SOLVING BIG CHALLENGES WITH BIG DATA

AUTHORS

Chris Vallyon MIR (Merit), LLB, PLSC, BA,
Senior Transportation Analyst, Beca
chris.vallyon@beca.com – Presenter

Richard Young, MICE, MBA, MSc, BSc (Hons)
Senior Project Manager, Beca
Richard.young@beca.com

Tyler Ross, B.Sc(Tech) - Presenter
Transport Outcomes Planner, New Zealand Transport Agency
Tyler.ross@nzta.govt.nz – Presenter

ABSTRACT

Following on from their 2013 IPENZ Paper on the challenges faced collecting data on the Roads of National Significance, Beca and the NZ Transport Agency team up again to discuss: Big Challenge that can be a Big Frustration - New Zealand has 95,000km of road network, which is relatively large for our population. How much do we really know about road user experience on the network?

This paper discusses the challenge for engineers and planners managing an extensive national road network in a country lacking monitoring infrastructure and how changes in technology and emerging ‘Big Data’ systems are making management more achievable. From the analysis of GPS data from over a million vehicle trips across the country in a recent study, this paper presents the study’s findings on the performance of 4,000km of rural strategic State Highway (8,000km of data). Lastly, we pose the question - how long before the performance of a road network that is largely invisible, will becomes revealed at the click of a mouse, and what will this mean for Road Controlling Authorities (RCAs), engineers and planners, and the travelling public?
INTRODUCTION (THE NEW ZEALAND CONTEXT)

New Zealand is a country that faces many infrastructure challenges. By population New Zealand is a small country, but in terms of land area, it is larger than Great Britain, more than three times larger than Ireland, Portugal or South Korea, and more than six times larger than Switzerland, Denmark, or Taiwan. This low population density poses significant challenges for providing an efficient and reliable transportation network.

The strategic state highway network connects New Zealand’s -mostly coastal- populations through terrain that is often rugged and sparsely populated. It would be possible to talk about population density. However, to put this into a road asset ownership perspective, a more interesting way of phrasing this issue is to look at the number of people per km of road network.

Table 1 – Population per km of road asset (selected countries).

<table>
<thead>
<tr>
<th>Country</th>
<th>Population (000s)</th>
<th>GDP per Capita (US$)</th>
<th>Road Assets (km)</th>
<th>People per km road assets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>22,260</td>
<td>43,300</td>
<td>823,200</td>
<td>27</td>
</tr>
<tr>
<td>New Zealand</td>
<td>4,360</td>
<td>30,200</td>
<td>94,200</td>
<td>46</td>
</tr>
<tr>
<td>United States</td>
<td>316,670</td>
<td>50,700</td>
<td>6,506,200</td>
<td>49</td>
</tr>
<tr>
<td>Ireland</td>
<td>4,775</td>
<td>42,600</td>
<td>96,000</td>
<td>50</td>
</tr>
<tr>
<td>Denmark</td>
<td>5,555</td>
<td>38,300</td>
<td>73,900</td>
<td>75</td>
</tr>
<tr>
<td>Portugal</td>
<td>10,800</td>
<td>23,800</td>
<td>82,900</td>
<td>130</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>63,395</td>
<td>37,500</td>
<td>394,400</td>
<td>161</td>
</tr>
<tr>
<td>Malaysia</td>
<td>29,630</td>
<td>17,200</td>
<td>144,400</td>
<td>205</td>
</tr>
<tr>
<td>Chile</td>
<td>17,215</td>
<td>18,700</td>
<td>77,800</td>
<td>221</td>
</tr>
<tr>
<td>South Korea</td>
<td>48,955</td>
<td>32,800</td>
<td>105,000</td>
<td>466</td>
</tr>
<tr>
<td>Taiwan</td>
<td>23,300</td>
<td>39,400</td>
<td>41,500</td>
<td>562</td>
</tr>
</tbody>
</table>

As can be seen in Table 1, above, New Zealand has a combination of a relatively low GDP per capita, combined with a low number of people for the amount road asset infrastructure required for our land area and agricultural land usage. When countries are ranked in this manner, New Zealand has one of the lowest rates of people per km of road asset infrastructure. Although ranked higher than Australia, New Zealand per capita GDP is lower than Australia’s.

To help resolve this challenge, New Zealand has a central road controlling authority (the NZ Transport Agency) responsible for almost 11,000km of road network vital for linking towns and cities, often in areas that would not have the population to fund these roads locally. The Agency has significant challenge in understanding the performance of its assets, particularly in rural areas.

The government’s “Roads of National Significance” programme has led to some significant investment in some of the key corridors linking urban areas, but often falling outside the coverage of existing infrastructure used for monitoring network performance. This creates interesting challenges.

How do we understand the performance of a large network of road assets? And how do we demonstrate effective operational and investment decisions without spending a lot of money? This paper summarises a proof of concept for moving from the historic floating vehicle methodology to using GPS data sourced from third party commercial vehicles, in order to test whether or not commercial vehicle GPS can be used to improve understanding of the road network, in particular, the strategic rural corridors where there is currently very limited data available.
HISTORIC PERFORMANCE MONITORING

New Zealand has had a national congestion monitoring programme since 2002. This collects information twice a year to understand urban congestion in Auckland, Wellington, Christchurch, Tauranga and Hamilton. The data is used for a variety of purposes, including understanding the impacts of investment in the road network.

In recent years, where they are not already included in the survey network, coverage of these urban areas has expanded to include the Roads of National Significance. This means that the Wellington coverage now stretches as far north as Foxton and the Auckland coverage includes rural strategic highway as far north as Wellsford.

Although these urban areas have expanded, the emphasis for these surveys has been on the original urban areas. The total monitored coverage of the biannual congestion monitoring programme is around 1,500km (3,000km of data, once both directions are counted) as compared to the roughly 95,000km of New Zealand road network. Other forms of infrastructure that have remained sparse outside of major urban corridors include speed loops, radar and CCTV. As a result there are significant gaps in our knowledge of the complete transport network.

There are also significant limitations to expanding the national programme. The floating vehicle survey methodology does not lend itself to the scalability required for filing the gaps between urban areas. While it is possible to use the floating vehicle methodology to understand broad annual travel time trends, the limited sample size misses a issue of reliability of travel times for rural roads.

This means that, unless alternative methodologies are developed, RCAs have little information to describe road user (customer) experience for the vast proportion of the New Zealand road network.

Why GPS?

The following Table 2 provides a concise summary of different RCA functions, the data requirements to assist those functions, and the key technologies currently available.

<table>
<thead>
<tr>
<th>RCA Functions</th>
<th>Temporal data Requirements</th>
<th>Spatial Data Requirements</th>
<th>Intensity / Impact Assessment</th>
<th>Current NZ Data Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) Real time operations</td>
<td>Importance: Very High</td>
<td>Importance: Moderate</td>
<td>Importance: High</td>
<td>CCTV, Aggregate of point speed technologies, ANPR/Bluetooth, Commercial GPS</td>
</tr>
<tr>
<td>(ii) Proactive operational decisions / Works planning</td>
<td>Importance: High</td>
<td>Importance: High</td>
<td>Importance: High</td>
<td>Aggregate of point speed technologies, ANPR/Bluetooth, Commercial GPS</td>
</tr>
<tr>
<td>(iv) Strategic KPI reporting</td>
<td>Importance: Moderate</td>
<td>Importance: Very High</td>
<td>Importance: High</td>
<td>ANPR/Bluetooth, Commercial GPS, Floating Vehicle Surveys, Volume data</td>
</tr>
</tbody>
</table>
One technology that has the potential to sit across all RCA functions is that of commercial/fleet GPS data. That is to say, GPS data derived from fleets of vehicles, rather than individual survey vehicles. For this reason, fleet/commercial GPS has been investigated further.

GPS: MOVING FROM THE FISHING LINE TO THE FISHING NET

Conducting a floating vehicle survey involves paying someone to drive a set route, with a set start time. Each time the survey is undertaken the route and start time are kept consistent. The strength of a floating vehicle surveys is that GPS usage allows the exact time and location of a vehicle to be known, which provides a very accurate assessment of travel speeds and delay. Speeds can then be compared between corridors and cities over time.

The primary limitation of the floating vehicle surveys are the limited sample size. To increase sample size while maintaining the floating vehicle survey methodology requires more cars, more drivers, and more petrol, which makes this an expensive proposition.

To move from one GPS survey vehicle to many vehicles requires a fundamental change in methodology, similar to the step change of moving from a fishing line to a fishing net.

To implement a cost effective sample size change requires procuring data from a third party, i.e. finding and purchasing data from a commercial company whose business practices result in the creation of GPS data.

A lot of vehicles now have some form of GPS device in them when they travel. These devices can be roughly split into three different categories:

- Off the shelf in-vehicle navigation systems
- Smart phones
- Fleet management systems

Of these, the first to enter the vehicle fleet were the off the shelf in-vehicle navigation systems. A road user pays a company for a system that uses mapping information to provide advice on route choice.

Navigations systems have the ability to also inform route choice based on the speeds of other vehicles along the intended route equipped with similar devices in a shared vendor network. For this purpose, even a small amount of data (delay vs freeflow) is useful for the navigation system. If the customer base grows, so too does the amount of data collected. As a source of data to understand asset performance, the customer base – or market share – determines the usefulness of the data source. The system relies on members of the public continuing to purchase, and continuing to use and update their in-vehicle navigation system, with the GPS data being produced as a by-product. If people stop using their devices, the system stops receiving information about traffic conditions. To counter a declining sample size, data from these systems data tends to be aggregated to averages over longer periods. This has the effect of diluting the quality/resolution. Typically, these systems anonymise data by aggregating based on a proprietary link model. This process loses the information that would inform an understanding of driver behaviour (e.g. route choice, origin-destination behaviour, demand studies, etc). Differences in link models between vendors can also introduce complications when trying to collate / aggregate data from multiple vendor sources.

Increasingly, smart phones are becoming a ubiquitous part of society. It is now possible to download a cell phone application to provide a navigation service (often for free) that in the past would have required a customer to go out and purchase purpose built hardware. The service provided by smart phones has made in-vehicle navigation hardware technology
vulnerable to changes in the market place and potentially redundant in the long term, and therefore an increasingly risky source of GPS data. Additionally, smart phones can be matched through a combination of GPS, WiFi and cellular phone mast triangulation. The availability of this location data, potentially for all mobile phones (not just smart phones) for traffic monitoring purposes is currently an area of much public debate relating to privacy. In New Zealand cellphone companies have yet to make their phone networks available for this type of service.

This leaves the third category, fleet management systems. These are proprietary systems, often provided by third parties, that enable a company to optimise the performance of its fleet of vehicles. The systems typically record driver behaviour, routing, public road and private road distances driven and enable the automatic processing of Road User Charges. To achieve these functions the systems incorporate a GPS. As with other sources of GPS data, the market share of the providing company is important, but unlike in-vehicle navigation, the GPS is a bi-product of commercial services that companies are willing to pay for. Private companies gain a commercial advantage from improved fleet management tools, and in doing so, create a potential source for useful GPS data.

For all three forms of GPS data, strongly public perceptions around individual privacy concerns, are a key limiting factor for data availability. Companies that own the data also wish to minimise perceptions around the use of the data (valid or entirely fanciful) that might lead to reputational damage and undermine the owning company’s market share.

GPS PROOF OF CONCEPT
The New Zealand Transport Agency has the task of operating a network of almost 11,000km of road assets, which together form the key links between the nation’s population centres. While the Agency does collect rural information relating to road safety, pavement quality and volume / demand, questions remain over route choice decisions and the average travel times, reliability and level of service provided to members of the public using the Agency’s assets.

The Agency recently completed a trial proof of concept for using commercial GPS data to assess 4,000km of rural strategic state highway (8,000km of data, if counting both directions of travel). The study network was selected based on parallel work being undertaken as part of the Agency’s Safer Journeys work. The goal was to identify a suitable means of assessing network performance for a predefined strategic network. It was not intended to assess the entire state highway network; however, the findings would ‘test the waters’ for a methodology for wider network performance monitoring.

In undertaking this trial, new challenges emerged relating to how the data could be managed.

Conducting the biannual floating car surveys involves managing a team of over 50 people in five urban areas. The results include about 3 million GPS points, which are aggregated to form the basis of reporting in map, graph, and network level formats.

Managing data suitable for the vast network of rural strategic assets, however, requires a whole different quantum of data. The Agency’s proof of concept was derived from over 950,000 unique commercial vehicle trips identified from a source of over 450,000,000 points of GPS data. Along with developing processes for protecting the privacy and commercial sensitivities, the sheer volumes of data provides a big technical challenge.

“Big Data” is often described as any dataset too large for Excel. This definition is perhaps insufficient in the transportation space, as even the biannual surveys generate enough data to require management in a purpose built environment.

To be able to report on a larger, rural strategic network it was necessary to find a third party source
of data. Once a suitable source was identified, a system then had to be created to manage and interrogate the raw GPS data.

The system used for interrogating commercial GPS data had to be sufficient to manage the current data set of roughly 2 billion GPS points per year. However, as vehicles increasingly adopt GPS based fleet management practices, the sample rate will continue to grow, and even this figure of 2 billion GPS points per year is likely to be dwarfed over time.

To achieve this feat in the transport space and create a system that was suitable, scalable, and sufficiently robust, the Agency called upon software expertise more often associated with banking software and military surveillance data warehouses. The results of the proof of concept reveal the performance of parts of the network where such data has never before been available.

The following Figure 1 provides an illustration of average travel speeds on rural strategic state highways for the North Island, for the month of June 2013, derived using commercial GPS data (this forms one of a series of similar maps for 2013).

As can be seen from Figure 1, the result successfully demonstrates a means of reporting on key corridors between urban areas and townships. Most of the intended coverage was possible with the commercial vehicle data source, with only a few locations (such as Milford Sound) having too little data to be considered reliable.

The expanded network coverage provides a means of comparing the performance of different parts of the rural strategic network, not previously possible. Along with assessing average speeds for several months, the proof of concept also mapped the reliability of these key routes (omitted here due to brevity).

In collating this data, it was identified that the sample size varied from location to location. In some cases, thousands of trips were available, and in other cases tens of thousands.
Figure 1 Rural Strategic State Highway Average Travel Speeds (June 2013)
The Austroads KPI include productivity
In parallel to the proof of concept for rural strategic roads, the Agency has also undertaken a proof of concept assessing the urban network (including some key non-state highway corridors). The Austroads Productivity results derived from the biannual floating vehicle surveys were compared to results derived from almost 2 million commercial vehicle trips in New Zealand's three largest urban areas.

The following figure provides an example of network level results for the Austroads Productivity formula (essentially a utilisation of road network comparing observed throughput and flow against benchmark targets). The histographs show the proportion of network operating within Austroads Productivity bands, although it should be noted that the Wellington network includes a significant portion of rural state highway, as far north as Levin. The histographs compare – at a network level – the results derived using floating vehicle data compared with the same report derived using commercial GPS data.

As commercial vehicles are often heavier and slower than private vehicles, it was initially expected that differences in the vehicle fleet would lead to significant differences between the two methodologies. However, at a network level, the Productivity results were relatively similar, presumably because both sets of vehicles were observing (and being delayed by) similar levels of congestion.

As expected, there is a slight decrease in observed Productivity using the commercial GPS data, but the difference was less than anticipated. However, the caveat is that there were differences observed at an individual corridor level. The suitability for urban corridor level reporting is still being explored.

Understanding Traveller Behaviour
A benefit of having access to data derived from actual use of the network is that, with the right data source, it is possible to identify aggregate driver behaviours such as preferred route choice. The following Figure 3, provides an example of an investigation into route choice, using the GPS data source as the rural road trial. The percentages and the width of the lines indicate the origins, destinations, and route choice along the way. Interestingly, the study encountered a high reliance on the state highway network, with very few vehicles leaving the state highway network, except when travelling on the Route K toll road or the toll free alternative. Based on these observed results, it can be inferred that commercial vehicles using the SH29 Welcome Bay Roundabout are almost certainly not accessing the ports, but rather, are more likely to be making an east-west trip bypassing urban Tauranga.
Figure 3 Commercial GPS route choice to the Ports of Tauranga

Travel by Time of Day
The GPS data was looked at in more detail to identify travel by time of day for commercial vehicles, in order to understand commercial vehicle usage of the rural road network. The results are in Figure 4 below.

Figure 4 Travel by Time of Day

If the temporal coverage of the commercial vehicle data was distributed evenly across the day (i.e. if commercial vehicle travel was entirely random) each hour would result in roughly 4% of trips (although in practice this would never happen, because demand is never random). Figure 4 identifies the non-random distribution for commercial vehicle travel using the GPS sample as a means of understanding demand on the network. The result indicates a weighting toward AM and interpeak travel, but not PM travel.
The results also indicate that a third of the trips identified from the Commercial GPS data happened overnight (between 7pm and 7am). This is a significant portion of the total trips and a far greater proportion than would be expected from ordinary vehicle travelled – which might be identified via (unclassified) AADT volume profiles. This finding provides quantifiable results to confirm anecdotal understanding that overnight trips are a significant feature of commercial activity.

The overrepresentation of AM trips may be a result of an overlap of trip purpose, e.g. some of the overnight trips finishing during the AM peak, and some of interpeak trips starting in the morning.

Understanding Route Resilience
From an absence of data, to having visibility of the performance of rural strategic corridors, the use of commercial GPS data provides a step change. The proof of concept identified that sufficient data is being collected on almost 4,000km of road network, sufficient for the purposes of reporting monthly average speeds. However, it is one thing to be able to report on average speeds, and another to be able to to identify when roads are closed due to crashes, floods, slips, or other reasons. Further analysis was undertaken to look at the relationship between spatial and temporal coverage. For the purposes of this analysis, we required a minimum of 2 observations per hour, or else it was treated as a ‘gap’ in the data.

Figure 5 shows that the results (with the current sample source and rate) identified that the GPS data was nearly continuous (during weekday, 7am-7pm hours) for just over 1,400km of the 4,000km road network studied, i.e. for around a third of the studied network, an absence of data during weekdays would indicate that the road was closed, rather than indicating a gap in the sample rate.

GPS vs Bluetooth Matching
Commercial GPS data is just one possible way to understand the performance of New Zealand’s long road network. Another way to fill the gaps in our understanding of the state highway network is to identify and match vehicles by deploying sophisticated traffic sensors. The Agency now has a network of Bluetooth matching sensors spanning more than 600km of state highway. This system relies on detecting the unique Bluetooth identifier emitted by mobile devices within a vehicle as it passes a detector and matching that device at multiple detectors to build up travel times and routes. The Transport Agency’s national system (BlipTrack) is fully encrypted and meets exceeds all NZ Government privacy requirements.

The current size of the Bluetooth network means that the coverage was insufficient for the desired 4,000km proof of concept. However, the existing Bluetooth network does provide a useful comparison with the Commercial GPS in order to inform future investment decisions.

A comparison of the two data sets found that the sample rate of the Agency’s Bluetooth matching system provides roughly 20 times more trips than were observed during the Agency’s commercial vehicle proof of concept study.
However, this difference dropped in areas with significant commercial activities, such as near ports.

To put this into perspective, the existing 600km Bluetooth network owned by the Agency is observing more than 2 million (anonymous) trips every month.

The commercial GPS sample rate was compared to Bluetooth on 26 overlapping corridors. The results are shown in Figure 6 below.

![Figure 6 Bluetooth vs GPS sample comparison](image)

The sample rate for the commercial GPS varied when compared to Bluetooth, with the higher GPS sample appearing to reflect the significance of the location to commercial vehicle movements, rather than the volume of the location itself (e.g. an increase near ports), i.e. the sample rate for GPS could be bias toward the function/usage of the road, rather than the net traffic volume.

Investigation into the Bluetooth sample rate has found that the Agency’s Bluetooth matching system is relatively consistent in detecting the equivalent of roughly 15% of all traffic regardless of expected changes in vehicle composition (this 15% value includes calculations for filtering out multiple Bluetooth devices carried in a vehicle – the raw comparison is around 17%).

This means that with the Bluetooth system it is possible to identify gaps in the data of less than an hour duration, as the supply of data is continuous even in moderate vehicle flows. It is also possible to study changes to congestion within a single peak for a single day. Figure 7 below provides an example, illustrating how the Bluetooth system in Wellington captured delays following earthquake disruption in August 2013, as people tried to leave the central city en masse. The Friday PM peak for previous weeks is compared against the profiles for previous weeks. Note the unusual delay during an early peak, followed by further delay in a second peak for those who left work at the usual time. Both the peak delay and the peak spread were far greater than usual.
Figure 7 Wellington commuter delays post-earthquake

With a 15% sample rate, it is possible to identify changes to travel time and delay, and even changes to delay within a single peak period. The results for Figure 7 show a clear change to the peak delay profile, even though the total traffic volumes (demand) for the day did not change substantially compared to a usual Friday i.e. the overall number of Bluetooth matches remained the same, at just under 2,000 Bluetooth observations. The Bluetooth system is therefore more useful than the GPS system for real time monitoring and route reliability and is now being trialled for real-time alerts and customer information.

The comparison suggests that while it is possible to identify road closures, resilience and reliability with the commercial GPS data for 1,400km of New Zealand road network, the Agency’s Bluetooth matching technology is probably more suitable, where the infrastructure is available.

A Brief Comparison of Data Sources

An advantage of commercial GPS over floating vehicle surveys is that – along with the significant increase in sample size - the data continues to be collected by the third party providers throughout the year. This provides the potential for continuous monitoring, or for assessing the impact of holiday peak loading and seasonal variances, without the need for additional data collection.

However, while this means commercial GPS data can be studied forensically throughout the year, it does not necessarily follow that it can also be used for real time monitoring. The requirement to procure the data from a third party can introduce delays.

These delays are not present for infrastructure owned directly by the RCA – e.g. loops, radar and Bluetooth matching infrastructure, etc.

One key difference is that commercial GPS data requires procurement from third parties, whereas the Bluetooth matching requires capital investment in infrastructure. Without this infrastructure it is impossible to ‘go back in time’ to collect the data. Although, once installed, significant volumes of data can be collected in real time and in a very cost effective manner.

Collecting sufficient real time data provides opportunities for road operators to move from a reactive network monitoring space into a real time operational environment. In the past, real time operational decisions have only been possible for some parts of the major urban centres restricted to locations where point speed and CCTV cameras provide some visibility of the performance of the network. However, expanding loops and CCTV coverage is a very expensive exercise, and therefore not practical for coverage of the rural strategic network. Storage and archival also creates a significant challenge.
The data being collected by the Agency’s Bluetooth system and GPS proof of concept are both sufficiently large to require cloud-based storage. However, compared to video footage, which requires terabytes of storage capacity, both are more efficient means of storing data for study of the network. Typically, CCTV footage is erased after only a few days, as the storage requirements and privacy concerns are onerous. This makes it suitable for real time operations (seeing the network perform in real time) but unsuitable for regular reporting of network performance. The costs and difficulties associated with video footage / analysis / analytics / storage are also a large part of the reason that numberplate recognition technologies tend to be prohibitively expensive.

CONCLUSIONS

GPS and Bluetooth matching both provide a cheaper alternative to expanding CCTV or loop based operations. The current key difference between the two technologies is that, while Bluetooth provides a much larger sample rate and real time data, it requires an investment in infrastructure before the data can be collected, whereas Commercial GPS data is already being collected by third parties.

The Commercial GPS obtained for the proof of concept was sufficient to provide an understanding of average travel times for most (although not all) of the 4,000km strategic network included in the study. The sample rate on 1,400km of network was sufficient that daytime road closures could potentially also be identified. This provides the Agency with significant potential to expand understanding of the network beyond key urban corridors to identify customer experience travelling between urban centres and townships.

However, the requirement to procure the data may introduce delays, which means that – unlike the Agency’s existing Bluetooth system – Commercial GPS data will in the short term remain unsuitable for a real time operations unless or until supply of the data from third parties is provided continuously and in real time (which may in turn introduce further privacy issues).

Whilst it could be said that the amount of data being collected for these proof of concept studies is already impressive, it is just the tip of the iceberg. GPS devices are becoming increasingly ubiquitous, particularly as they are now carried in Smartphones. Just a few years ago, the idea of every person carrying a GPS device on their person at all times might have seemed like science fiction. If it was possible to build a dataset using all of the GPS devices in the country, the number of GPS points generated per month might run into the billions.

The concept of the “driverless car” was also once restricted to the realm of science fiction, but is now being developed by several major companies. It may be that in a few years’ time the amount of GPS data being generated by vehicles today will be dwarfed by the data generated by an increasingly sophisticated, technology driven vehicle fleet.

In the short term, the findings of the proof of concept suggest that existing road users are already creating more than enough GPS data to fill in some of the big gaps in understanding the performance of major corridors between our urban centres. However, whether there is also sufficient data for less heavily trafficked parts of the rural state highway network remains to be seen, as this fell outside the remit of the proof of concept study. However, combining GPS data sources will improve the sample rate, and therefore provides a likely means of increasing the proportion of roads with a sufficiently large sample for analysis. However, this introduces additional challenges around consistency of data formats and network/ link-models.

In the long term, New Zealand therefore has a twofold challenge: firstly, to source GPS data from the companies that are already collecting it, in order to fill in gaps in our understanding of the network. And then secondly, to procure, collate, store, standardise and convert it into information useful for RCA engineers and planners. This second challenge will grow as increasing amounts of GPS data becomes available.
REFERENCES

i Compiled from the CIA World Factbook

Land area ranking

ii Compiled from the CIA World Factbook

Road Assets:

Population:

GDP per Capita: