LEVERAGING BIG DATA TO OPTIMISE NETWORK PERFORMANCE & CONSTRUCTION DELIVERY

Author
Pamela Ward BDes
Product Manager – Transport
Qrious Ltd
Pam@qrious.co.nz

ABSTRACT

Big Data is a buzzword that has permeated the transport industry in recent years, yet the value that can be achieved through its application still remains a hidden art. By introducing the 'Design Thinking' practises adopted by global technology startups at the forefront of innovation, we can accelerate the productisation of data to enable operational gains.

Recent studies have set out to evaluate whether real-time data sources can accurately measure the disruption roadworks and incidents inflict on road users. These studies have proven the ability to successfully monitor the network through various data sources, however actioning these insights to drive change, has proven to be a challenge the industry is still grappling with.

To unlock the benefits of real-time location data, for road users, asset owners, and traffic management staff, Qrious conducted a trial with Fulton Hogan. By utilising the 'Design Thinking' methodology the final output was a software application that: 1. enables proactive delay and queue mitigation, through real-time alerts and, 2. detects abnormal congestion, and identifies the exact location of an incident. This functionality helps to optimise operational costs, intuitively assists with traffic management, and minimises disruption for road users.

INTRODUCTION

Historically the success of transport industry initiatives, were measured by the kilometres of tarmac laid, or the magnitude of the infrastructure constructed. However, in recent years there has been a significant shift in the methodology applied to Transport policy, with a new focus towards 'Customer Outcomes'. This new approach aligns with emerging global innovation methods developed through the discipline of Human Centred Design (HCD). The adoption of HCD methods is evident throughout the New Zealand Transport Agency's (NZTA) policy framework, and features heavily in their latest Annual Report, prefaced with the 'Purpose' statement: 'Creating transport solutions for a thriving New Zealand'.

To measure the success of government funds invested into the industry, a new set of metrics have been developed by the NZTA to enable 'customer outcomes' to be measured. These outcomes include: the effectiveness of people and freight movement, the resilience and efficient delivery of infrastructure and services, and the safety gains – both to users of the network, and contractors alike. With the introduction of new success metrics, the need for more advanced data collection methods, analytics, and reporting capability, has emerged.

While an appetite for more advanced data capability presents an exciting opportunity for the development of new technology solutions, the rate at which these new technologies are being explored is hindered by the politically sensitive nature of data application. Qrious' transport data research has been focused on the application of location data, however, these common concerns are applied by both national and local government to many transport-related datasets. Particular concerns include a) data accuracy – more specifically, sample size and collection method, b) data ownership, to ensure value for money and industry accessibility, and c) data compatibility, to enable consistent reporting and reusability. While these are all valid concerns, there is a significant discrepancy between the rate at which they apply to various use cases. In order to optimise outcomes for the industry, government, and general public, it is important that the concerns above are applied on a case-by-case basis so that the adoption of data-powered technology solutions is not inhibited.

Opportunities to Improve Network Performance with GPS Probe Data

Preferences for data collection methods are commonly being driven by a desire to 'own' a dataset. Due to this desire, alongside a residual tendency to measure success by the installation of physical infrastructure, it is no surprise that the data collection methods gaining traction tend be those of a physical nature. These data collection methods typically take longer to implement, and require ongoing testing and maintenance.

In the case of data sources for monitoring traffic flow, NZTA and many of the Road Controlling Authorities (RCA's) in New Zealand deem Bluetooth and Wi-Fi sensor data, more suitable than GPS probe data. This is evident the NZTA's 'NIEMS project' that mandated the recent installation of Bluetooth sensors across much of the Auckland road network, as a means to improve network operations and performance.

Despite a notable reluctance from transport governing bodies to adopt GPS data sources for network performance use cases, Qrious has undertaken a series a data trials to test the capability of Google data. This dataset was selected due to it's market-leading accuracy in real-time 'travel time' data across the vast majority of New Zealand's road network, as well as its ease of scalability across regions without the overhead of installing sensors. Google cleanses the data that is supplied via their API's to filter out noise, such as devices used by pedestrians, so that it is ready to be consumed by web applications that require data for travel times across specified sections of road.

The Influence of Technology Development Methodologies

The rate of development of new technologies, has not only been impacted by the preference for physical sensors, but also the approach to Research and Development, and procurement. Historically, ‘Waterfall’ has been the most common methodology applied to the development of technology solutions. In this process, all functional and nonfunctional requirements are determined upfront by project stakeholders. Rarely end users of the system are consulted, and little opportunity is allowed for feedback or changes in scope to be introduced during the delivery phase. With Waterfall, when a project is ‘delivered’ the solution is considered complete, therefore ongoing improvements are seldom budgeted for. This approach to product development typically leads to solutions that don’t address the root problem, or fail to solve the problem in an intuitive way for end users of the system.

The Waterfall process can become even more problematic when used to develop data-powered solutions. Through the requirements gathering phase, the stakeholders’ focus tends to be diverted to the mechanism at which data is collected and presented rather than identifying the root problems their contractors, customers, and end users are needing to solve. Often, relatively simple datasets can adequately solve the problem at hand, however, the aforementioned common concerns regarding the source of the data, can unjustifiably influence these decisions and add significant overhead to the project. By the time a project is fully scoped, the requirements are frequently biased towards providing better performance visibility to stakeholders, with an assumption-based secondary set of features for the end user to support operations. This Waterfall approach often leaves stakeholders with a bitter taste in their mouths after many expensive solutions are developed that don’t deliver meaningful operational gains. Instead these technologies tend to result in low user engagement, leaving governement with little choice than to legislate to enforce adoption of the technology, or to develop a replacement.

The pain of this process has been a catalyst for power struggles between the transport industry, government, and technology developers – whom all share the common goal of creating valuable solutions, but have been hindered by a methodology that inhibits collaboration. One example is the way the Waterfall methodology encourages stakeholders to start ‘solutioning’ right from project inception. It is difficult to gather meaningful requirements collaboratively, because stakeholders tend to quickly become fixated on data and technology choices, driving their politcal agendas, and voicing their knowledge and ideas. Consequently, the functionality that is needed to solve the important problems faced by their team, is seldom captured in the requirements. In order to combat this result, the Agile and Design Thinking methodologies were developed, which seek to include a broad spectrum of end users and stakeholders when identifying problems and iterating on solutions throughout the delivery phase.

Application of Design Thinking and Agile Development Methods

One of the fundamental differences between Waterfall, and Agile (and Design Thinking) methods, is that Agile adopts an iterative approach to developing technology. This means that very little time is spent upfront trying to define every feature, instead a set of problems and possible solutions are determined and prioritised, then the design, development, and testing begins. While this process is often scary for stakeholders who are used to all requirements being defined from the outset, it is considerably lower risk, as ‘shippable’, valuable features are delivered at a much higher frequency, allowing opportunities for changes in scope throughout development phase.

At the inception of a project, Design Thinking methods are used to explore the problem space by empathising with key stakeholders and end users. This involves interviews, observation, focus groups, and workshops, to gain a holistic understanding of the problems they face. Throughout this phase, insights into the current solutions are usually uncovered, as are ideas for possible solutions.

The solution ideas are then filtered through a technical feasibility assessment, where possible feature sets are identified. These features typically go through a phase of lightweight prototyping and user testing to refine functionality, and determine priority. A 'Minimum Viable Product' (MVP) feature set is identified based on the minimum amount of functionality required for delivering value to a segment of end users. For data-powered solutions, this usually involves displaying, or supplying the dataset's in a way that adds extra value on top of it's raw form.

Once the MVP is determined, features are prioritised based on value to stakeholders and end users, with consideration towards the technical complexity and design. One of the most popular Agile frameworks known as Scrum, is based on Sprint cycles – the time-boxed delivery of features. Each Sprint runs for 1-4 weeks, and the scope of the work is determined by what is feasibly 'shippable' at the end. This ensures that new functionality is being supplied to an end user for testing and feedback at regular intervals throughout the development phase. This feedback is surfaced through the product backlog as improvements or new features, many of which would never have been identified without the opportunity to test, and iterate accordingly.

Completing the MVP is a key milestone for considering the direction and future scope, as well as scaling to a larger set of end users. The entire process aims to be as 'lean' as possible, where rapid iterations minimise wasted time building features that don't add value. While these methods challenge the Waterfall nature of government and corporate procurement processes, there is a growing appetite to explore the value and cost savings that can be realised.

CASE STUDIES

Qrious adopted Lean Agile and Design Thinking methods to test and develop data-driven solutions with Fulton Hogan, for 1. minimising the disruption during construction delivery for road users, and 2. improving incident response capability.

The Design Thinking process began by identifying the 'location-based' data-related problems, that key stakeholders were looking to solve. This involved an initial brainstorming session of the various contract KPI's Fulton Hogan were trying to measure and prove capability against. Two common themes were highlight that fitted within the umbrella of traffic management for planned (i.e. roadworks), and unplanned (i.e. incidents) events.

Following the brainstorming session, Qrious ran workshops with a broad spectrum of Fulton Hogan staff that were involved in the planning, delivery, and reporting of road construction projects, and incident response. In the workshop participants pain points, current solutions, and solution requirements were drawn out using Design Thinking methods (see **Figure 1**).

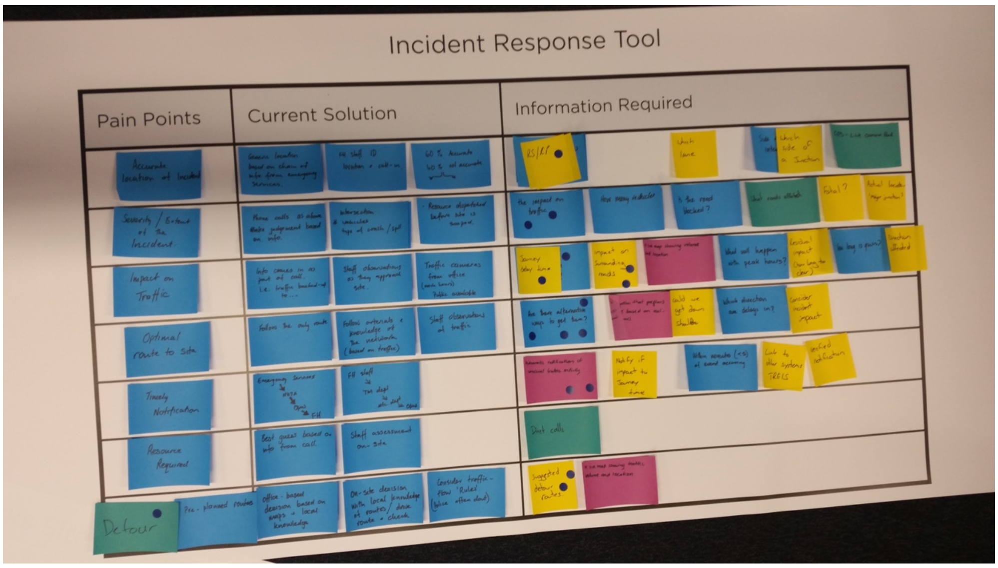

Figure 1. Workshop Template for Incident Response use case

Follow up interviews were conducted with various Fulton Hogan ‘end user’ segments, across the country. The research also included input from Traffic Operation Centres (TOC’s), NZTA, and industry experts. These interviews began by validating the ‘problems-to-solve’, followed by a second round for testing prototypes of proposed solution ideas.

1. Minimising Disruption during Construction Delivery

The MVP for the Construction Delivery solution was defined based on the output of the research process described above. An initial goal of testing the suitability of Google data for identifying delay times and queue length surrounding roadworks sites was determined. To ensure that the dataset was providing accurate, real-time travel time data, a trial was implemented on the Riccarton Road/Deans Ave Roadworks site. To test the dataset, an SMS message was sent to Traffic Management staff, notifying them of queue length and delay times, for the various routes through the site. Data was also presented to them on a graph, to validate against daily activity reports. Further measures were taken to test data accuracy, including comparisons against CCTV camera data, and congestion heatmaps from the Google Traffic application.

The Agile development process enabled learnings to be taken from the testing and immediately applied to the algorithms as they were being developed. Through this testing process many needs were identified that didn’t surface through the workshops, including:

- The need for alerts when congestion on the surrounding network was likely to cause adverse effects on traffic flow through the site
- The ability to share data with clients to inform the scheduling of Roadworks
- Contextual information to highlight the cause of congestion

The initial SMS solution, proved the capability of the Google dataset to accurately identify travel time delay, queue length, effect on the surrounding network, and the rate at which these factors are fluctuating in real-time. These functions enabled traffic management staff to prepare for unpredicted traffic flow, and mitigate disruption before it escalated.

Upon proving this capability, more advanced features have now been developed for visualising the data in a way that addresses a series significant challenges faced during the planning, delivery, and reporting, of road construction projects.

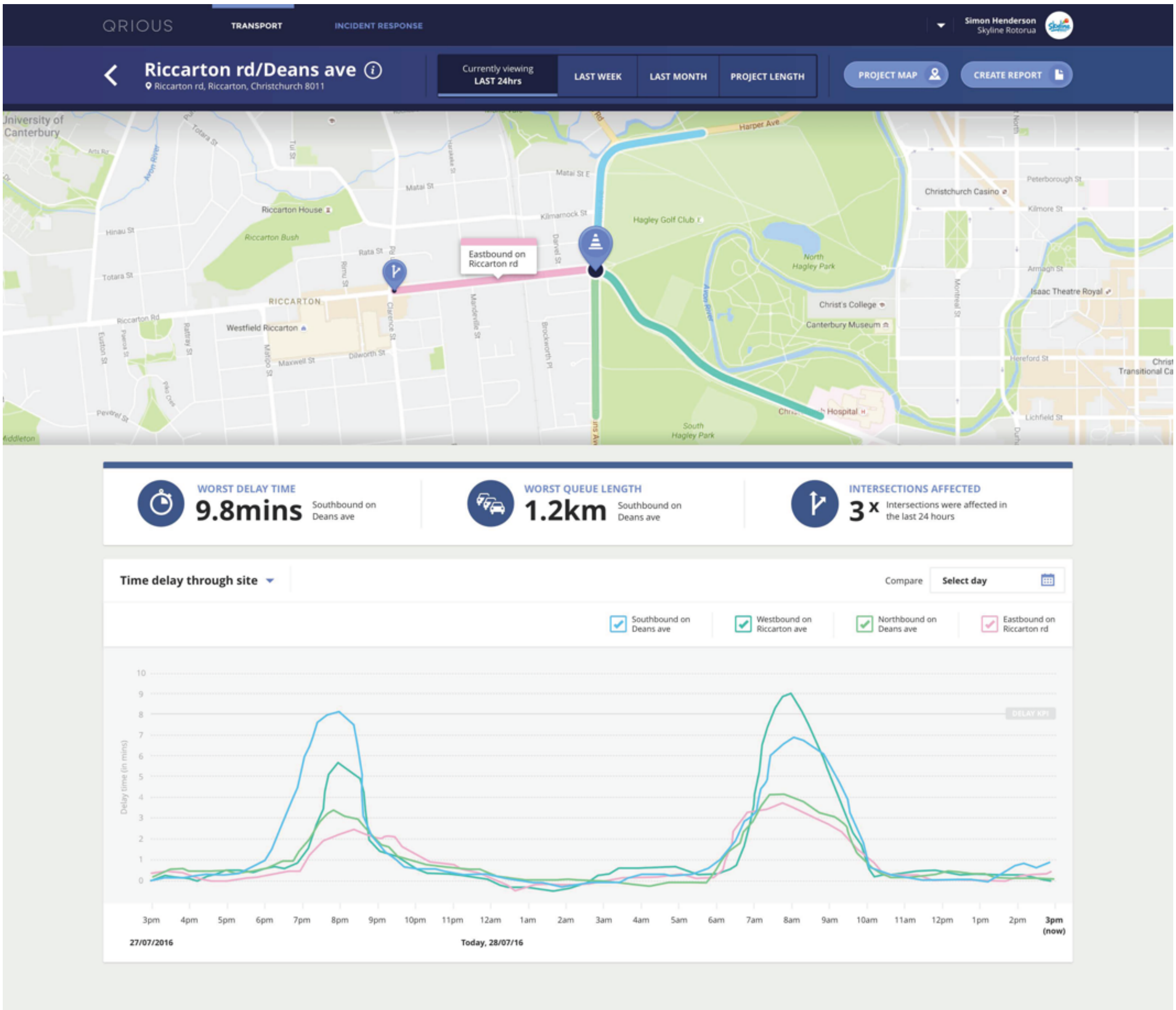

Figure 2. Application Dashboard: Travel time delay graph

2. Improving Incident Response Capability

For the Incident Response use case, the initial problem raised by stakeholders, was the difficulty in measuring response time for KPI reporting (i.e. the time it took from being notified about an incident, to arriving at the scene). Through the workshops a further set of problems were identified, including:

- A verified, accurate location of the incident
- Optimal route to the incident location
- Best detour route
- Effectiveness of traffic management strategy
- Coordination of response effort with other response units, traffic operations centre's and other external stakeholders

An initial trial was conducted to test the ability of Google data to identify if an incident location could be pinpointed down to the nearest 50m, and whether the flow on effects could be dynamically monitored and attributed accordingly. To validate this trial, Fulton Hogan response teams provided information about incidents that occurred across 4 trial locations, it was also tested against TREIS data and crash reports.

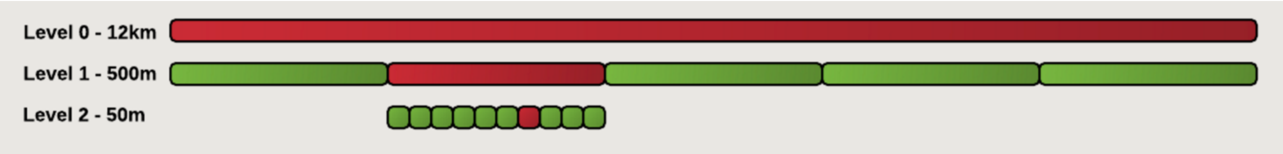

Figure 3. Dynamic Polling logic visualisation

In order to identify an incident within 50m of it occurring, a Dynamic Polling algorithm was developed. The logic fueling this algorithm, has been designed to identify 'abnormal congestion' with the following parameters:

- When a long road segment (i.e. 12km) reaches a travel time duration that is x% above normal, then activate the polling at a more granular level (i.e. 500m).
- When a 500m segment has reached a travel time that is x% above normal, then activate the polling of a 50m segment (as seen in **Figure 3**)

This logic is applied for both directions of traffic to ensure an accurate understanding of traffic flow activity can be determined.

A further degree of sophistication must be applied to the logic to determine whether an instance of 'abnormal congestion' is worth flagging, and the probability that an incident has occurred. These parameters include:

- Severity of delay (i.e. % above 'normal')
- Length of queued traffic
- A static 'starting point' for the location of the abnormal congestion
- Congested segments ahead of the 'starting point'
- Location-specific delay trends (i.e. is this a recurring issue)
- Directions affected
- Congestion on surrounding roads
- Length of time 'abnormal congestion' has been occurring

This logic relies on an accurate baseline for 'normal traffic flow' to be defined for each segment, in order to filter out cases of peak congestion, traffic lights, turning lanes, motorway on-ramps, intersections, merging traffic, changing speed limit zones, etc. See **Figure 4** for an example of an 'abnormal congestion' event. The challenge in this case, is determining the probability that an incident has occurred. At 5.34pm you can see an instance where an incident may have been detected, had the algorithm not been designed to factor in the preceding congestion trends. This is just one example of the complexity of the logic required to fuel an algorithm for cases such as this.

To apply this algorithm to traffic management and operations use cases, a series of thresholds must be determined for each of the parameters listed above. Setting these thresholds to the correct sensitivity, is critical for ensuring that only meaningful alerts are sent to an end user. In the case where the sensitivity is set too low, and the user is spammed with too many alerts, there is a risk that they disengage with the system altogether. However, if it is set too high, then the risk is that incidents occurs without triggering an alert in time. Naturally the user will allow for a margin of error in either case, as the system is not designed to be the only means in which an incident is identified, however to promote adoption of the technology for these use cases, a degree of accuracy is imperative.

In order to refine this algorithm, a further round of testing has commenced, which will use traffic camera data, TREIS data, and feedback from Incident Response staff for validation. The Agile methodology promotes lightweight testing of the technology, through the rapid delivery of new iterations, so that no time is wasted in over-engineering the solution before it is tested with users.

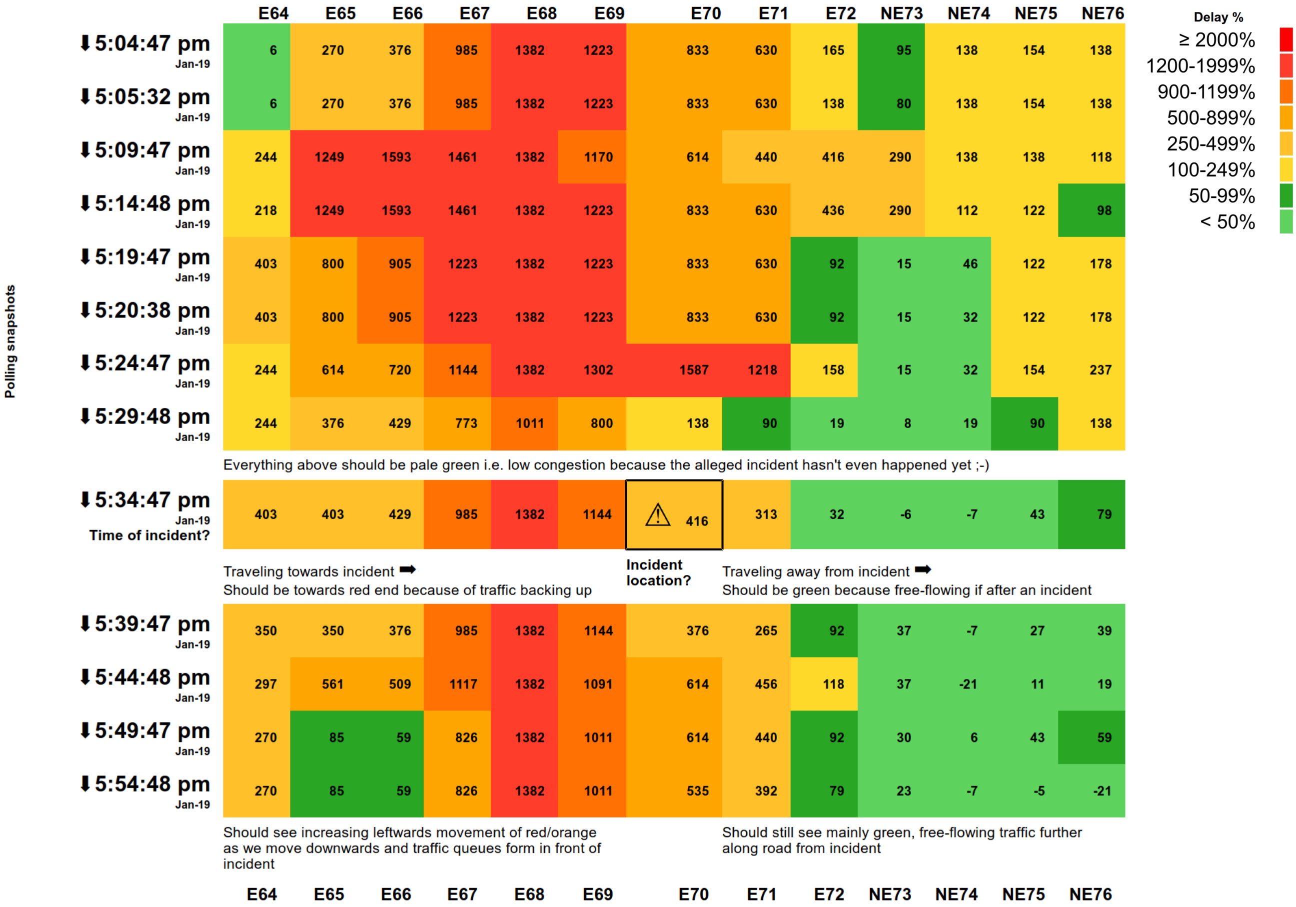

Figure 4. Visualisation of real Abnormal Congestion Occurrence

The underlying capability this system provides, can be consumed by a software application which displays a visual representation of abnormal network congestion. This interface can provide the contextual information to assist Incident Response staff and Traffic Operations Centres, with managing their response (see **Figure 5.**)

Through the research, a series of interviews were conducted with staff in these roles, to explore the ways this dataset could to solve the challenges they face when identifying and responding to incidents. The following needs were identified as problems that could be solved through the application of sophisticated logic applied to location datasets:

- Timely notification and accurate location of an incident
- Knowing the impact on traffic
- Determining the optimal route to access the incident site
- Knowing how many staff are required to manage traffic
- Coordinating the response
- Identifying the best detour route and how it's performing

The software application in **Figure 5** has been designed to resolve this set of challenges, and has been tested by Incident Response and Traffic Operation staff for feedback before the development begins.

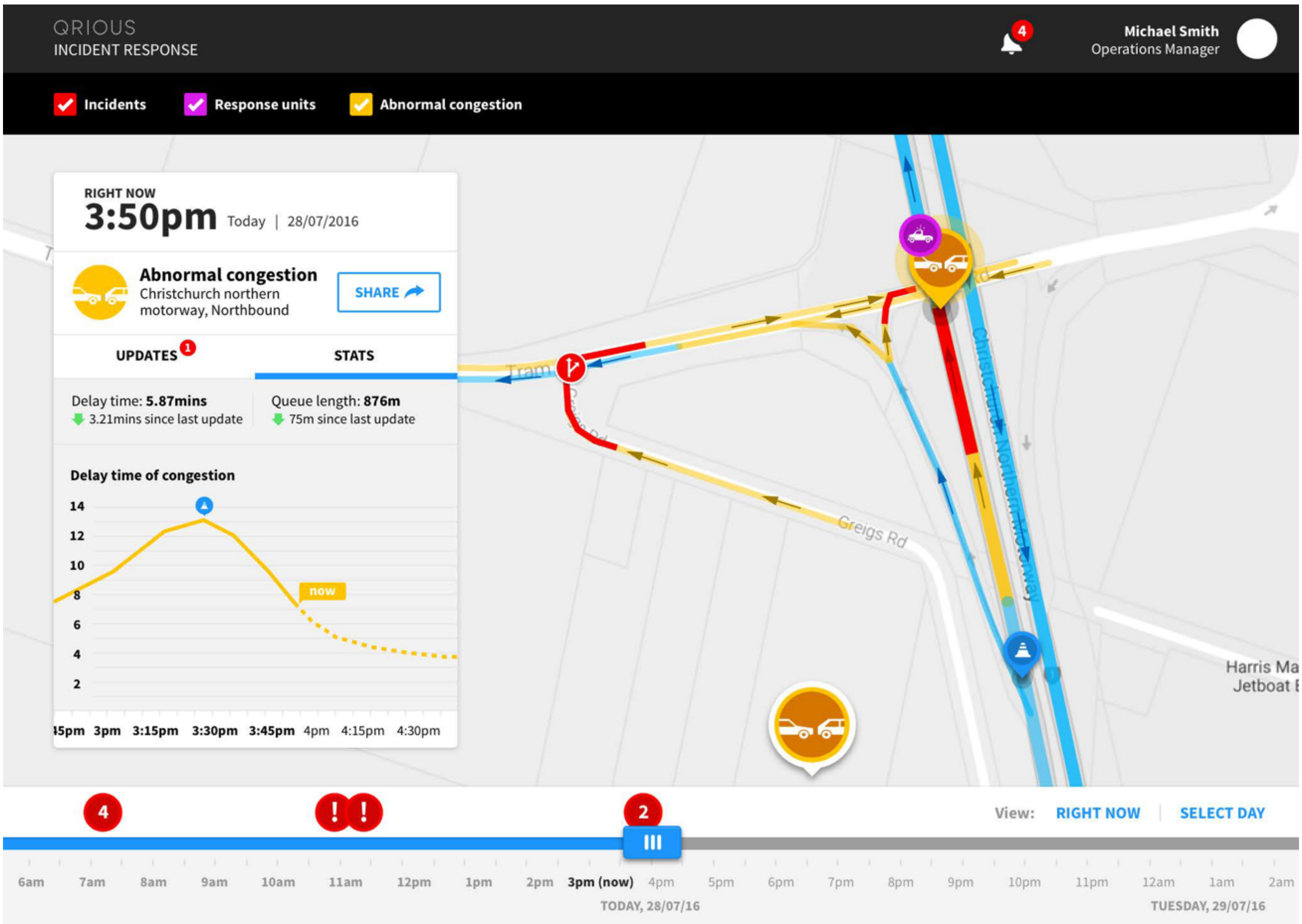

Figure 5. Incident Response Prototype: monitoring detour route performance

CONCLUSIONS

Through the adoption of Design Thinking and Agile methodologies, technology development processes can be significantly faster and more cost effective than those experienced in Waterfall projects. Stakeholder engagement throughout the project delivery phase, results in a lower barrier to adoption and enables solutions to be created, that address the needs of end users as well as business stakeholders.

In the case studies presented, the datasets applied to seemingly complex use cases, were incredibly simple. This was enabled by a pragmatic approach to solving problems, rather than a quest to finding the perfect dataset. Through the smart use of data coupled with the methods of Design Thinking, we were able to build solutions that have the potential to optimise operational costs, improve traffic management, and minimise disruption for road users.

Findings surfaced through these data trials provides sufficient evidence that GPS probe data can be applied to various network performance use cases, and can be rapidly tested using Agile and Design Thinking methods. Industry and government collaboration is critical as we strive towards better transport outcomes, even if it does mean enduring the pain that comes with challenging traditional approaches to procurement and project delivery processes.