NSW, Australia 2200 both divisions of the international Transdev Group S.A. (PLC) - www.transdev.com/en - headquartered in Immeuble Crystal, 3 Allée de Grenelle, 92442 Issy-les-Moulineaux, Paris, France.
TUB	University of Technology Berlin (Technische Universität Berlin)
UoW	University of Wollongong - www.uow.edu.au - Northfields Ave, Wollongong NSW Australia, 2522

B.2 Laws

CoA	Commonwealth of Australia
FOI	Freedom of Information Act 1982 - see GIPA
	Privacy Act 1988 - (original) - see PPIPA • see Australian Privacy Principles
NSW	New South Wales
GIPA	Government Information (Public Access) Act 2009 - see FOI
HRIPA	Health Records and Information Privacy Act 2002 (NSW)
PPIPA	Privacy and Personal Information Protection Act 1998 (NSW) • see Information Protection Principles

B.3 Measurement Units

It is important to use a single system of measurement; to prevent ambiguity and thus mistakes. The definitions of measurements are a large and complicated topic, and this thesis will use the latest standards of measurements exactly as defined.

The International System of Units (SI, abbreviated from the French Système international (d'unités)) is the standard system of measurement for scientific and engineering work. On 16th November 2018 the 26th General Conference on Weights and Measures in Versailles, France voted to update the SI standard to be effective from 20th May 2019. For this thesis, the exact definition of the second (as a number of vibrations of a caesium-133 atom) is not required; and the change in definition of the metre and kilogramme is not significant.

Figure 113: The seven SI base units

Symbol	Name	Quantity
s	second	time
m	metre	length
kg	kilogramme	mass
A	ampere	electric current
K	kelvin	temperature
mol	mole	amount of substance
cd	candela	luminous intensity

Source: (Source BIPM 2018)

This thesis will use the units in Australia commonly derived from these SI units, such as minutes, hours, kilometres, and kilometres per hour. That is because those units are widely used in the transport industry and by passengers.

The World Geodetic System (WGS) Is the international standard for navigation, including by the Global Positioning System (GPS). The latest version, known as WGS 84, is the geospatial coordinates (latitude and longitude) used in NSW machine readable timetables and in the Opal dataset. WGS 84 is also known as WGS 1984, EPSG:4326

The realities of our complicated physical environment add interesting caveats to the measurement systems that we use. For example, time is relative to speed. The position coordinates of a bus stop will move as the tectonic plate it rests upon moves. The US Navy periodically adds a second to certain days (leap-seconds); making an abomination called 23:60 by computers. However, within the context of operating a terrestrial public transportation system these caveats and boundary conditions can be ignored.

Other issues must be considered when undertaking a transport analysis, as they will cause errors when using the datasets. For example, Opal is stored in local time and Sydney has daylight savings time, thus the unqualified timestamp of 2020-04-05 02:30 in the database can actually refers to two different minutes an hour apart in reality.

B.4 General Terminology

- **Tuple**
  In computer science, a tuple is a finite ordered list (sequence) of elements.
- **Date Stamp**
  The date recorded for an action, a tuple of the year, month, and day.
- **Time Stamp**
  The date and time recorded for an action. A tuple of the year, month, day, hour, and minute
- **(Database) Table**
  A collection of tuples (rows) sharing the same sequence of elements (columns).
- **(Database) Field**
  A single column in a table with a single type and purpose.
- **(Database) Record**
  A single tuple (row) in a table with reference to a single entity.
- **(Database) Key**
  An element in a record that uniquely identifiers an entity. A field in a database that contains unique ordered identifiers, so a record can be quickly found by searching all keys in a table.

B.5 Transport Terminology

- **Passenger**
  The entity engaged in using the transit system. In Sydney, the tickets can only be used for one individual passenger at a time. Analysis that uses card pseudo-identifiers must consider that a ticket is physically transferrable, even though operators may seek to forbid sharing by regulation.
- **Trip**
  The movement of a passenger from one location to another location using zero or one mass transport vehicles. If a passenger transfers from one vehicle to another the analysis should consider that as multiple trips.
- **Board (verb)**
  A passenger entering the vehicle at the start of a trip.
- **Alight (verb)**
  A passenger leaving the vehicle at the end of a trip.

- **Journey**

  A collection of trips by a passenger for a single purpose, such as commuting to work or going home. Transferring from one vehicle to another is considered making multiple trips.

- **Stop**

  A location where a passenger can board or alight a transit vehicle, such as a station, wharf, or bus stop.

- **Stop Number**

  Each stop in NSW has a spatially unique numeric identifier; they are also given non-unique names. At any moment there will not be two stops with the same identifier. Usually a stop number is persistent; however, TfNSW reuses identifiers if a stop is removed or its location is changed.

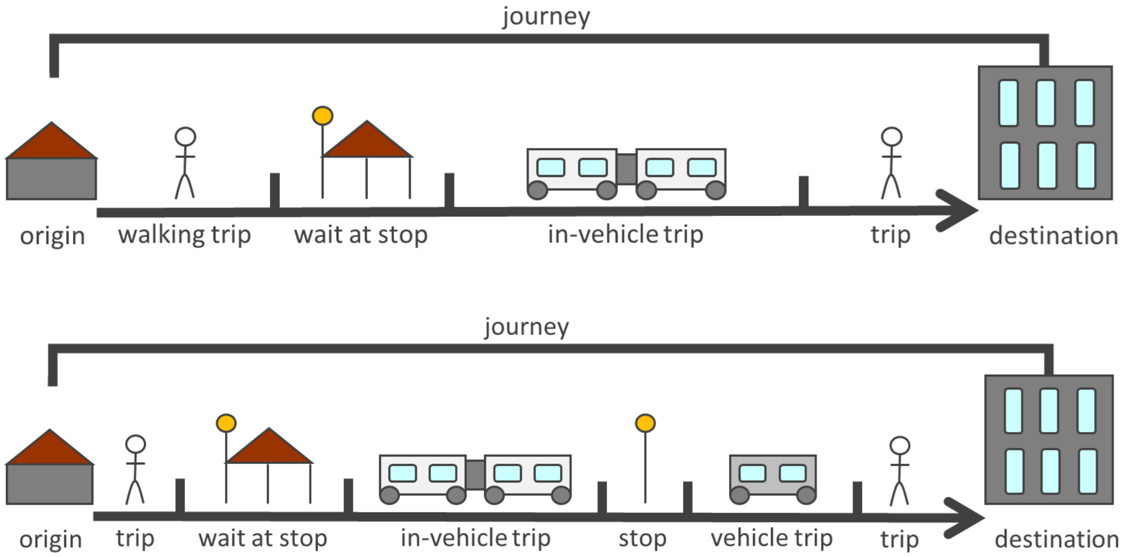
- **Origin**

  The time and location where the passenger started a journey, such as the passenger's home or work.

- **Destination**

  The time and location where the passenger ended a journey, such as the passenger's home or work.

Figure 114: Diagram of a transport journeys from origin to destination.



- **Route**

  A sequence of stops to be visited in order by a vehicle serving passengers; e.g. the 333 Bondi-Beach to Sydney CBD bus route in Sydney (Hounsell 2018b).

- **Vehicle Headway**

  The duration (time) expected between vehicles for a particular route at a stop; usually measured in minutes.

- **Service Frequency**

  The number of services expected for a given route at a given stop within one hour, the functional inverse of Vehicle Headway.

- **Distance**

  The length along the path of the vehicle a passenger moved between stops; see usually measured in kilometres or metres.

- **Displacement**

  The length along a straight arc over the Earth's surface a passenger moved between stops. Calculated using the great circle arc distance between the geolocation of the stops. Displacement measures the passenger's intent and not the operator's network design.

- **Duration**

  The time taken to complete a trip, or journey; usually measured in minutes.

- **Speed**

  The velocity of a passenger's trip; at an instant or over a subsection; usually measured in kilometres per hour (km/h).

- **Displacement Speed**

  The velocity achieved over a passenger's displacement. Calculated using the displacement divided by the duration. Displacement speed measures the passenger's experience and not the systems infrastructure limits.

B.5.1 Ticketing Terminology

- **Authority to Travel**

  An agreement between a service provider and an external person giving the person the right to make either a specified trip or an unlimited trip on the services provide, e.g. a train ticket, or an airline boarding pass.

- **ETS – Electronic Ticketing System**

  The combination of technologies used to process payments, check accounts and provide a valid authority to travel. Passenger tickets in Sydney are now purely virtual instrument and are created upon request and stored on a Ticket Device.

- **Virtual Ticket**

  A ticket is a secure unforgeable record from an authorised issuer that represents an authority to travel using the transit system. Virtual Tickets are purely electronic records that replaced physical records such as magnetic strip tickets and paper tickets. NSW transit uses Virtual Tickets that can only be used by one individual at a time.

Figure 115: Opal Electronic Ticketing Systems

Smart-card ticket devices, with NSW MST tickets.

Opal ticket reader for on–board buses (Giles 2013a)

Opal ticket reader and access control gates at Kellyville Metro station (MDRX 2019)

- **Ticket Device**

  An approved contact-less smartcard, smartphone, or server that is capable of storing Virtual Tickets.

- **Ticket Reader**

  A device with a radio transceiver that authenticates a Ticket Device and manages fare processing. Ticket readers may be either fixed (at stops) or remote (on-board a vehicle). Ticket readers may be free-standing or integrated into access-control / revenue-protection gates.

B.5.2 Transport Analytics Terminology

- **Embark (verb) -- Embarkation**

  Refers to the time and location where the passenger Taps-On. Based on the transit mode either when boarding the vehicle (bus), or when entering the transport stop (train, tram, and ferry).

- **Disembark (verb) -- Disembarkation**

  Refers to the time and location where the passenger Taps-Off. Based on the transit mode either when alighting from the vehicle (bus), or when exiting the transport stop (train, tram, and ferry).

- **Segment**

  A collection of trips by a passenger recorded as an embarkation-disembarkation pair by the ticketing system. Segments may include several trips. It may not be possible to derive the exact trips from a passenger's segment records.

- **Embarkation Stop**

  The first stop in a segment; the stop where the passenger Tapped-On.

- **Disembarkation Stop**

  The last stop in a segment; the stop where the passenger Tapped-Off.

- **Origin Stop**

  The first stop in a journey. In a TOTOR analysis, the Origin Stop can be determined but the Origin cannot.

- **Destination Stop**

  The last stop in a journey. In a TOTOR analysis, the Destination Stop can be determined but the Destination cannot.

- **UNKNOWN (-1)**

  A value that fills records when a value was not known, recorded, or computed.

- **Travel History**

  A biographical record of a person's movement that includes locations and/or timestamps. Since Travel Histories are descriptive of a person's activities, they do not need to record an identifier to be biographical.

Equation 1: Key fleet equations

$r=v\times h$	$h=r/v$	$v=r/h$	$V=\lceil r/h \rceil$

The number of vehicles increases proportionally as the return runtime increases or the vehicle-headway decreases. Return runtime is also called Cycle Time in Vuchic (2005)

Notes on Equations: Return runtime (r), vehicle headway (h), and number of vehicles (v) are positive rational numbers. Integral number of vehicles (V) is a positive integer.

$\{ r \in \mathbb{Q} \mid 0< r < \infty \}, \{ h \in \mathbb{Q} \mid 0< h < \infty \}, \{ v \in \mathbb{Q} \mid 0 < v < \infty \}$, and $\{ V \in \mathbb{N} \mid 0< V < \infty \}$

The ceiling operation finds the smallest integer greater than or equal to the given rational number. Ceiling is written in Vuchic (2005) as []+.

$\lceil v \rceil = \min \{ V \in \mathbb{Z} \mid V \geq v \}$

Appendix C – Explanatory Notes

C.1 Why use the term dis\embarkation?

The implementation of the public transport system and the Opal ETS in NSW has created an analytic ambiguity around the terms origin, trip, and journey.

The NSW system allows a person to Tap-On a train, tram, or ferry and undertake several trips before Tapping-Off. On the railways, the access control gates are at the entrance to stations and passengers can quickly transfer between trains within the station. For example, at several locations, such as Central & Town Hall, Dr. John Jacob Crew Bradfield CMG and his colleagues designed Sydney's railway infrastructure to deliberately facilitate cross-platform interchange (i.e. interchange without stairs).

In addition, trips within the transit network represent only part of the person's journey. A person will at a minimum have to walk from their origin and walk to their destination. Persons can also undertake multiple inter-modal transfers during the journey.

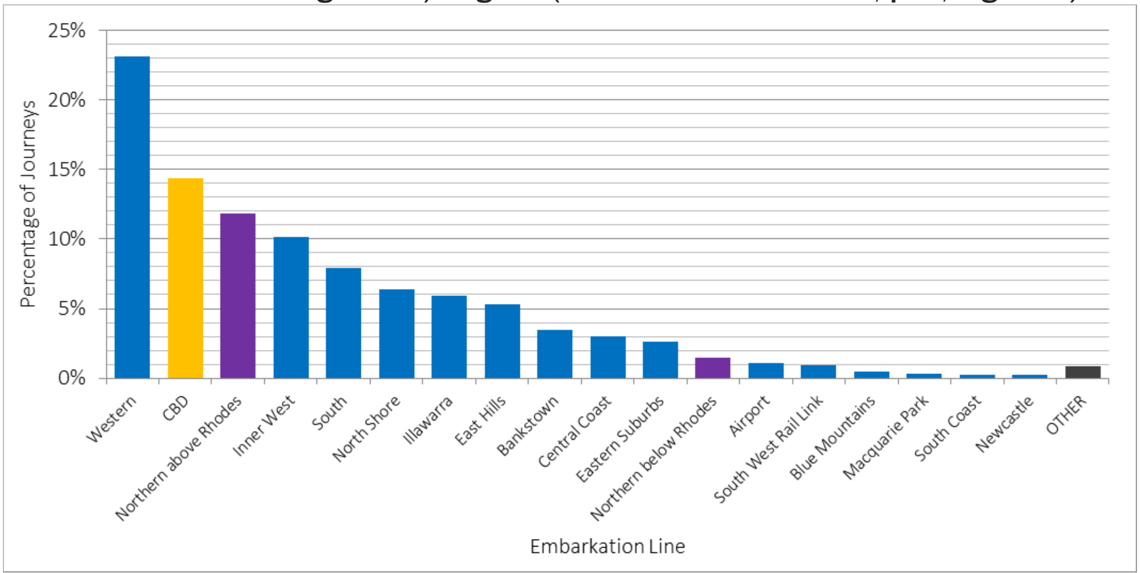
As such, an Opal Tap-On or Tap-Off does not necessarily relate one-to-one to either a journey or a trip. Therefore, the separate terminology of journey segment, embarkation, and disembarkation is required to remind analysts that the number of Taps does not relate 1:1 to the number of trips or journeys.

C.1.1 Examples

Passengers travelling to Rhodes embarked in locations all over the NSW Greater Metropolitan Area (GMA); see Figure 87 above. During an analysis of the flow of passengers in and out of Rhodes station during the morning peak, I developed the embarkation-disembarkation categorisation technique discussed in the case studies above. Using the categorisation technique, the Tap-On line was identified for each passenger Tapping-Off at Rhodes. Rhodes is on the "Northern Line" (purple) in Figure 116. That chart shows that the embarkation–disembarkation pairs represented by the journey segments in many cases represent multiple train trips.

Remember, embarkation and disembarkation analysis, as in Figure 87, cannot determine actual origin location as that could be anywhere within the reported walking distances in Sydney of up to 1,500m; (Daniels & Mulley 2013).

Figure 116: Embarkation line for 08:00 disembarkations at Rhodes (Hounsell 2017a Figure 27) original (Glazebrook et al. 2017, p. 7, Figure 4)



Thorpe (2017) used an implementation of the embarkation-disembarkation categorisation technique to analyse the number of transfers to/from the T4 (Illawarra / Eastern Suburbs) line. Thorpe empirically demonstrated that passengers on the Eastern Suburbs line often transfer and undertake multiple train trips for a single embarkation.

These research projects used empirical measurements to demonstrate that passengers in NSW often transfer between trains in order to access employment destinations. This empirical measurement runs contrary to certain key parameters built into the STM that penalise network designs which utilise passenger transfers for efficiency. The empirical results also show that the expressed preference of Sydney's passengers runs contrary to the stated preferences in the surveys used to construct the four-stage model currently central to NSW departmental planning; (TfNSW 2001a, 2001b, 2018c).

These are good example of empirical data being more accurate at describing reality than the model -- that is, the data more accurately described the operator's Blind-Spot, Unknowns, and Misknowns.

C.2 Data cleaning

There is a risk with each data analysis that the data will contain inaccuracies, measurement artefact, or environmental artefacts (events) that distort the results of a naive analysis. These inaccuracies could cause an incorrect analysis which could lead to a cascade of incorrect actions. The likelihood of this in any uncontrolled system in which humans are a crucial factor in the operations is 100% -- an absolute certainty. Therefore, the risk of an incorrect analysis needs to be controlled.

Fare Minimisation was discussed in Section 1.3.4 above, and is one of the artefacts introduced by human (environmental) behaviour into an Opal dataset analysis. Fare Minimisation is an example is one of the clearest and most impactful artefacts observed with the TOTOR dataset. The artefact is a correct measurement, but it is a distortion created by human behaviour attempting to minimise price.

The way to control for and reduce the risk of data inaccuracies corrupting an analysis, and subsequent conclusions, is to clean the dataset before analysis to remove known errors, and artefacts introduced by the system or by human behaviour.

The severity of the impact is proportionate to the number of incorrect data points that are introduced to the dataset. For example, if the analysis would not be impaired by an error rate of 5% than 1% of data points being distorted may not be significant.

C.2.1 A clean workspace is a safe workspace

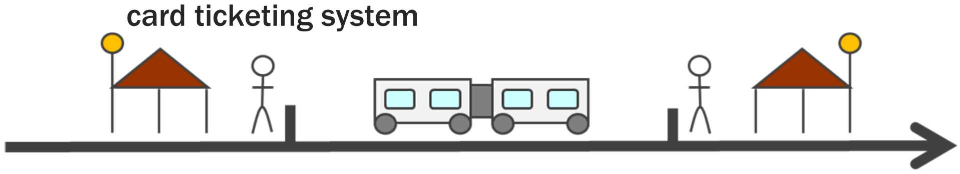
As the Fare Minimisation example demonstrates it is important to understand the limitations and artefacts that exist within each dataset. That is especially important for transport datasets which record human behaviour and operations in a complex environment. There are several limitations and complications with the Opal system that need to be accounted for before analysis can be undertaken. The following sections details some methodological considerations that are required during analysis.

There are significant technical differences between the approaches used in the paper-based ticketing era and the smart-card based ticketing era. The primary consideration is that the authority to travel during the paper era was encoded on a literal ticket and could be verified by the operator examining the literal ticket. Many jurisdictions still use paper tickets today because they are a cost-effective and proven technology.

The smart-card based ticketing systems, such as Opal, use virtual tickets stored on a ticket device (such as smart-cards). However, since the ticket is virtual it is possible to expand the system and store ticket on other devices such as a phone or in the cloud (i.e. on a networked server). In this way, a virtual ticket can be created in the cloud when a customer taps with an arbitrary credit card. Some locations, such as Greater London, are transitioning to cloud-based system to remove the need for dedicated cards.

Figure 117 below outlines the ticketing framework in the paper versus smart-card era. It is necessary to understand the ticketing framework to understand the artefacts the framework can create in the dataset. This is essential for understanding the dataset cleaning required to prevent analysis errors.

Figure 117: Sequence of ticketing events in the old paper versus the new smart-card ticketing system



	Enter Station	In-Vehicle	Exit Station
Paper Ticket	At ticket counter • specify end station • pay fare • take literal ticket	Authority to Travel stored on literal-ticket.	Show or relinquish literal ticket.
Opal	Tap-On at ticket reader • reader opens new virtual ticket on ticket device (card)	Authority to Travel stored in virtual ticket on device.	Tap-Off at ticket reader • reader closes virtual ticket • reader deducts fare from device balance based on distance

There are certain artefacts that can be created in the dataset from humans interacting with the technology in a non-compliant manner. Figure 118 below outlines the sequence of events in the compliant case; and in the case where the customer does not Tap-Off, or does not Tap-On. Since customers can fail to Tap, the Opal dataset will contain artefacts where the location of a Tap is registered as UNKNOWN.

As an example artefact, if there are 100 passengers travelling return between Tram Stop A and Stop B but half do not Tap-Off, then the dataset would record 50 pairs of A-B, 50 pairs of B-A, 50 pairs of A-UNKNOWN, and 50-pairs of B-UNKNOWN. If the analysis were to count the number of tram trips by summing all the trips between pairs of known tram stops (A-B, B-A) that would be 100 trips. However, if the analysis were to sum all the embarkations at tram stops (A, B) that would be 200 trips.

Figure 118: Sequence of ticketing events if Opal users do or do not Tap-On/Off

All tickets in this scenario are virtual-tickets stored on a smart-card ticket-device from (Tap-On) at a stop to (Tap-Off) another stop.

Scenario	Enter Station (Stop A)	Exit Station (Stop B)	Board Bus (Stop C)	Alight Bus (Stop D)
Usual Opal	Tap at ticket reader • reader detects no open ticket on device ∴ Tap-On • reader opens new ticket on device; from stop A	Tap at ticket reader • reader detects open ticket ∴ Tap-Off • reader closes ticket; to Stop B • reader deducts fare from device balance based on distance • reader uploads travel record for A-B	Tap at ticket reader • reader detects no open ticket on device ∴ Tap-On • reader opens new ticket on device, from Stop C	Tap at ticket reader • detect Tap-Off, reader closes ticket; to Stop D • reader deducts fare from device balance • reader uploads travel record for C-D
No Tap Off	Tap at ticket reader • reader detects no open ticket on device • reader opens new ticket on device, from Stop A	*If holder can exit without tapping, the virtual ticket remains open but is no longer a valid authority to travel*	Tap at ticket reader • reader detects open ticket • reader closes ticket, to stop UNKNOWN • reader deducts **maximum** fare from device balance • reader opens new ticket on ticket device; from Stop C • reader uploads travel record for A-UNKNOWN.	*Virtual ticket remains open but is no longer a valid authority to travel*
No Tap On	*Card holder has no authority to travel*	Tap at an exit gate ticket reader • reader detects no open ticket on device • reader deducts **maximum** fare from device balance • reader uploads travel record for UNKNOWN-B	*Card holder has no authority to travel.*	Tap at open ticket reader (i.e. could be an entry/exit) • reader detects no open ticket on device • reader opens new ticket on device, from Stop D

There is a risk when undertaking a TOTOR analysis that changing how a question is framed (how it is coded into the software) will generate different answers. Therefore, some answers may be inaccurate because the question was framed too narrowly and has excluded data affected by behavioural, environmental, or implementation artefacts.

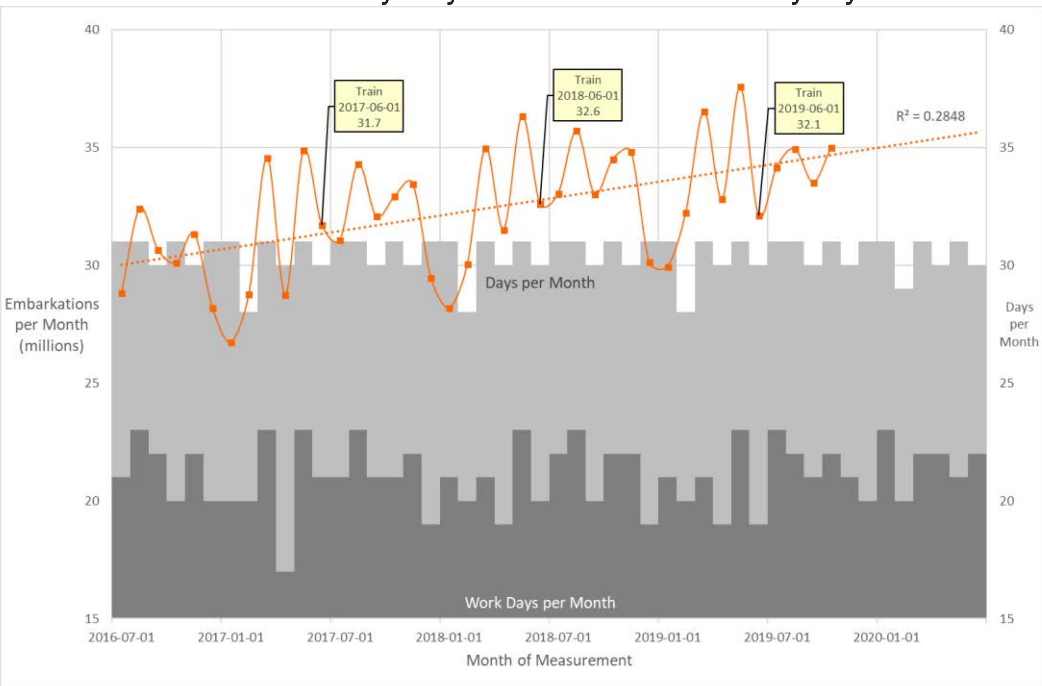
Remember, that machines break, and things can go wrong -- sometimes a person does not Tap-On or Tap-Off because the ticket readers are broken, or there is a blackout, or a severe hailstorm. And sometimes there are errors introduced by bugs in the software.

There are many subtle artefacts that exist within the Opal dataset and some of those may carry through to the primary dataset. It is the task of the creator of the primary dataset to compensate for problems and clean the dataset as best they can.

C.3 Measurements by the month

Without a doubt, the Gregorian month is a terrible unit of measurement and should not be used for datasets. Consider Figure 119 below which shows the number of embarkations recorded per month compared to the number of days per month, and the number of workdays per month (which excludes weekends and public holidays). The chart shows that the number of workdays days varies significantly from month-to-month ranging between 17 and 23 days with a standard deviation of 1.44 days.

Figure 119: Embarkations on Sydney & NSW trains versus days by month



Source: • (TfNSW 2019b)

Consider the case of November, this month was chosen as an indicative month during discussions with TfNSW as it has no public or school holidays and warm weather.

Table 48: Number of Each Day Type in November per Year

November	2016	2017	2018	2019	2020
Days of Month	30	30	30	30	30
Workdays	22	22	22	21	21
Monday	4	4	4	4	5
Tuesday	5	4	4	4	4
Wednesday	5	5	4	4	4
Thursday	4	5	5	4	4
Friday	4	4	5	5	4
Saturday	4	4	4	5	4
Sunday	4	4	4	4	5

In Table 48 above the number of each type of day for November is listed for the years from 2016 to 2020. Although every year the month has 30 days[35], in 2019 & 2020 it only has 21 workdays while in every other year it had 22 workdays. That is because in 2019 it had five Saturdays[36], which happens only one year out of every seven.

The reason the number of workdays is so significant to understanding public transport patronage is shown in Table 49 below. Table 49 shows that patronage on NSW trains for an average Saturday has less than 52% the patronage of an average workday, while the average Sunday does not even have 41% of an average workday's patronage. The average Saturday represents 9% of the week's railway patronage versus 7% on Sunday. On the buses, Saturdays is 6% of the week's patronage and Sundays are almost 5%.[37]

Table 49: Number of Passengers on Trains per Day

	Tuesday	Wednesday	Thursday	Saturday	Sunday
	9 Aug 2016	10 Aug	11 Aug	13 Aug	14 Aug
Train (thousands)	1,257	1,258	1,280	644	551
	22 Nov 2016	23 Nov	24 Nov	26 Nov	27 Nov
Train (thousands)	1,232	1,239	1,295	654	503
Average			(TWT) 1,260	649	527
Ratio to Average Workday TWT			100%	**51.5%**	**41.8%**

Source: • (TfNSW 2017a) • (TfNSW 2017b)

To convey this point further consider this rough heuristic. Assume that each workday equals 1 unit of train (or bus) patronage. As such, we have observed that an average Saturday is 0.515 units, and a Sundays is 0.418 units of workday patronage. Then from

[35] Months having 28, 30, or 31 days adds a degree of complexity to analytics using months

[36] 2020 is a leap year so the extra day in February changes the cycle.

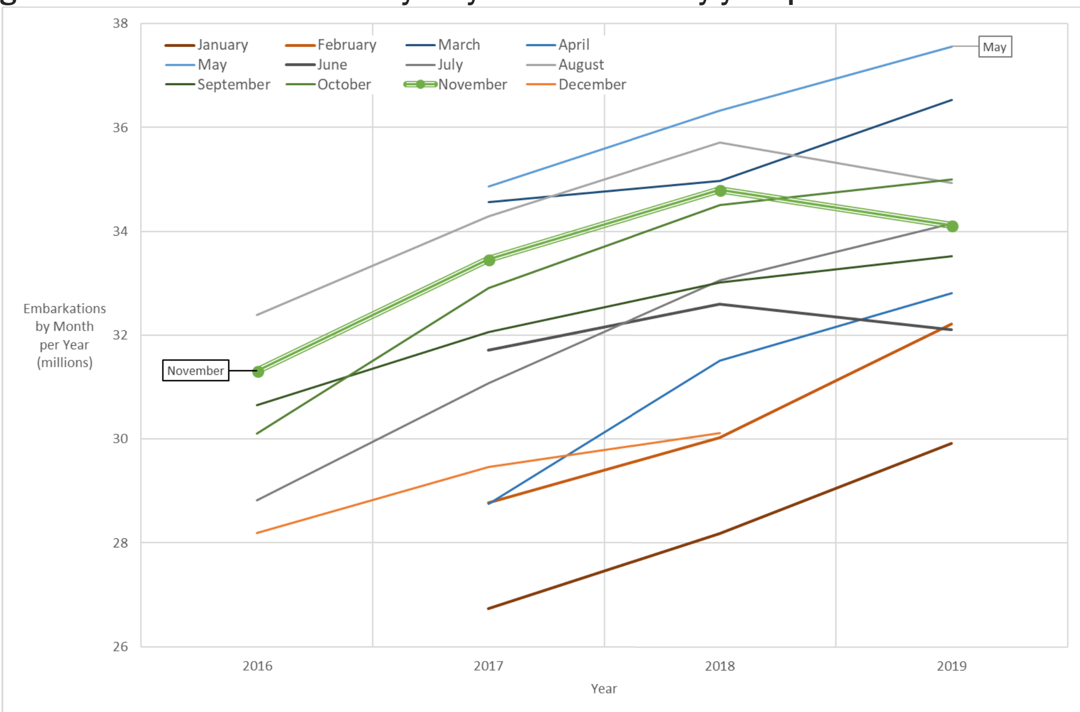
[37] Caveat: Humans like three-day weekends; so, Monday and Friday patronage can vary compared to the plain workdays of the midweek – Tuesday, Wednesday & Thursday (TWT). That is why the convention in NSW is to use only Tuesday, Wednesday, and Thursday to calculate the average workday patronage.

2016 to 2018, November would have 25.732 units of patronage but less in 2019 and 2020, with November patronage dropping to 25.15 workday units. Therefore, even if workday and weekend patronage were exactly the same then 2019 would record a 'drop' in November patronage using a naive monthly analysis.

Units of Patronage Heuristic	2016	2017	2018	2019	2020
Train	25.732	25.732	25.732	25.247	25.15
Bus	34.856	34.856	34.856	33.856	34.34

Figure 120 below was created to further highlight this point. The thick green line with markers is November, which suffers a significant year-on-year reduction in the monthly patronage in 2019 despite other months growing and the trend of overall growth in 2019. The patronage change is just a side effect of reporting by the Gregorian month.

Figure 120: Embarkations on Sydney & NSW trains by year per month



Source: • (TfNSW 2019b)

A possible reason for the prevalence of monthly datasets, despite their known disutility, is that most software has inbuilt support for manipulating date-stamps as a tuple, making it convenient to filter and aggregate records into years and months. However, months are not the only choice; e.g. the ACT releases patronage data by the week.

Reporting data in Gregorian months is ill-advised as it severely complicates analytics. However, a monthly aggregate is better than no data. Therefore, if a monthly dataset is the only realistic choice, the monthly dataset would be acceptable.

Appendix D – Additional Information

D.1 Top-100 Locations for Intermodal Continuation Embarkations

Table 50: Top-102 Locations for Transfer Embarkations on Wed 2016-11-16

Locations of Intermodal Transfers	Number of Segments	Percentage of Continuations
Central Station	42,270	18.1%
Bondi Junction Station	20,690	8.8%
Wynyard Station	15,440	6.6%
Parramatta Station	12,530	5.4%
Circular Quay Station	11,340	4.9%
Town Hall Station	10,400	4.4%
Blacktown Station	6,780	2.9%
Strathfield Station	6,310	2.7%
Hurstville Station	4,330	1.9%
Burwood Station	4,230	1.8%
Chatswood Station	4,000	1.7%
Liverpool Station	3,960	1.7%
North Sydney Station	3,150	1.3%
Edgecliff Station	3,010	1.3%
Rockdale Station	2,930	1.3%
Redfern Station	2,680	1.1%
Mount Druitt Station	2,660	1.1%
Fairfield Station	2,520	1.1%
Lidcombe Station	2,470	1.1%
Bankstown Station	2,460	1.1%
Epping Station	2,360	1.0%
Museum Station	2,340	1.0%
Green Square Station	2,200	0.9%
Campbelltown Station	2,010	0.9%
St Leonards Station	1,820	0.8%
Kogarah Station	1,730	0.7%
Penrith Station	1,630	0.7%
Martin Place Station	1,600	0.7%
Eastwood Station	1,600	0.7%
Gordon Station	1,570	0.7%
Macquarie University Station	1,550	0.7%
Hornsby Station	1,440	0.6%
Gosford Station	1,400	0.6%
Newtown Station	1,350	0.6%
Westmead Station	1,310	0.6%
Cabramatta Station	1,250	0.5%
St Marys Station	1,230	0.5%
Rhodes Station	1,190	0.5%
Ashfield Station	1,190	0.5%
Campsie Station	1,140	0.5%

Woy Woy Station	1,100	0.5%
Seven Hills Station	1,090	0.5%
Auburn Station	1,060	0.5%
Padstow Station	1,040	0.4%
Pennant Hills Station	1,030	0.4%
Mascot Station	1,030	0.4%
Turramurra Station	1,000	0.4%
West Ryde Station	990	0.4%
Sutherland Station (#50)	990	0.4%
Granville Station	970	0.4%
Glenfield Station	790	0.3%
Merrylands Station	750	0.3%
Dulwich Hill Station (#54)	740	0.3%
Sydenham Station	720	0.3%
Manly Wharf	650	0.3%
Taronga Zoo Wharf	620	0.3%
St James Station	550	0.2%
Ingleburn Station	520	0.2%
Macquarie Park Station	510	0.2%
Riverwood Station	500	0.2%
Minto Station	470	0.2%
Holsworthy Station	450	0.2%
Capitol Square Light Rail (#63)	430	0.2%
Lindfield Station	410	0.2%
Wyong Station	400	0.2%
Katoomba Station	400	0.2%
Westfield Hurstville	350	0.2%
Marrickville Station	320	0.1%
Revesby Station	300	0.1%
Rooty Hill Station	290	0.1%
Petersham Station	290	0.1%
Cremorne Point Wharf	290	0.1%
Miranda Station	280	0.1%
Mosman Bay Wharf	270	0.1%
Tuggerah Station	250	0.1%
Mortdale Station	250	0.1%
Quakers Hill Station	240	0.1%
Jannali Station	240	0.1%
Kings Cross Station	210	0.1%
Berowra Station	210	0.1%
Leumeah Station	200	0.1%
Kingsgrove Station	200	0.1%
Lakemba Station	180	0.1%
Milsons Point Station	180	0.1%
Balmain East Wharf	180	0.1%
Schofields Station	170	0.1%
Canley Vale Station	170	0.1%
Windsor Station	160	0.1%
Leppington Station	160	0.1%

Wollongong Station	150	0.1%
St Peters Station	150	0.1%
Cronulla Station	150	0.1%
Hamilton Station	150	0.1%
Engadine Station	150	0.1%
Carlingford Station	150	0.1%
Punchbowl Station	140	0.1%
Meadowbank Station	140	0.1%
Caringbah Station	140	0.1%
Canterbury Station	130	0.1%
Marion Light Rail (#100)	120	0.1%
Guildford Station	120	0.1%
Beecroft Station	120	0.1%

Note: Table 50 above was correlated into geographic clusters using simple association based on names, with some manual assistance. As such, about 1,000-1,500 (0.2-0.3%) of intermodal continuations were not associated to a geographic area.

102 locations were included in the above table to note the position of Marion Light Rail which should be an interchange location between the IWLR and western buses. On inbound running, there is a bus stop both 300m and 100m before the light rail station; as well as stops 200m and 400m after the station. During the community consultation the project team revealed that a bus stop at the light rail station was ruled out by the RTA. It should also be remembered that at the moment the fully grade separated IWLR has a slower journey time to the CBD than the Parramatta Rd buses running in mixed traffic.

D.2 Bus Priority Infrastructure Program (BPIP)

NSW Roads and Maritime Services describes the location of works undertaken to improve reliability on Sydney's strategic bus corridors as follows:

> 'The NSW Government released its roadmap to deliver the state's future transport network, the Future Transport Strategy 2056, in 2018. The Bus Priority Infrastructure Program supports this vision by delivering infrastructure that make bus services faster and more reliable, such as bus lanes, bus priority at intersections or more efficient bus stop placement.
>
> ### Bus Priority Infrastructure Program Projects
>
> **Projects with changes to bus stops**
> - Arlington Street, Croydon Road and Parramatta Road intersection improvements (routes 490 and 492)
> - Woniora Road and King Georges Road, South Hurstville (routes 970/971)
> - Princes Highway, Blakehurst and Sylvania (routes 970/971)
> - Menai Road, Princes Highway, Kirrawee and the Kingsway, Miranda (routes M92/962)
> - Gibson Avenue and Ryan Road, Padstow (routes M92/962)
> - Gibbons Street, Redfern bus zone extensions
> - Parramatta to Macquarie Park via Epping (route M54)
> - Terrey Hills, Belrose and Davidson to City (routes 270/271)
> - Chatswood to Manly (route 136)
> - Botany to Gore Hill (route M20)
> - Bondi Junction to Chatswood (route M40)
> - Peakhurst, Penshurst and Hurstville (route M91)
> - Parramatta, Granville and South Granville (route M91)
> - Guildford, Chester Hill, Sefton, Bass Hill and Yagoona (route M91)
> - Padstow (route M91)
> - Blacktown to Hornsby via Castle Hill (routes M60, T70 and T71)
> - Castle Hill and Baulkham Hills (route M61)
> - Macquarie Park to Hurstville (route M41)
> - Canterbury, Earlwood and Kingsgrove (routes 412, 423 and 428)
> - Camperdown, Stanmore, Dulwich Hill and Marrickville (routes 412, 423, 426 and 428)
> - Arncliffe, Kogarah, Rockdale and Wolli Creek (route 422)
> - Vaucluse to Bondi (route 389)
> - Rozelle to Annandale (route 440)
> - Darlinghurst to Bondi (routes 333/380)
>
> **Other bus priority projects**
> - Bus priority improvements on Smith Street and Wilde Avenue, Parramatta
> - Intersection upgrade at Showground road and Carrington Road, Castle Hill
> - Macquarie Park Bus priority and Capacity Improvement[38]

[38] Note that the "Macquarie Park Bus priority and Capacity Improvement" was initiated to support the closure of the Epping to Chatswood Rail Line so that could be converted to Sydney Metro.

D.3 The rates of crimes in NSW

Table 51: Crime recorded by the NSW Police Force: rates^ and trends^^

Offence type	Rate per 100,000 population FY 2018-2019	5-year trend and average annual percent change (Jul 2014-2019)
Transport regulatory offences	1,390.4	Stable
Malicious damage to property	732.2	Down 3.2%
Fraud	649.5	Stable
Breach bail conditions	567.4	Up 6.2%
Steal from motor vehicle	484.2	Down 2.6%
Intimidation, stalking and harassment	424.8	Up 2.7%
Non-domestic violence related assault	402.1	Stable
Domestic violence related assault	387.1	Stable
Other theft	344.4	Down 3.4%
Steal from retail store	331.1	Up 5.0%
Break and enter dwelling	324.5	Down 6.1%
Steal from dwelling	239.0	Down 4.3%
Possession and/or use of cannabis	219.1	Down 1.5%
Breach Apprehended Violence Order	212.9	Up 5.1%
Prohibited and regulated weapons offences	190.5	Up 6.4%
Other offences	186.0	Stable
Motor vehicle theft	168.5	Down 3.2%
Liquor offences	151.0	Stable
Trespass	132.7	Up 3.4%
Break and enter non-dwelling	127.1	Down 6.2%
Indecent assault, act of indecency and other sexual offences	103.1	Up 5.5%
Receiving or handling stolen goods	103.0	Stable
Possession and/or use of amphetamines	95.1	Up 4.6%
Other drug offences	78.4	Up 5.1%
Resist or hinder officer	76.9	Stable
Sexual assault	73.5	Up 4.5%
Possession and/or use of other drugs	69.2	Up 11.8%
Arson	64.5	Stable
Offensive conduct	57.4	Down 7.2%
Steal from person	49.5	Down 10.0%
Possession and/or use of ecstasy	36.9	Up 8.6%
Assault Police	31.6	Stable
Offensive language	31.4	Down 12.5%
Possession and/or use of cocaine	29.9	Up 20.5%
Criminal intent	29.7	Stable
Dealing, trafficking in amphetamines	22.7	Down 6.9%
Robbery without a weapon	19.0	Down 5.1%
Cultivating cannabis	16.0	Stable
Other offences against the person	14.9	Down 2.3%

Dealing, trafficking in cannabis	14.6	Stable
Other offences against justice procedures	13.5	Up 20.5%
Possession and/or use of narcotics	13.1	Up 5.2%
Dealing, trafficking in narcotics	13.0	Stable
Robbery with a weapon not a firearm	11.6	Down 4.2%
Dealing, trafficking in cocaine	10.2	Up 28.7%
Pornography offences	8.1	Up 6.4%
Dealing, trafficking in ecstasy	6.9	Stable
Dealing, trafficking in other drugs	6.6	Up 16.2%
Fail to appear	6.3	Down 9.9%
Stock theft	5.6	Down 3.2%
Abduction and kidnapping	2.6	Down 5.8%
Escape custody	2.3	Stable
Betting and gaming offences	2.1	Stable
Robbery with a firearm	2.0	Down 12.6%
Blackmail and extortion	1.2	Stable
Murder *	1.0	Stable
Importing drugs	1.0	Up 20.5%
Manufacture drug	0.8	Down 13.4%
Prostitution offences	0.3	Down 44.1%
Attempted murder	0.3	nc
Manslaughter *	0.1	nc
Murder accessory, conspiracy	0.0	nc

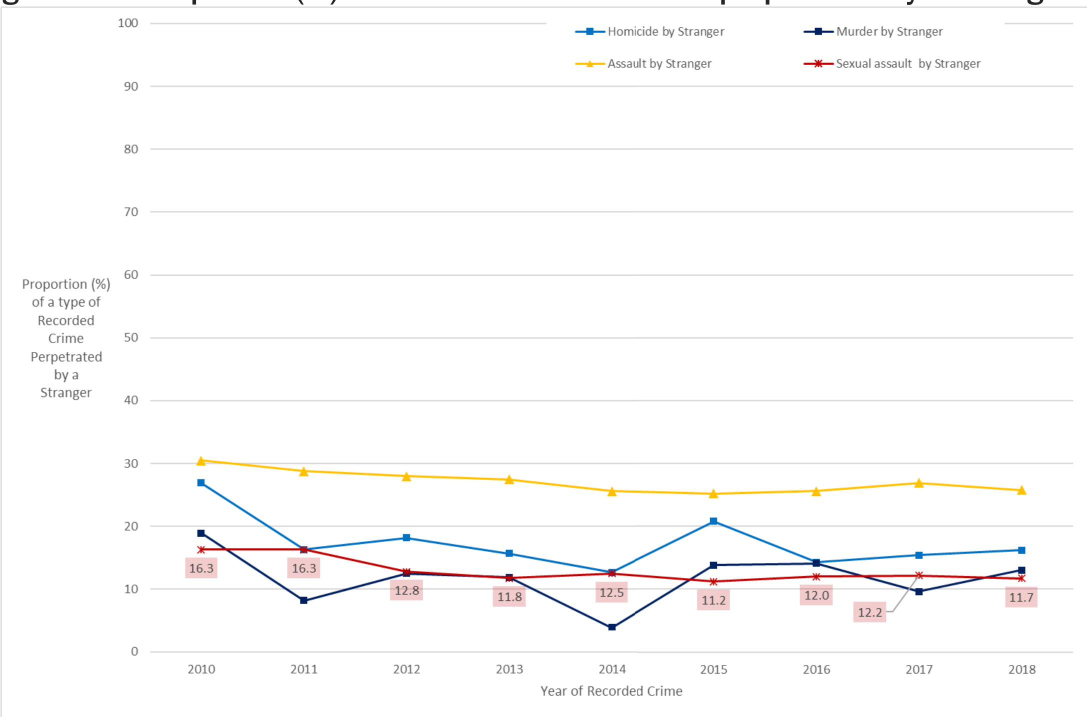
Source: NSW Bureau of Crime Statistics and Research – NSW_trends19Q2; (BCSR 2019)

Note *the* source contained no footnote to explain the asterisks (*) or single caret (^).

For the rate calculations, specialised population data were prepared and provided to BOCSAR by the Australian Bureau of Statistics.

^^ Shows the results of a statistical test for a significant upward or downward trend in the monthly number of criminal incidents recorded. Where the trend is significant (i.e. p<0.05) the average annual percentage change over the five and two-year period is shown. Significant upward trends are highlighted in red; significant downward trends are highlighted in yellow. ‘Stable’ indicates there was no significant upward or downward trend and ‘nc’ indicates that the number of incidents recorded was too small for a reliable trend test to be performed.

Figure 121: Proportion (%) of Recorded Crime in NSW perpetrated by a Stranger



Source: • (ABS 2019) -- 45100DO003_2018 Recorded Crime – Victims, Australia, 2018 -- Table 11 VICTIMS, Relationship of offender to victim by offence, Selected states and territories, 2018

Appendix E - Reference Texts

Engineers Australia General Regulations

These extracts are from (EA 2018a, pp. 88-91) and (EA 2018b). They are provided to show a clear example of how an institutional body of engineers have identified the need for practitioners to protect the community during the execution of their duties. Note that judgement was consistently mis-spelt as judgment [sic] in the original text.

The text highlighted in yellow reflect that engineering is Normative and based on evaluations and thus values. The blue highlight is for text that acknowledge rights and diversity. Whereas text focusing on the need for an environmentally sustainable approach are highlighted in green.

Schedule 1 Code of Ethics and Guidelines

Part 1 – Code of Ethics (sub-regulation 2.20(1))

As engineering practitioners, we use our knowledge and skills for the benefit of the community to create engineering solutions for a sustainable future. In doing so, we strive to serve the community ahead of other personal or sectional interests.

Our Code of Ethics defines the values and principles that shape the decisions we make in engineering practice. The related Guidelines on Professional Conduct provide a framework for members of Engineers Australia to use when exercising their [judgement] in the practice of engineering. …

In the course of engineering practice we will:

1. Demonstrate integrity
    - 1.1 act on the basis of a well-informed conscience
    - 1.2 be honest and trustworthy
    - 1.3 respect the dignity of all persons
2. Practise competently
    - 2.1 maintain and develop knowledge and skills
    - 2.2 represent areas of competence objectively
    - 2.3 act on the basis of adequate knowledge
3. Exercise leadership
    - 3.1 uphold the reputation and trustworthiness of the practice of engineering
    - 3.2 support and encourage diversity
    - 3.3 communicate honestly and effectively, taking into account the reliance of others on engineering expertise

4. Promote sustainability

4.1 engage responsibly with the community and other stakeholders

4.2 practise engineering to foster the health, safety and wellbeing of the community and the environment

4.3 balance the needs of the present with the needs of future generations.

Part 2 – Guidelines on Professional Conduct (sub-regulation 2.20(2))

These Guidelines on Professional Conduct provide a framework for members of Engineers Australia to use when exercising their [judgement] in the practice of engineering.

The Guidelines are not intended to be, nor should they be interpreted as, a full or exhaustive list of the situations and circumstances which may comprise compliance and non-compliance with the Code of Ethics. If called upon to do so, members are expected to justify any departure from either the provisions or spirit of the Code.

Ethical engineering practice requires [judgement], interpretation and balanced decision-making in context.

Engineers Australia recognises that, while our ethical values and principles are enduring, standards of acceptable conduct are not permanently fixed. Community standards and the requirements and aspirations of engineering practice will develop and change over time. Within limits, what constitutes acceptable conduct may also depend on the nature of individual circumstances.

…

1. Demonstrate integrity

1.1 Act on the basis of a well-informed conscience

(a) be discerning and do what you think is right

(b) act impartially and objectively[39]

(c) act appropriately, and in a professional manner, when you perceive something to be wrong

(d) give due weight to all legal, contractual and employment obligations

1.2 Be honest and trustworthy

(a) accept, as well as give, honest and fair criticism

(b) be prepared to explain your work and reasoning[40]

(c) give proper credit to those to whom proper credit is due

(d) in managing perceived conflicts of interest, ensure that those conflicts are disclosed to relevant parties

(e) respect confidentiality obligations, express or implied

(f) do not engage in fraudulent, corrupt, or criminal conduct

1.3 Respect the dignity of all persons

(a) treat others with courtesy and without discrimination or harassment

(b) apply knowledge and skills without bias in respect of race, religion, gender, age, sexual orientation, marital or family status, national origin, or mental or physical handicaps

[39] As discussed in previous sections it is impossible for a human to act fully objectively.

[40] Honesty and peer-review are mechanism to achieve transparency.

2 Practise competently

2.1 Maintain and develop knowledge and skills

(a) continue to develop relevant knowledge and expertise

(b) act in a careful and diligent manner

(c) seek peer review

(d) support the ongoing development of others

2.2 Represent areas of competence objectively

(a) practise within areas of competence

(b) neither falsify nor misrepresent qualifications, grades of membership, experience or prior responsibilities

2.3 Act on the basis of adequate knowledge

(a) practise in accordance with legal and statutory requirements, and with the commonly accepted standards of the day

(b) inform employers or clients if a task requires qualifications and experience outside your areas of competence

3 Exercise leadership

3.1 Uphold the reputation and trustworthiness of the practice of engineering

(a) advocate and support the extension of ethical practice

(b) engage responsibly in public debate and deliberation

3.2 Support and encourage diversity

(a) select, and provide opportunities for, all engineering practitioners on the basis of merit

(b) promote diversity in engineering leadership

3.3 Communicate honestly and effectively, taking into account the reliance of others on engineering expertise

(a) provide clear and timely communications on issues such as engineering services, costs, outcomes and risks

4 Promote sustainability

4.1 Engage responsibly with the community and other stakeholders

(a) be sensitive to public concerns

(b) inform employers or clients of the likely consequences of proposed activities on the community and the environment

(c) promote the involvement of all stakeholders and the community in decisions and processes that may impact upon them and the environment

4.2 Practise engineering to foster the health, safety and wellbeing of the community and the environment

(a) incorporate social, cultural, health, safety, environmental and economic considerations into the engineering task

4.3 Balance the needs of the present with the needs of future generations

(a) in identifying sustainable outcomes consider all options in terms of their economic, environmental and social consequences

(b) aim to deliver outcomes that do not compromise the ability of future life to enjoy the same or better environment, health, wellbeing and safety as currently enjoyed.in the practice of engineering.

Bibliography

Abbott, T. 2009, *Battlelines*, Melbourne University Publishing, Carlton, Vic.

ABS, A.B.o.S. 2016, *Australian Statistical Geography Standard (ASGS)*, Canberra, Australian Capital Territory, Australia.

ABS, A.B.o.S. 2019, '4510.0 - Recorded Crime - Victims, Australia, 2018', Series 4510.0 - Recorded Crime - Victims, Australia, 2018 vol. Commonwealth of Australia, Canberra, Australian Capital Territory, Australia, <https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4510.0Main+Features12018?OpenDocument>.

Anderson, S. 2017, *Coercion*, Metaphysics Research Lab, Stanford University, viewed 2019-08-07 2019, <https://plato.stanford.edu/entries/coercion/>.

ARLC, A.L.R.C. 2010, *State and territory regulation of privacy*, Commonwealth of Australia, viewed 2019-11-30, <https://www.alrc.gov.au/publication/for-your-information-australian-privacy-law-and-practice-alrc-report-108/2-privacy-regulation-in-australia/state-and-territory-regulation-of-privacy/>.

AS, S.A. & NZS, S.N.Z. 2004, *AS/NZS 4360:2004 - Risk management-Guidelines*, Standards Australia.

ATC, A.T.C. 2006, *National Guidelines for Transport Systems Management in Australia*, vol. 4 - Urban Transport, Commonwealth of Australia.

Auditor-General, N. 2019, *Transport Access Program - Performance Audit - New South Wales Auditor General's Report*, 2019-02-19 edn, Audit Office of New South Wales, Sydney, New South Wales, Australia, <https://www.audit.nsw.gov.au/our-work/reports/transport-access-program>.

Bamford, J. 2002, *Body of secrets : anatomy of the ultra-secret National Security Agency*, 1st Anchor Books edn, Anchor Books, New York.

Bartels, R. 1988, 'The history of marketing thought'.

BCSR, N.B.o.C.S.a.R. 2019, 'Trends for major offences', electronic data set, <https://www.bocsar.nsw.gov.au/Pages/bocsar_datasets/Datasets-.aspx>.

Bechara, A. & Damasio, A.R. 2005, 'The somatic marker hypothesis: A neural theory of economic decision', *Games and Economic Behavior*, vol. 52, no. 2, pp. 336-72.

Bell, K. 2015, 'From Squaresville to Triangle Town: Geometries for public transport network planning'.

BIPM, B.I.d.P.e.M. 2018, *Logo of the Intenational System of Units that was introiduced in May 2019.*, Bitmap, Wikimedia Commons, <https://en.wikipedia.org/wiki/File:International_System_of_Units_Logo.png>.

BITRE, B.o.I., Transport and Regional Economics 2016, *Australian Infrastructure Statistics — Yearbook 2016*, Commonwealth of Australia, Canberra, Australian Capital Territory, Australia, <https://www.bitre.gov.au/publications/2016/yearbook_2016.aspx>.

BITRE, B.o.I., Transport and Regional Economics 2018, *Australian Infrastructure Statistics — Yearbook 2018*, Commonwealth of Australia, Canberra, Australian Capital Territory, Australia, <https://www.bitre.gov.au/publications/2018/yearbook_2018.aspx>.

BITRE, B.o.I., Transport and Regional Economics 2019, *Australian Infrastructure Statistics — Yearbook 2019*, Commonwealth of Australia, Canberra, Australian Capital Territory, Australia, <https://www.bitre.gov.au/publications/2019/yearbook_2019>.

Blaug, M. 1980, *The methodology of economics : or, How economists explain*, Cambridge University Press, Cambridge ; New York.

Branson, C.M. 2008, 'Achieving organisational change through values alignment', *Journal of Educational Administration*, vol. 46, no. 3, pp. 376-95.

BTS, B.o.T.S. 2013, *2011/12 Household Travel Survey Summary Report 2013 Release*, State of NSW, Sydney, New South Wales, Australia.

BTS, B.o.T.S. 2014, *Household Travel Survey Report: Sydney 2012/13*, State of NSW, Sydney, New South Wales, Australia.

Bunge, M. 2003, *Emergence and Convergence : Qualitative Novelty and the Unity of Knowledge*, University of Toronto Press, Toronto, UNKNOWN.

Bunge, M. 2012, 'Wealth and Well-being, Economic Growth, and Integral Development', *International Journal of Health Services*, vol. 42, no. 1, pp. 65-76.

Bunge, M. 2016, *Between Two Worlds: Memoirs of a Philosopher-Scientist*, Springer International Publishing.
Canada-OPCC, O.o.t.P.C.o.C. 2019, *PIPEDA fair information principles*, Canada, viewed 2019-11-30, <https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/p_principle/>.
Carbery, G. 2010, 'Towards homosexual equality in Australian criminal law: a brief history', *Sydney: Australian Lesbian and Gay Archives Inc*.
Carrel, A., Halvorsen, A. & Walker, J.L. 2013, 'Passengers' Perception of and Behavioral Adaptation to Unreliability in Public Transportation', *Transportation Research Record*, vol. 2351, no. 1, pp. 153-62.
CEN, C.E.d.N.E.C.f.S. 2002, *European Standard EN 13816*.
Chao, H.-K., Chen, S.-T. & Millstein, R.L. 2013, 'Towards the Methodological Turn in the Philosophy of Science', in H.-K. Chao, S.-T. Chen & R.L. Millstein (eds), *Mechanism and Causality in Biology and Economics*, Springer Netherlands, Dordrecht, pp. 1-16.
Churchill, W. & Churchill, W.S. 1986, *Closing the ring*, vol. 5, Houghton Mifflin Harcourt.
CIA, C.I.A.U.S.o.A., *World Factbook*, <https://www.cia.gov/library/publications/the-world-factbook>.
Cialdini, R.B. 2007, *Influence : the psychology of persuasion*, Rev. edn, Morrow, New York.
CoA-OAIC, O.o.A.I.C. 2019, *Australian Privacy Principles quick reference*, Commonwealth of Australia, viewed 2019-11-30, <https://www.oaic.gov.au/privacy/australian-privacy-principles/australian-privacy-principles-quick-reference/>.
Cohon, R. 2018, *Hume's Moral Philosophy*, Metaphysics Research Lab, Stanford University, viewed 2019-06-04 2019.
Constantinides, E. 2006, 'The Marketing Mix Revisited: Towards the 21st Century Marketing', *Journal of Marketing Management*, vol. 22, no. 3-4, pp. 407-38.
Cook, B., David, F. & Grant, A. 2001, *Sexual violence in Australia*, Australian Institute of Criminology Canberra.
Council, W. 2016, *The Great Aussie Cossie: 1880s to today*, Waverly Council, <http://web.archive.org/web/20151217223416/http://www.waverley.nsw.gov.au/__data/assets/pdf_file/0008/8729/The_Great_Aussie_Cossie_1880s_to_today.pdf>.
Crayen, C.R.M. 2018, 'Mobility-as-a-Service: are current players delivering value?: a qualitative consumer analysis'.
Croome, R. 1995, 'Sexual (Mis) conduct', *Alternative LJ*, vol. 20, p. 282.
Crowe, J. & Stephenson, P. 2014, 'An Express Constitutional Right to Vote-The Case for Reviving Section 41', *Sydney L. Rev.*, vol. 36, p. 205.
Curtis, C.a.S., J. 2016, *Planning for public transport accessibility: an international sourcebook*, Routledge.
Daniels, R. & Mulley, C. 2013, 'Explaining walking distance to public transport: The dominance of public transport supply', *Journal of Transport and Land Use*, vol. 6, no. 2, pp. 5-20.
Devlin, H. 2019, ''We are hurtling towards a surveillance state': the rise of facial recognition technology ', *The Guardian (Surveillance)*, 2019-10-05.
Diallo, R. 2020, 'Coronavirus exposed the real reasons behind France's 'burqa ban'', *Al Jazeera*, 2020-05-15.
EA, E.A. 2018a, *Engineers Australia General Regulations* pamphlet, Canberra, ACT, Australia, <https://www.engineersaustralia.org.au/sites/default/files/resource-files/2018-10/EA%20General%20Regulations_0.pdf>.
EA, E.A. 2018b, *Our code of ethics*, pamphlet, Canberra, ACT, Australia, <https://engineersaustralia.org.au/ethics

http://eacorp.prod.acquia-sites.com/sites/default/files/resource-files/2018-10/EA%20Code%20of%20Ethics%20Final.pdf>.
ECJ, C.o.J.o.t.E.U. 2007, *The Treaty of Lisbon and the Court of Justice of the European Union - Organisation of the Court of Justice of the European Union*, European Union, <https://curia.europa.eu/jcms/jcms/P_69328/en/>.
ECtHR, E.C.o.H.R. 2010, 'European Convention on Human Rights (Convention for the Protection of Human Rights and Fundamental Freedoms)', Series European Convention on Human Rights (Convention for the Protection of Human Rights and Fundamental Freedoms) Council of Europe, <https://www.echr.coe.int/Documents/Convention_ENG.pdf

https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/005>.

Fercoq, A., Lamouri, S. & Carbone, V. 2016, 'Lean/Green integration focused on waste reduction techniques', *Journal of Cleaner Production*, vol. 137, pp. 567-78.
Fieser, J. 2019, *David Hume (1711—1776)*, University of Tennessee at Martin, U.S.A, viewed 2019-05-21 2019.
Gearin, M. 2019, *'Shocking' myki privacy breach for millions of users in data release*, ABC, Australian Broadcasting Corporation, viewed 2020-01-18 2020, <https://www.abc.net.au/news/2019-08-15/myki-data-spill-breaches-privacy-for-millions-of-users/11416616>.
GHD 2010, *Sydney Light Rail - Inner West Extension Study Draft Report*.
Gilbert, D. 2014, 'Britain 'Sleepwalked' into Becoming a Mass Surveillance State', *International Business Times*, 2014-10-17.
Giles, B. 2013a, *594 Opal Card [Reader] Bus Trial IMG 5392*, Photo, Wikimedia Commons, <https://commons.wikimedia.org/wiki/File:594_Opal_Card_Bus_Trial_IMG_5392.jpg
Author: https://www.flickr.com/photos/beaugiles/14620338518/in/album-72157631848123485/>.
Giles, B. 2013b, *Opal Card [with others] IMG_5889*, Flikr, <https://www.flickr.com/photos/beaugiles/8474774377/in/album-72157631848123485/>.
Glazebrook, D.G., Wilmot, K., Hounsell, M. & Bos, P.V.D. 2017, *Station Precinct Rhodes Planning Proposal 2 – Transport Capacity Assessment*, Institute for Sustainable Futures, Sydney, New South Wales, Australia.
Green, L. 2018, *Legal Positivism*, Metaphysics Research Lab, Stanford University, viewed 2019-11-15 2019, <https://plato.stanford.edu/entries/legal-positivism/>.
Greenleaf, G. 2007, 'Australia's proposed ID card: Still quacking like a duck', *Computer Law & Security Review*, vol. 23, no. 2, pp. 156-66.
Grönroos, C. 2006, 'Adopting a service logic for marketing', *Marketing theory*, vol. 6, no. 3, pp. 317-33.
Gummesson, E. 2008, 'Customer centricity: reality or a wild goose chase?', *European Business Review*, vol. 20, no. 4, pp. 315-30.
Halpern, H. 2009, 'Supervision and the Johari Window: a framework for asking questions', *Education for Primary Care*, vol. 20, no. 1, pp. 10-4.
Hasan, S., Schneider, C.M., Ukkusuri, S.V. & González, M.C. 2013, 'Spatiotemporal Patterns of Urban Human Mobility', *Journal of Statistical Physics*, vol. 151, no. 1, pp. 304-18.
Heawood, J. 2014, 'Writerly rights: free speech, privacy and H.G. Wells', *Critical Quarterly*, vol. 56, no. 4, pp. 71-82.
Heinze, J.E., Krusky-Morey, A., Vagi, K.J., Reischl, T.M., Franzen, S., Pruett, N.K., Cunningham, R.M. & Zimmerman, M.A. 2018, 'Busy Streets Theory: The Effects of Community-engaged Greening on Violence', *American journal of community psychology*, vol. 62, no. 1-2, pp. 101-9.
Himma, K.E. 2019a, *Legal Positivism*, University of Tennessee at Martin, U.S.A, viewed 2019-11-15 2019, <https://www.iep.utm.edu/legalpos/>.
Himma, K.E. 2019b, *Philosophy of Law*, University of Tennessee at Martin, U.S.A, viewed 2019-11-15 2019, <https://www.iep.utm.edu/legalpos/>.
Hines, P., Holweg, M. & Rich, N. 2004, 'Learning to evolve: a review of contemporary lean thinking', *International journal of operations & production management*, vol. 24, no. 10, pp. 994-1011.
Hounsell, M. 2014, 'Submission to SSI 13_6148 - Regarding the M4 Western Motorway - WestConnex M4 Widening', Series Submission to SSI 13_6148 - Regarding the M4 Western Motorway - WestConnex M4 Widening NoW Public Transport Inc, Sydney, New South Wales, Australia, 2014-09-30.
Hounsell, M. 2017a, 'Mass Transport Customer Characteristics – NSW GMA', Data Analysis Report, University of Technology Sydney, Ultimo, Sydney, NSW.
Hounsell, M. 2017b, 'Resolving transport and land-use questions with data driven enquiry', Short Course, University of Technology Sydney, Ultimo, NSW, Australia.
Hounsell, M. 2017c, 'Stage 1 to 2: How can TOTOR systems be used to improve customer experience on mass transport services; using the Inner West Light Rail as a case study?', Stage 1 to 2 Transition, University of Technology Sydney, Ultimo, NSW, Australia.
Hounsell, M. 2018a, 'Better service through runtime savings – Inner West Light Rail case study', paper presented to the *Australasian Transport Research Forum*, Darwin, <https://atrf.info/papers/2018/files/ATRF2018_paper_61.pdf>.

Hounsell, M. 2018b, 'Implications of running time variability for passengers and operators–a case study using Bondi’s 333 buses', paper presented to the *Australasian Transport Research Forum (ATRF)*, Darwin, <https://atrf.info/papers/2018/files/ATRF2018_paper_60.pdf>.
Hounsell, M. 2018c, 'Relationship between transport & landuse system components and feedback processes on urban transport networks', Short Course, University of Technology Sydney, Ultimo, NSW, Australia.
Hounsell, M. 2018d, 'Stage 2 to 3: How can the lessons learnt from a multidisciplinary andragogic programme to improve public transport operations and governance?', Stage 2 to 3 Transition, University of Technology Sydney, Ultimo, NSW, Australia.
Howard, J. 2018, *Cognitive Errors and Diagnostic Mistakes: A Case-Based Guide to Critical Thinking in Medicine*, Springer International Publishing.
Hymon, S. 2020, 'Frequent service plan released for NextGen Bus Plan', *The Source*, weblog, Metro - Los Angeles County Metropolitan Transportation Authority, <https://thesource.metro.net/2020/01/10/frequent-service-plan-released-for-nextgen-bus-plan/>.
India, R.o. & Ambedkar, B.R. 1949, 'The Constitution of India', Series The Constitution of India Republic of India, <https://www.india.gov.in/my-government/constitution-india/constitution-india-full-text>.
IPART, I.P.a.R.T.o.N.S.W. 2019, *Maximum Opal Fares 2020-2024 - Draft Report*, State of New South Wales, Sydney, New South Wales, <https://www.ipart.nsw.gov.au/files/sharedassets/website/shared-files/pricing-reviews-transport-services-publications-opal-public-transport-services-to-june-2024/draft-report-maximum-opal-fares-2020-2024-december-2019.pdf>.
ISO, I.S.O. & AS, S.A. 2018, *AS ISO 31000:2018 - Risk management-Guidelines*, Standards Australia.
Jittrapirom, P., Caiati, V., Feneri, A.-M., Ebrahimigharehbaghi, S., Alonso González, M.J. & Narayan, J. 2017, 'Mobility as a service: A critical review of definitions, assessments of schemes, and key challenges'.
Jones, T.H. 1994, 'Legal protection for fundamental rights and freedoms: European lessons for Australia?', *Federal Law Review*, vol. 22, no. 1, pp. 57-91.
Keoghan, S. & Mcdonough, K. 2017, 'Women choosing to wear skimpy bikinis or nothing at all risk heavy fines but people aren’t offended anymore', *The Sunday Telegraph*.
Kitamura, R., Fujii, S. & Pas, E.I. 1997, 'Time-use data, analysis and modeling: toward the next generation of transportation planning methodologies', *Transport Policy*, vol. 4, no. 4, pp. 225-35.
Kittelson, Associates, Administration, U.S.F.T., Program, T.C.R. & Corporation, T.D. 2003, *Transit capacity and quality of service manual*, Second edn, Transportation Research Board.
Knaus, C. 2016, 'Steep rise in police requests for Opal data to track people on NSW', *The Guardian (Australia)*, 2016-12-02.
Kyodo, J. 2019, 'Japan kicks off Reiwa Era's first energy-saving Cool Biz casual wear campaign', *Japan Times*, 2019-05-07.
LAPD, L.A.P.D. 2019, *Design Out Crime - Los Angeles Police Department*, City of Los Angeles, State of California, United States of America, <http://www.lapdonline.org/crime_prevention/content_basic_view/8852>.
Legara, E.F., Monterola, C., Lee, K.K. & Hung, G.G. 2014, 'Critical capacity, travel time delays and travel time distribution of rapid mass transit systems', *Physica A: Statistical Mechanics and its Applications*, vol. 406, pp. 100-6.
Levitt, T. 1984, 'Marketing myopia', *Journal of Library Administration*, vol. 4, no. 4, pp. 59-80.
Li, Y. & Voege, T. 2017, 'Mobility as a service (MaaS): Challenges of implementation and policy required', *Journal of transportation technologies*, vol. 7, no. 02, pp. 95-106.
Liedtka, J. 2015, 'Perspective: Linking Design Thinking with Innovation Outcomes through Cognitive Bias Reduction', *Journal of Product Innovation Management*, vol. 32, no. 6, pp. 925-38.
Llamas, J. 2016, 'Houston's Groundbreaking New Transit System', *Strong Towns*, weblog, <https://www.strongtowns.org/journal/2016/1/25/texas-transportation-funding-its-not-all-bad-news>.
Logan, D.C. 2009, 'Known knowns, known unknowns, unknown unknowns and the propagation of scientific enquiry', *Journal of Experimental Botany*, vol. 60, no. 3, pp. 712-4.
Long, Y. & Thill, J.-C. 2015, 'Combining smart card data and household travel survey to analyze jobs–housing relationships in Beijing', *Computers, Environment and Urban Systems*, vol. 53, pp. 19-35.
Luft, J. & Ingham, H. 1961, 'The Johari Window', *Human relations training news*, vol. 5, no. 1, pp. 6-7.

Lusch, R.F. & Vargo, S.L. 2014, 'Evolving to a new dominant logic for marketing', *The Service-Dominant Logic of Marketing*, Routledge, pp. 21-46.
Ma, X., Wu, Y.-J., Wang, Y., Chen, F. & Liu, J. 2013, 'Mining smart card data for transit riders' travel patterns', *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 1-12.
MacAskill, E. 2016, ''Extreme surveillance' becomes UK law with barely a whimper', *The Guardian (Surveillance)*, 2016-11-19.
Marr, D. 1999, *The high price of heaven*, Allen & Unwin, St Leonards, N.S.W.
Martin, L.A. & Thornton, S. 2017, 'Can Road Charges Alleviate Congestion?', *Available at SSRN 3055428*.
Maslow, A.H. 1943, 'A theory of human motivation', *Psychological review*, vol. 50, no. 4, p. 370.
Maybee, J.E. 2016, *Hegel's Dialectics*, Metaphysics Research Lab, Stanford University, viewed 2019-07-15 2019, <https://plato.stanford.edu/entries/hegel-dialectics/>.
McNally, M.G. & Rindt, C.R. 2007, 'The Activity-Based Approach', in R. Rindt Craig, A.H. David & J.B. Kenneth (eds), *Handbook of Transport Modelling*, vol. 1, Emerald Group Publishing Limited, pp. 55-73.
MDRX 2019, *SydneyMetro39 [Opal Card Reader Gates]*, Photo, Wikimedia Commons, <https://commons.wikimedia.org/wiki/File:SydneyMetro39.jpg>.
Mees, P. 2015, *A very public solution*, Melbourne Univ. Publishing.
Metcalfe, R. & Dolan, P. 2012, 'Behavioural economics and its implications for transport', *Journal of Transport Geography*, vol. 24, pp. 503-11.
Mill, J.S. 1859, *On Liberty*, Batoche Books.
Mishevska, E., Najdenov, B., Jovanovik, M. & Trajanov, D. 2014, *Open Public Transport Data in Macedonia*.
Moonan, L. 1975, 'Hume on is and ought', *Journal of the History of Philosophy*, vol. 13, no. 1, pp. 83-98.
Morris, W.E. & Brown, C.R. 2019, *David Hume*, Metaphysics Research Lab, Stanford University, viewed 2019-0522 2019.
Munizaga, M.A. & Palma, C. 2012, 'Estimation of a disaggregate multimodal public transport Origin–Destination matrix from passive smartcard data from Santiago, Chile', *Transportation Research Part C: Emerging Technologies*, vol. 24, pp. 9-18.
Murdoch, N.J. 2017, 'Predictions of Level of Service Improvements Resulting From Bus Stop Removal', Capstone thesis, University of Technology Sydney, Ultimo, New South Wales, Australia.
Murphy, M. 2019, *The Natural Law Tradition in Ethics*, Metaphysics Research Lab, Stanford University, viewed 2019-08-07 2019.
NAA, T.N.A.o.A. 2009, *Transportation to Australia*, The National Archives of Australia.
Naznin, F., Currie, G. & Logan, D. 2016, 'Exploring the key challenges in tram driving and crash risk factors on the Melbourne tram network: tram driver focus groups'.
Neumann, N. & Böckenholt, U. 2014, 'A Meta-analysis of Loss Aversion in Product Choice', *Journal of Retailing*, vol. 90, no. 2, pp. 182-97.
NSW-RMS, R.a.M.S. 2012, *WestConnex travel time savings fact sheet - December 2012* State of NSW.
NSW-RMS, R.a.M.S., INSW, I.N. & TfNSW, T.f.N. 2012, *WestConnex – Sydney's next motorway priority*, State of NSW.
NSW-Treasury 2017, *NSW Budget 2017-18 - Budget Paper No. 3 - Budget Estimates*, State of New South Wales, Sydney, New South Wales, Australia, <https://www.treasury.nsw.gov.au/documents-resources/documents-library?field_category_target_id=23&field_date_year_value=2017

https://www.treasury.nsw.gov.au/sites/default/files/2017-09/2017-18%20Budget%20Papers%20-%20Budget%20Paper%20No.%203%20-%20Budget%20Estimates-2.pdf>.
NSW, D.o.F.a.S. 2019a, 'Spatial Information eXchange (SIX) Maps', Series Spatial Information eXchange (SIX) Maps State of NSW, <http://maps.six.nsw.gov.au/>.
NSW, H. 2019b, 'Injury and poisoning hospitalisations by leading cause - NSW by Year', Series Injury and poisoning hospitalisations by leading cause - NSW by Year State of NSW, <http://www.healthstats.nsw.gov.au/Indicator/inj_hos_cat/inj_hos_cat?&topic=Injury%20and%20poisoning&topic1=topic_inj&code=inj>.
NSW, S.o.N. 1974a, *Sydney Area Transportation Study (SATS) - Volume 1*, State of New South Wales, <https://www.opengov.nsw.gov.au/publications/17498>.
NSW, S.o.N. 1974b, *Sydney Area Transportation Study (SATS) - Volume 2*, State of New South Wales, <https://www.opengov.nsw.gov.au/publications/17513>.
NSW, S.o.N. 1974c, *Sydney Area Transportation Study (SATS) - Volume 3*, State of New South Wales, <https://www.opengov.nsw.gov.au/publications/17515>.

NSW, S.o.N. 1974d, *Sydney Area Transportation Study (SATS) - Volume 4*, State of New South Wales, <https://www.opengov.nsw.gov.au/publications/17516>.
NSW, S.o.N. 2019c, 'NSW Globe', Series NSW Globe State of NSW, Sydney, New South Wales, Australia, <http://globe.six.nsw.gov.au/>.
NSW.IPC, I.a.P.C. 2017, *Information Protection Principles (IPPs) - for agencies*, State of New South Wales, viewed 2019-11-30, <https://www.ipc.nsw.gov.au/information-protection-principles-ipps-agencies>.
Waters v Transport for NSW [2018] NSWCATAD 40 (15 February 2018) (2018).
Transport for New South Wales v Waters (No 2) [2019] NSWCATAP 96 (18 April 2019) (2019).
OECD, O.f.E.C.a.D. 1980, *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*.
OECD, O.f.E.C.a.D. 2013, *The OECD Privacy Framework - Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*, viewed 2019-11-30, <http://www.oecd.org/internet/ieconomy/privacy-guidelines.htm

http://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf>.
Oshinsky, D. 2012, 'Strange Justice: The Story of Lawrence v. Texas, by Dale Carpenter', 2012-26-03.
Paine, T. & Philp, M. 1998, *Rights of man ; Common sense ; and other political writings*, Oxford University Press, Oxford ; New York.
Petersen Peter, B. 1999, 'Total quality management and the Deming approach to quality management', *Journal of Management History*, vol. 5, no. 8, pp. 468-88.
Pressman, R.S. 1997, *Software Engineering: A Practitioner's Approach*, Fourth Edition edn, McGraw-Hill Education.
Purdy, G. 2010, 'ISO 31000:2009—Setting a New Standard for Risk Management', *Risk Analysis*, vol. 30, no. 6, pp. 881-6.
QLD-TMR, T.a.M.R. 2019, 'TransLink Public Transport Origin-Destination trips ', Series TransLink Public Transport Origin-Destination trips 2019-10-08 edn, vol. November 2018 State of Queensland, 2020-01-03, <https://www.data.qld.gov.au/dataset/go-card-transaction-data/resource/ea4ac597-f948-4347-8fb9-c6abcf9f677f>.
Quinn, B. 2016, 'French police make woman remove clothing on Nice beach following burkini ban ', *The Guardian (Africa)*, 2016-08-24.
Reveley, W. 1791, *Plan of Jeremy Bentham's panopticon prison*, Wikimedia Commons, <https://en.wikipedia.org/wiki/Panopticon#/media/File:Panopticon.jpg>.
Rihova, I., Buhalis, D., Moital, M. & Gouthro, M.B. 2015, 'Conceptualising customer-to-customer value co-creation in tourism', *International Journal of Tourism Research*, vol. 17, no. 4, pp. 356-63.
RMS-WDA, W.D.A., Jacobs & AECOM 2014, *WestConnex M4 Widening – Environmental Impact Statement – Appendix D – Traffic and transport working paper – working paper 4*, NSW Government, Syndey, New South Wales, Australia.
Robertson, G. 2006, *Crimes against humanity : the struggle for global justice*, Third edition, Australian publication. edn, New Press, New York.
Robertson, G. 2009, *The statute of liberty : how Australians can take back their rights*, Vintage Books, North Sydney, N.S.W.
Roess, R.P., McShane, W.R. & Prassas, E.S. 2011, *Traffic engineering*, 4th edn, Pearson, Upper Saddle River, NJ.
Rooney, S.A. & Rooney, J.J. 2005, 'Lean glossary', *Quality Progress*, vol. 38, no. 6, p. 41.
Rose, J., Mulley, C., Tsai, C.-H.P. & Hensher, D. 2013, 'Will bus travellers walk further for a more frequent service? A Stated preference investigation', paper presented to the *Australasian Transport Research Forum*, Brisbane, Queensland, Australia, <https://www.sciencedirect.com/science/article/pii/S0967070X17309022>.
Rosenfeld, M.J., Thomas, R.J. & Hausen, S. 2019, 'Disintermediating your friends: How online dating in the United States displaces other ways of meeting', *Proceedings of the National Academy of Sciences*, p. 201908630.
SACTRA, T.s.a.c.o.t.r.a. 1994, *Trunk Roads and the Generation of Traffic*, London, United Kingdom.
Salmelainen, P. 1992, *Crime & Justice Bulletin Stealing in NSW*, State of NSW, <https://www.bocsar.nsw.gov.au/Documents/CJB/cjb16.pdf>.
Samra, O. 2017, 'Investigating operating procedures on the Inner West Light Rail to Improve Level Of Service', Capstone thesis, University of Technology Sydney, Ultimo, New South Wales, Australia.

Saulwick, J. 2018, '[NSW] Government creates list, but does not follow it, on train station upgrades', *Sydney Morning Herald*, 2018-04-02.

Saulwick, J. & Kontominas, B. 2015, 'Opal card shuttle: Sydney Light Rail stops in Pyrmont help commuters beat the system', *Sydney Morning Herald*, 2015-04-15.

Schmitt, A. 2016, 'A Year After Houston's Bus Network Redesign, Ridership Is Up', *Streetsblog USA*, weblog, <https://usa.streetsblog.org/2016/08/17/a-year-after-houstons-bus-network-redesign-ridership-is-up/>.

Schwab, K. 2017, *The fourth industrial revolution*, Crown Business.

Shafer-Landau, R. 2012, *Ethical theory: an anthology*, vol. 13, John Wiley & Sons.

Shah, D., Rust, R.T., Parasuraman, A., Staelin, R. & Day, G.S. 2006, 'The path to customer centricity', *Journal of service research*, vol. 9, no. 2, pp. 113-24.

Smalley, J. 2017, 'Improving the level of service of the Sydney Inner West Light Rail in the Sydney city district', Capstone thesis, University of Technology Sydney, Ultimo, New South Wales, Australia.

Smith, A., 1723-1790 2002, *An Inquiry into the Nature and Causes of the Wealth of Nations*, vol. 1, Project Guttenberg.

Smith, J.B. & Colgate, M. 2007, 'Customer value creation: a practical framework', *Journal of marketing Theory and Practice*, vol. 15, no. 1, pp. 7-23.

Smith, S. 2014, 'Muda, Muri and Mura', *Lean & Six Sigma Review*, vol. 13, no. 2, p. 36.

SMPO, S.M.P.O. 2013, *[Westconnex] Business Case Executive Summary*, State of NSW.

Spolsky, J. 2004, *Joel on Software*, Apress.

Steer 2020, 'Estimates of station usage [Great Britain]', Series Estimates of station usage [Great Britain] Datasets and Dashboard, United Kingdom edn, United Kingdom - Office of Road and Rail, 2020-01-19, <https://dataportal.orr.gov.uk/statistics/usage/estimates-of-station-usage/>.

Suárez-Barraza, M.F., Smith, T. & Dahlgaard-Park, S.M. 2012, 'Lean Service: A literature analysis and classification', *Total Quality Management & Business Excellence*, vol. 23, no. 3-4, pp. 359-80.

Sun, L., Lee, D.-H., Erath, A. & Huang, X. 2012, *Using smart card data to extract passenger's spatio-temporal density and train's trajectory of MRT system*, Association for Computing Machinery, Beijing, China.

Sutcliffe, A. & Sawyer, P. 2013, 'Requirements elicitation: Towards the unknown unknowns', pp. 92-104.

Tao, S., Rohde, D. & Corcoran, J. 2014, 'Examining the spatial–temporal dynamics of bus passenger travel behaviour using smart card data and the flow-comap', *Journal of Transport Geography*, vol. 41, pp. 21-36.

Tarafder, A. & Basu, A. 2017, 'For the Many and the Few: What a Fundamental Right to Privacy Means for India', 2017-08-25.

Taylor, J. 2019, 'Myki data release breached privacy laws and revealed travel histories, including of Victorian MP ', *The Guardian (Australia)*, 2019-08-15.

TfNSW, T.f.N. 2001a, *Sydney Strategic Travel Model - Technical Report - Stage 1*, vol. 00-08, State of New South Wales, Sydney, New South Wales, Australia.

TfNSW, T.f.N. 2001b, *Sydney Strategic Travel Model - Technical Report - Stage 2*, vol. 00-09, State of New South Wales, Sydney, New South Wales, Australia.

TfNSW, T.f.N. 2011, *New era in transport from today*, New South Wales, Australia, Sydney, NSW, Australia, <http://web.archive.org/web/20120326100148/http://www.transport.nsw.gov.au/sites/default/files/b2b/releases/20111101-Day-One-TfNSW-joint-release.pdf>.

TfNSW, T.f.N. 2012, *NSW Long Term Transport Master Plan*, vol. 1, New South Wales, Australia, Sydney, NSW, Australia.

TfNSW, T.f.N. 2013, *Sydney's Bus Future - Simpler, faster, better bus services*, State of NSW.

TfNSW, T.f.N. 2014a, 'Rail Journeys and Ticket Issues Data - May 2014 Release', electronic data set.

TfNSW, T.f.N. 2014b, *Transport for NSW Annual Report 2013-14* State of New South Wales, Sydney, NSW, Australia, <https://www.transport.nsw.gov.au/sites/default/files/media/documents/2017/tfnsw-annual-report-2013-2014.pdf>.

TfNSW, T.f.N. 2015, *Opal Data Dictionary - Ticketing Derived Information Project Phase 2*, State of NSW.

TfNSW, T.f.N. 2016a, *Network Maps*, State of New South Wales, Sydney, New South Wales, Australia, <https://opendata.transport.nsw.gov.au/dataset/network-maps>.

TfNSW, T.f.N. 2016b, 'Timetables Complete GTFS', electronic data set, <https://opendata.transport.nsw.gov.au/dataset/timetables-complete-gtfs>.

TfNSW, T.f.N. 2016c, *Train Load Survey: March 2016 - By Line*, State of NSW, Sydney, NSW, Australia, <https://www.transport.nsw.gov.au/data-and-research/passenger-travel/train-patronage/train-loads/train-loads-by-line>.

TfNSW, T.f.N. 2017a, 'Opal Tap On and Tap Off Release 1', electronic data set, <https://opendata.transport.nsw.gov.au/dataset/opal-tap-on-and-tap-off>.

TfNSW, T.f.N. 2017b, 'Opal Tap On and Tap Off Release 2 ', electronic data set, <https://opendata.transport.nsw.gov.au/dataset/opal-tap-on-and-tap-off-release-2>.

TfNSW, T.f.N. 2017c, *Sydney Metro City & Southwest - Sydenham to Bankstown upgrade - Environmental Impact Statement*, vol. 1A – Parts A and B, State of NSW.

TfNSW, T.f.N. 2018a, *About us - Who we are - Culture and values*, viewed 2019-01-17, <https://www.transport.nsw.gov.au/about-us/who-we-are/culture-and-values>.

TfNSW, T.f.N. 2018b, *Future Transport Strategy 2056*, vol. 1, New South Wales, Australia, Sydney, NSW, Australia.

TfNSW, T.f.N. 2018c, *Principles and Guidelines for Economic Appraisal of Transport Investment and Initiatives - June 2018*, vol. 2018, State of New South Wales, Sydney, New South Wales.

TfNSW, T.f.N. 2019a, 'Fare Compliance Survey Results Data', electronic data set, <https://opendata.transport.nsw.gov.au/dataset/fare-compliance-survey-results-data>.

TfNSW, T.f.N. 2019b, 'Opal Trips - All Modes', electronic data set, <https://opendata.transport.nsw.gov.au/dataset/opal-trips-all-modes>.

TfNSW, T.f.N. 2019c, 'Opal Trips - Light Rail', electronic data set, <https://opendata.transport.nsw.gov.au/dataset/opal-trips-light-rail>.

Thornton, S. 2014, *Karl Popper*, Metaphysics Research Lab, Stanford University, viewed 2019-07-15 2019.

Thorpe, V. 2017, 'Sydney Trains: Investigation Of Passenger Behaviour', Capstone thesis, University of Technology Sydney, Ultimo, New South Wales, Australia.

Transurban 2016, *Changed Conditions Ahead - The Transport Revolution And What It Means For Australians - Melbourne Road Usage Study - Report*, Melbourne, Victoria, Australia.

UK, U.K. 1998, 'Human Rights Act', Series Human Rights Act no 19980042, United Kingdom, <http://www.legislation.gov.uk/ukpga/1998/42#reference-c18865841>.

UN, U.N. 1948, 'Universal Declaration of Human Rights ', Series Universal Declaration of Human Rights OHCHR, Office of the United Nations High Commissioner for Human Rights, <https://www.ohchr.org/EN/UDHR/Pages/Language.aspx?LangID=eng>.

UN, U.N. 2016, *The World's Cities in 2016 - Data Booklet*.

Unsworth, B. 2004, *Review of bus services in New South Wales*, State of NSW, <Metrobuses>.

USA, U.S.o.A. 1776, 'Declaration of Independence', Series Declaration of Independence National Archives, <https://www.archives.gov/founding-docs/declaration-transcript>.

Vargo, S.L., Maglio, P.P. & Akaka, M.A. 2008, 'On value and value co-creation: A service systems and service logic perspective', *European Management Journal*, vol. 26, no. 3, pp. 145-52.

VEOHRC, T.V.E.O.a.H.R.C. 2019, *Australia's human rights framework* State of Victoria, viewed 2019-11-18 2019, <https://www.humanrightscommission.vic.gov.au/human-rights/the-charter/australian-human-rights-framework>.

Victor, B. & Cullen, J. 2008, *The Organizational Bases of Ethical Work Climates*, vol. 33.

Vuchic, V.R. 2005, *Urban Transit: operations, planning and economics*, John Wiley & Sons.

Walker, J. 2012, *Human transit : how clearer thinking about public transit can enrich our communities and our lives*, Island Press, Washington, DC.

Walker, J. 2014, 'houston: transit, reimagined', *Human Transit*, weblog, United States of America, <https://humantransit.org/2014/05/houston-a-transit-network-reimagined.html>.

Warren, S.D. & Brandeis, L.D. 1890, 'Right to Privacy', *Harvard Law Review*, no. 5, pp. 193-220.

Whittaker, B. 1999, 'What went wrong? Unsuccessful information technology projects', *Information Management & Computer Security*, vol. 7, no. 1, pp. 23-30.

WHO, W.H.O. 2004, *Guidelines For Medico-Legal Care For Victims Of Sexual Violence — Chapter 2: Sexual Violence: Prevalence, Dynamics And Consequences*, World Health Organisation.

Wolf, B. & Zuckerman, P. 2012, 'Deviant Heroes: Nonconformists as Agents of Justice and Social Change', *Deviant Behavior*, vol. 33, no. 8, pp. 639-54.

Wolmar, C. 2008, *Fire & steam: how the railways transformed Britain*, Atlantic Books London.

Wolmar, C. 2009, *The Subterranean Railway: how the London Underground was built and how it changed the city forever*, Atlantic Books Ltd.

Wood, T. & Porter, E. 2019, 'The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence', *Political Behavior*, vol. 41, no. 1, pp. 135-63.

WSP, R.E.a. 2018, *Latest evidence on induced travel demand: an evidence review*, United Kingdom, <https://www.gov.uk/government/publications/induced-travel-demand-an-evidence-review>.

Zeibots, M.E. 2007, 'Space, time, economics and asphalt: An investigation of induced traffic growth caused by urban motorway expansion and the implications it has for the sustainability of cities'.

Zeibots, M.E. & Petocz, P. 2005, 'The relationship between increases in motorway capacity and declines in urban rail passenger journeys: a case study of Sydney's M4 Motorway and Western Sydney Rail Lines', paper presented to the *Australasian Transport Research Forum*, <https://www.atrf.info/papers/2005/2005_Zeibots_Petocz.pdf>.

Zhong, C., Huang, X., Müller Arisona, S., Schmitt, G. & Batty, M. 2014, 'Inferring building functions from a probabilistic model using public transportation data', *Computers, Environment and Urban Systems*, vol. 48, pp. 124-37.

Zhong, C., Manley, E., Müller Arisona, S., Batty, M. & Schmitt, G. 2015, 'Measuring variability of mobility patterns from multiday smart-card data', *Journal of Computational Science*, vol. 9, pp. 125-30.