# 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.

Signature: Production Note: 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