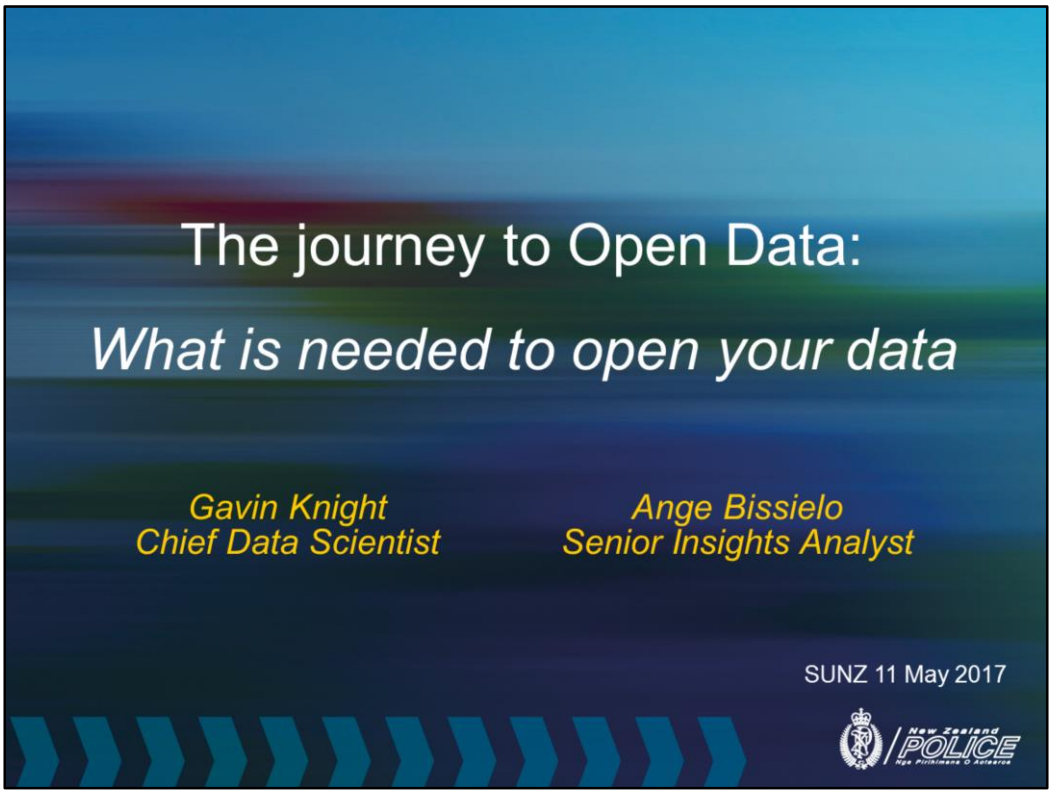
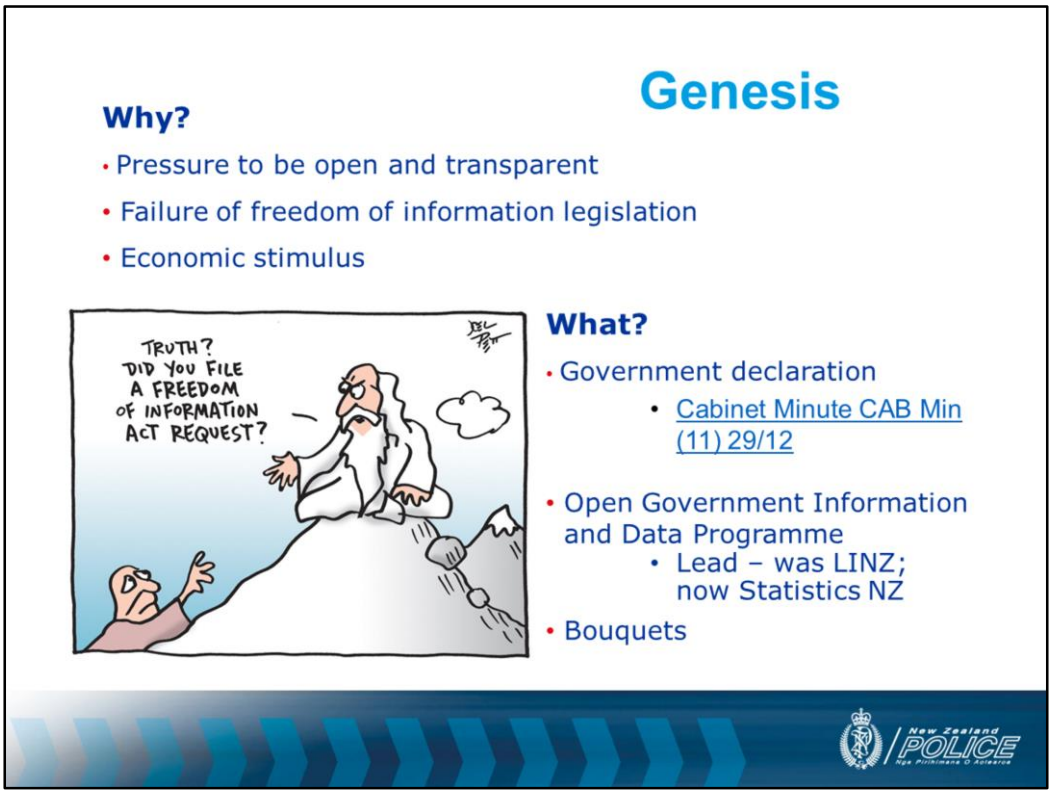



Most of us will have heard of Open Data. Many of us are working to implement it.

New Zealand Police is well progressed on this journey. We have learned some things along the way that others who are also contemplating opening their data might like to know.

So in the few minutes available to us today, we would like to share some of the key points about our journey so far.

I will outline this journey, and my colleague Ange Bissielo, will give a quick demonstration of how we are using SAS Visual Analytics to provide public access to Police data.




Following frustration experienced by citizens attempting to get information out of their government, numerous governments around the world made public commitments to be more open and transparent.

In particular, it was felt that freedom of information legislation established to promote government transparency was not working. So, in the face of this public dissatisfaction, governments directed public sector agencies proactively publish their data rather than wait to receive requests under freedom of information legislation. Some governments, including New Zealand's also recognised that data has economic value. They saw the potential to stimulate economic growth through opening data for reuse by third parties.

In New Zealand, on 8 August 2011, Cabinet directed government agencies to commit to releasing high value public data actively for reuse.

LINZ was given the mandate to drive the opening of public data. They conducted training sessions for officials from other agencies, produced reports on the progress being made by agencies, and consulted with agencies with the aim of helping to stimulate progress in opening data.

I would like to acknowledge two previous LINZ employees in particular for their fantastic vision, energy and support to helping New Zealand Police and, I'm sure, numerous other agencies to make progress in opening their data. These are Keitha Booth who, perhaps more than anyone else in New Zealand, got the ball rolling, and her protégé Paul Stone, who then took over the reigns when Keitha retired, and who has been working behind the scenes with agencies to help us overcome problems by offering advice and connecting people in different agencies who are doing things that might help others.

## Strategic decision

- Business Intelligence Statement of Direction
- Proactively embrace - Open by default
- New BI operating model
- New technology
- Birth of the data scientist in Police
  - Drivers:
    - Compliance
    - Economic stimulus
    - Transparency
    - Growth in OIA requests
      - Efficiency & Risk

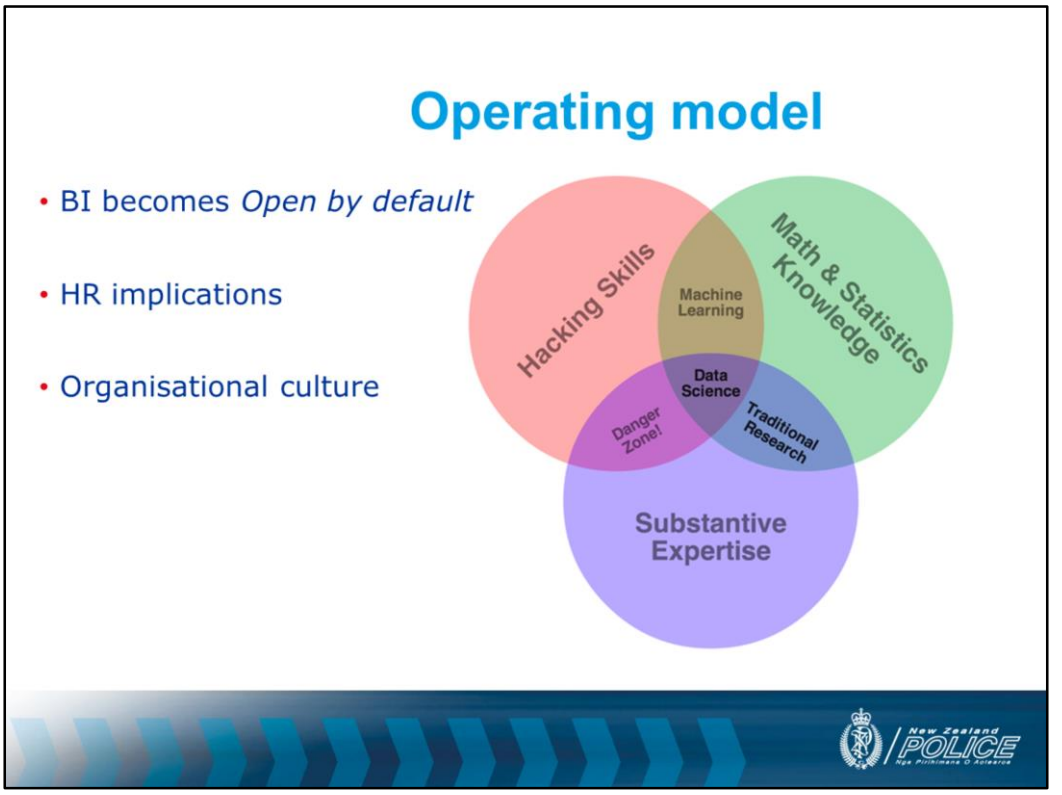

In April 2013, following approaches from Keitha Booth, the New Zealand Police, as part of its business intelligence strategy, decided to proactively embrace Big Data and Open data.

Rather than be reluctant followers – forced by cabinet directive to release some data – we would enter into the spirit of the government declaration and proactively transform our business intelligence ethos and operating model, to develop datasets not just for internal operational purposes, but also with the expectation that they would be published. We called this 'Open by Default'.

In 2015 we restructured our organisation, transformed our operating model, and changed our core BI technology to enable this to occur.

The impetus behind Police deciding to proactively embrace Open Data, involved two predominant factors: First, for Police to be effective it is important to maintain public trust and confidence, so the transparency aspect to the Open Data strategy resonated strongly with Police. However, perhaps even more pressing, we were experiencing massive growth in demand for data (often through requests made under the official information act), placing pressure on the organisation to service this demand. This also increased the risks around releasing data.

I will expand on these risks shortly, but first, I'll briefly touch on the primary change to our BI operating model – the injection of data scientists into the process.




To be open by default we want to design our data sets and reporting so they can be consumed by our staff in the same way we can reasonably expect is possible for the general public (not just tech-savy data experts).

To achieve this we now have an SQL-free user environment, where users access data in pre-counted flat file tables that are easy to describe and understand without having to have technical skills. In doing so it has become easier for us to replicate our internal reporting in the public domain. The only differences are in the level of detail available to the public. In particular, personal data must be anonymised and to some extent confidentialised.

We introduced data science not just for modelling and analysis of data, but into the design of our data integration and ETLs in the data warehouse function. Design of data sets in the new NZ Police data warehouse environment requires skills in statistical classification design. We need our people who design BI data sets to understand the relevance of concepts like mutual exclusivity, exhaustiveness, statistical balance, statistical bias, and statistical errors. They also need to understand the operational business processes that generate data in IT systems. Only then will the resultant data be meaningful.

We have also trained the ETL developers who have worked in the old data warehouse environment in how to use SAS Data Integration Studio to stage data into SAS, but the design of the data domains we use to report and publish data is overseen by a data scientist because of the need for statistical thinking skills.

It is easier to train a statistician to produce SAS code – particularly with modern SAS tools like DI Studio – than it is to train an IT developer in the necessary statistical skills.

Not only are there changes for the personnel who develop data sets, there are also changes for BI users. Analysts in Police are used to developing their own database queries and producing their own reports. This approach produces significant variation in output and quality of output, depending on the skills and approach of the analyst. This is difficult to manage, and creates confusion.

In the new model the datasets used for analysis are designed centrally by highly skilled people. To meet the needs of analysts, they must have flexibility to drill into data about what they need to analyse. This approach provides uses with less flexibility than they are used to but sufficient functionality to reach the point where the loss of flexibility is less than the value of the improved quality and consistency in the data, and resultant efficiency in the business.


Operational agencies like Police often need to release data, particularly at a local level, to community groups, so that we can partner with them to achieve shared objectives.

For example a local Neighbourhood Support Group might find it helpful to know how many burglaries have occurred in their neighbourhood recently, and whether local crime levels are on the increase.

Increasingly, data being released by government agencies is being scooped up and published – For example many OIA responses are appearing on the FYI site. This enables others to collate all this data, potentially linking data from different releases to reveal a greater level of detail about those crimes. If the data contains personal information, this carries some risk of enabling re-identification of individuals.

So, we have competing pressures around the release of data. On one hand the intent of the OIA and the Open Data strategy is to release more data. However, privacy risks limit how much can be released. The more variables contained in a data set, the greater potential exists to cross-tabulate and link data sets to identify individuals.

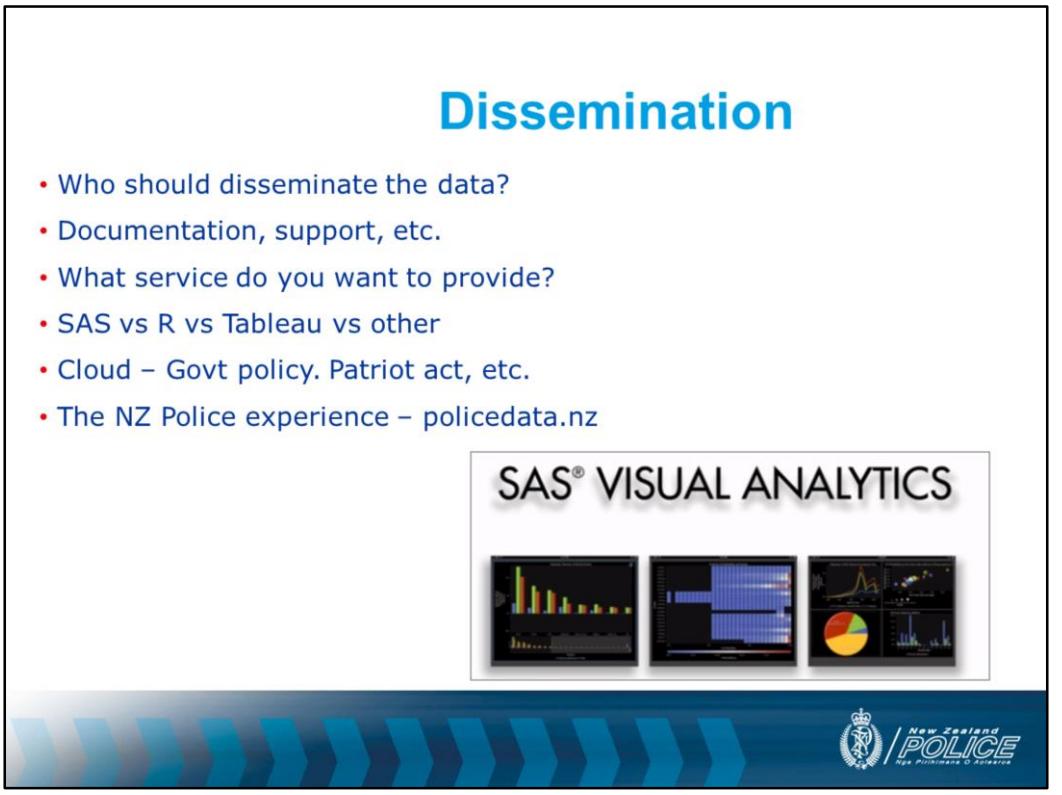
For such multi-dimensional data sets, some traditional data confidentialisation methods used by statisticians – such as random rounding and cell suppression – become less useful, because they are designed to work on data sets with larger counts and less variables. They also create computational performance problems that limit the range of data dissemination tools that can be used. Their application to more complex datasets can limit the utility of the data.

Given these developments, continuing with our historic practice of releasing data at a local level with sufficient detail for operational purposes, carries greater privacy risk than it used to. So, Police has opted to try to proactively publish data sets that are at the threshold of releasing as much detail as we can without creating undue privacy risk, thereby enabling us to refuse requests for more detail on privacy grounds. This strategy enables us to ensure consistency in the data we release, while acting in alignment with the principles of all of the relevant legislation.

However, designing such data sets is a challenging problem. It is mathematically complex to set the right level of detail in all dimensions of a 20 or 30 dimensional data set. For a given level of privacy risk, greater detail in one dimension (e.g. geographical location) means you need to reduce the detail in another dimension (e.g. monthly verses weekly counts). The choices in which way to restrict this detail affects the utility of the data for different purposes. So, we need to develop principles for deciding where to put the restrictions.

Police has worked with both Statistics NZ and the Office of the Privacy Commissioner to explore these issues. It would be fair to say that this work has stimulated a lot of thinking about how privacy should be protected in an Open Data world. There is a growing awareness that we may need to move from data design principles that remove the theoretical possibility of identifying individuals to an approach based on the practical risk of such events actually occurring.

Australia is introducing legislation that puts the onus on users of data not to re-identify people from anonymised data *{Ref. Privacy Amendment (Reidentification Offence) Bill 2016}*. Such an approach has the potential to reduce the effective risk of releasing more detailed data. It is designed to create a freer environment for producers of data to release their data in relative safety. It will be interesting to see whether citizens accept the trade-off of reduced privacy protection in return for being able to get more detailed government data.




As well as these challenging problems, we also have technical options for how to actually release the data.

Should all government data be released by a central agency, rather than have everyone build their own tools? No such centralised mechanism currently exists with all of the functionality required to deliver a good user experience.
So should agencies make do with the functionality that can be provided by central agencies, or try to do something better themselves? If they go it alone, should they disseminate data directly themselves of hire a third party specialist to do it?

Either way, in practice more is required than simply publishing data as prescribed in the Open Data Strategy. Customers can experience problems or have questions, so the producing agency needs to develop an effective service delivery model with customer service, service performance management, and user documentation designed for public consumption.

The Open Data strategy in its current form simply requires that data be made available, ideally in a machine-readable form. However, doing this only puts the data in the hands of the few with the technical expertise to access and process such data. To make data accessible to more people, some easy way for a non-technical person to consume data is required.

So, the producing agency must consider solutions for achieving this – interactive data visualisation tools immediately come to mind. There is a large and growing range of such tools. Many organisations use tools like Tableau, which are easy to use and give a great user experience. Others opt for a more complex but open-source solution such as R. Police at this point in time have opted to use SAS Visual Analytics running in the cloud on Amazon Web Services. I would like to Acknowledge SAS NZ for helping to make this easy for us to do.

Visual Analytics is our core enterprise BI tool that we use internally, so it is not much work for us to simply tweak our internal reports and publish them externally. This approach also enables us to give the public the data in the same form we produce it for our own use, so we are truly being transparent.

I'd now like to hand over to Ange Bissielo, to give you a quick demonstration of a couple of public Visual Analytics reports that he developed.

***{Ange's demo of policedata.nz goes here}***

It would be fair to say that our use of Visual Analytics for public dissemination of Police data has been a learning experience – both for SAS and Police. When we launched our policedata.nz service six months ago, we immediately experienced a raft of technical problems. For example, the firewall settings of some businesses prevented them accessing the policedata.nz site. Others encountered problems because the reports use Flash, which requires that browsers are up to date.

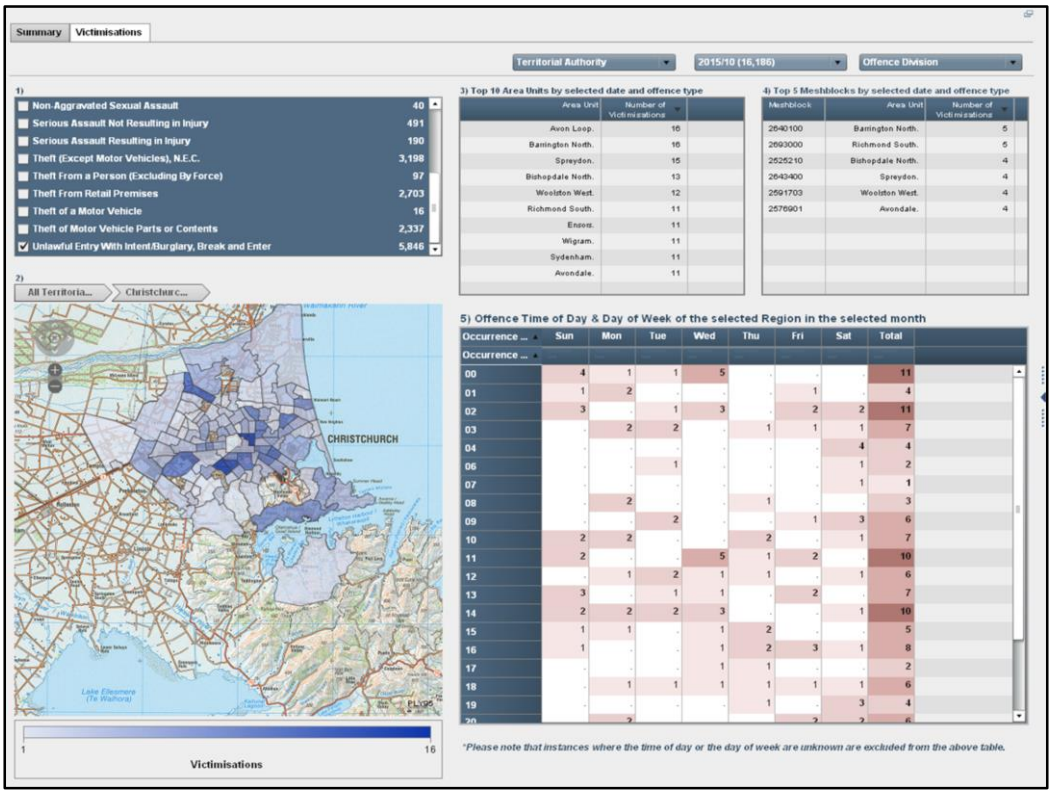
We used Flash because the HTML5 functionality of our current Visual Analytics version does not permit the data download functionality we want to support the Open Data principles. We are anticipating this to be fixed when we upgrade to version 8.2 later his year.

We also encountered a problem with a user hacking the URL to alter the user preferences of the guest user, changing the default language to Russian.

We managed to fix all of these teething problems very quickly, yet it was apparent that different complexities exist in a public environment than exist within an enterprise BI environment. SAS, like all technology providers in the hot space of tools that can provide the dissemination service Police needs, faces challenges. I envisage the development of low cost, highly functional and secure solutions.

However, SAS is a good partner to work with on the challenges Police has faced. Police is currently working with SAS to find the best way to address these challenges using Visual Analytics. We hope to be able to continue and improve on what we have started. It will be interesting to see where we are at in six months' time.


Summary
Offender Demography
Offenders by Iwi
Time Series
Repeat Offenders
Offender List
Unique Offender List
2014/11
2015/10
2015/07
2015/10
Person (173,198)
Ethnicity
Bay Of Plenty (6,172)
Police Area
Police Station
Age Group
5
80
5
80
1
Homicide and Related Offences
Acts Intended to Cause Injury
Sexual Assault and Related Offences
Dangerous or Negligent Acts Endangering Persons
Abduction, Harassment and Other Related Offences Agai...
Robbery, Extortion and Related Offences
Unlawful Entry With Intent Burglary, Break and Enter
Theft and Related Offences
Fraud, Deception and Related Offences
Illicit Drug Offences
Prohibited and Regulated Weapons and Explosives Offen...
Property Damage and Environmental Pollution
Public Order Offences
Traffic and Vehicle Regulatory Offences
Offences Against Justice Procedures, Govt Sec and Govt...
Miscellaneous Offences
0
200
400
600
800
1,000
2
63.2%
2.0%
2.1%
3.5%
29.1%
Ethnicity
Maori
European
Pacific Island
Unknown
Other
3
Family Member N.F.D. 62
Partner 235
Parent 68
Step Parent 15
Grandparent 4
Child 97
Stepchild 6
Grandchild 6
Sibling 40
Boy/Girlfriend 132
Care Giver/Receiver - Relative 1
Ex-Partner 50
Ex-Boy/Girlfriend 180
4
25
20
15
10
5
0
Offender Age
Sex
Female
Male
89.9%
1.4%
8.7%
Method of Proceeding Division
Court Action
Non-Court Action
Not Proceeded With


Summary
Victimisations
Territorial Authority
2015/10 (16,186)
Offence Division
1)
Non-Aggravated Sexual Assault 40
Serious Assault Not Resulting in Injury 491
Serious Assault Resulting in Injury 190
Theft (Except Motor Vehicles), N.E.C. 3,198
Theft From a Person (Excluding By Force) 97
Theft From Retail Premises 2,703
Theft of a Motor Vehicle 16
Theft of Motor Vehicle Parts or Contents 2,337
Unlawful Entry With Intent/Burglary, Break and Enter 5,846
2)
All Territoria...
Christchurc...
CHRISTCHURCH
1
16
Victimisations
3) Top 10 Area Units by selected date and offence type
Area Unit
Number of Victimisations
Avon Loop. 16
Barrington North. 16
Spreydon. 15
Bishopdale North. 13
Woolston West. 12
Richmond South. 11
Enoos. 11
Wigram. 11
Sydenham. 11
Avondale. 11
4) Top 5 Meshblocks by selected date and offence type
Meshblock
Area Unit
Number of Victimisations
2640100 Barrington North. 5
2693000 Richmond South. 5
2525210 Bishopdale North. 4
2640400 Spreydon. 4
2591703 Woolston West. 4
2576901 Avondale. 4
5) Offence Time of Day & Day of Week of the selected Region in the selected month
Occurrence ...
Occurrence ...
Sun
Mon
Tue
Wed
Thu
Fri
Sat
Total
00 4 1 1 5 11
01 1 2 1 4
02 3 1 3 2 2 11
03 2 2 1 1 1 7
04 4 4
06 1 1 2
07 1 1
08 2 1 3
09 2 1 3 6
10 2 2 2 1 7
11 2 5 1 2 10
12 1 2 1 1 1 6
13 3 1 1 2 7
14 2 2 2 3 1 10
15 1 1 1 2 5
16 1 1 2 3 1 8
17 1 1 2
18 1 1 1 1 1 1 6
19 1 3 4
*Please note that instances where the time of day or the day of week are unknown are excluded from the above table.

## Future possibilities

- Social licence
- The death of privacy as we have known it?
- Legislation change
- Open data strategy refresh?
- Centralised delivery or centralised funding models?
- The death of Official Statistics?


Having initially launched only a limited amount of crime data via policedata.nz, we are now developing plans to roll out more Police data over the coming months.

As we move forward and start hitting the boundaries of the right level of detail, we will take into account the expectations citizens have of the level of detail that should be published.

Mark Zuckerberg has talks of the acceptance people have that smart Data Science applications will collect and use data about them – for example, Google deciding which advertisement to present in your browser, You Tube suggesting which channels you might be interested in, Facebook suggesting who you might want to be friends with. Zuckerberg suggests this acceptance signals then end of privacy. If he is even partially right, then the social licence government has to release more detailed data might increase.

If legislation such as Australia's Privacy Amendment Bill is implemented, of if the Official Information Act, the Privacy Act, or Statistics Act are amended, such changes could also impact the principles guiding the release of public data.
The Open Government & Information and Data Strategy when next refreshed may go beyond simply opening data, to making data accessible through usability features.

As agencies continue to open their data, a centralised dissemination service, or at least a centralised funding model, may emerge to make it easier for agencies to release their data.

Irrespective of the dissemination mechanism or mechanisms, one question all of this leaves us with is: What might an open-by-default release of detailed operational data produced by government agencies do to the traditional framework for official statistics?