Speaking Notes

|  |  |
| --- | --- |
| **Event title:** | Public Sector Network - Data Innovation Executive Session |
| **Speaker:** | Sven Bluemmel |
| **Event date:** | 23 June 2021 |

Introduction

* My office – the Office of the Victorian Information Commissioner, or OVIC – is the primary regulator for information privacy, information security and freedom of information in Victoria, administering both the Privacy and Data Protection Act 2014 and the Freedom of Information Act 1982.
* One of OVIC’s key strategic goals is to promote fair access to public sector information while ensuring proper use and protection of the information rights of Victorian community.
* I will discuss the benefits of sharing public sector data data, how it can go wrong, how to do it right and balance government transparency and data sharing with maintaining public trust in government.
* While the sharing of public sector data raises obvious privacy risks, privacy law recognises that there will be instances where the collection and sharing of citizens’ information can benefit of the wider community.

Benefits of sharing public sector data

* Whether data is being shared to the public at large or between public sector organisations, it can have direct benefits to the safety, health and wellbeing of members of the public when done properly. For example:
  + Collecting and publishing data about the age, gender, suburb and recent whereabouts of confirmed COVID-19 cases has assisted controlling and suppressing the spread of the virus.
  + The Child Information Sharing Scheme under the *Child Wellbeing and Safety Act* improves the early identification of risk to children’s safety and wellbeing thereby directly benefiting children.
    - The scheme allows public sector organisations and services prescribed by the legislation to share personal and health information about children where such collaboration is necessary to maintain their safety and wellbeing.
  + The Family Violence Information Sharing Scheme under the *Family Violence Protection Act* directly improves the health and safety of adults and children at risk of family violence.
    - The scheme allows prescribed public sector organisations and services to share information related to assessing or managing a family violence risk to prevent family violence from occurring or protect individuals from being subjected to further family violence.
* There are also several ways that sharing public sector datasets can indirectly benefit the public. For example, access to public sector data can:
  + Support more advanced and accurate research, leading to, among other things, more informed public policy development.
  + Improve public administration by identifying workforce capabilities and data governance gaps that therefore promotes innovation to address gaps and ultimately improve productivity and customer service outcomes.
  + Support public sector organisations to further engage in evidence-based decision making practices.
  + Reduce the over-collection and storing of overlapping datasets across public sector organisations, increase transparency of government operations and minimise the risk of data breaches that may affect the public.
  + Stimulate growth in the Victorian economy by identifying industries that need additional resources to function effectively and thereby create jobs and economic opportunities for individuals.

How it can go wrong

* While there are many benefits to sharing public sector data, there are also inherent risks that can have serious consequences.
* One such risk is that of re-identification of de-identified datasets. Given the need to protect individuals’ personal information, the sharing of public sector datasets generally relies on the proper de-identification of personal information.
* However, de-identifying personal information to the point where it is permanent or cannot be re-identified, is especially difficult.
* As improvements in technology increase the type and rate at which data is generated, the possibility of re-identification of publicly released data is greater than ever and it is also very difficult to determine the likelihood of re-identification for any given dataset.
* For example, auxiliary information can be used to connect an individual to seemingly de-identified data, enabling an individual’s identity to be ascertained. Auxiliary information can come from anywhere, including other publicly available sources online.

Kevin Rudd – MBS/PBS data example

* + In 2016, the Commonwealth Department of Health released a de-identified dataset of Medical Billing System and Pharmaceutical Benefits Scheme information.
  + This included two types of data:
    - Group statistics, including a complete list of the frequencies of each medical test/procedure’s billing code for the whole population, separated by patient age ranges of 10 years, state and month for a 30 year period.
    - Longitudinal records that include patients’ gender, year of birth and item code for a 10% sample of the Australian population.
  + Because longitudinal records include more detailed information, they can be easily re-identified when paired with auxiliary information, which can have serious consequences for the privacy of the patient.
  + Group statistics provide less information, but still prove risky where there are unique billing codes among the dataset that are paired with auxiliary information.
  + For example, former Australian Prime Minister Kevin Rudd had an aortic valve replacement in Brisbane in August 2011, which was described in an online news article that gave the exact date of his surgery.
  + In the group statistics, this surgery was unique in that month and age range so researchers analysing the data could identify the exact record that corresponded to Mr Rudd’s record. In this case, the record only provided his age, gender and state in which the surgery was performed, which had already been reported in the article.
  + However, if Mr Rudd’s record had been chosen for the longitudinal sample (and there was a 10% chance it would have been), the re-identification of that record would have implied the retrieval of the entire 30 years’ worth of his medical billing history – information that was not otherwise available online and would have amounted to a significant breach of his privacy.

Myki dataset example

* + For the 2018 Melbourne Datathon (an annual event where Victorian data scientists compete to find innovative uses for public data) the dataset to be examined was a record of public transport trips recorded on the myki ticketing system over a three year period (amounting to travel information of millions of Victorians).
  + The dataset recorded ‘touch on’ and ‘touch off’ events on the myki public transport ticketing system. The data was released as a linked unit level dataset, consisting of records of individual transactions (trips) linked through a numerical identifier assigned to each card.
  + Academics from the University of Melbourne downloaded the dataset from the Datathon’s website (where it was published for Datathon participants) and were able to use the data to identify their own travel movements and the travel movements of other people known to them.
    - Two of the academics had registered their myki cards. This meant they were able to access historical, ‘to the second’ trip data for the previous six months through PTV’s website. They matched this with information in the dataset. To confirm they had found their own records, they cross-checked the registration dates of the cards with the first recorded trips.
    - One academic also successfully re-identified a ‘co-traveller’ who had travelled with one of the academics on a single occasion. They did so by identifying everyone who had touched on or off the relevant tram at about the same time as the academic and narrowed down potential candidates by looking at their travel patterns. To confirm they had re-identified the correct person, they cross-matched the record with some further information obtained from the person (the expiry date of the myki card).
* These two examples show how easily seemingly de-identified data can be made re-identifiable and therefore how critical it is to ensure data sharing is done in a safe and secure way. Otherwise, we risk a reduction in both protection of privacy and public trust in government.

Public trust

* Public sector organisations need to handle personal information in line with their obligations under privacy law, as well as community expectations. They also need to show that they respect and understand the value of this information. In doing so, they build trust and the social licence needed to continue to collect and use personal information.
* Trust in government is essential to the success of public policy. It influences public attitudes and responses to government actions and impacts government’s ability to govern effectively.
* It therefore follows that a high level of public trust is an important element of the public sector’s ability to successfully share data to gain insights whilst also the upholding privacy of the public.

How to get it right: balance data sharing and maintaining public trust

Privacy laws as a mechanism for balance

* There is a misconception that privacy laws are a barrier to the sharing of information, but this is not true. Rather, privacy laws ensure that information is shared in a manner that recognises and protects the privacy rights of individuals.
* Our privacy laws are robust enough to balance competing objectives, such as protecting individual privacy with facilitating the reasonable and necessary access to, and flow of, information.
* The principle-based approach in our privacy laws is designed to cater to the range of information handling practices that public sector organisations need to engage in to deliver public services in a privacy enhancing way. These principles are flexible enough to be adapted and applied over time to new technologies, initiatives, and types of personal information.

Victorian Centre for Data Insights

* The Victorian Centre for Data Insights (VCDI) is a resource that Victorian public sector organisations can leverage to ensure they are sharing data in a safe and secure manner.
* Established by legislation, the VCDI aims to promote the ethical and safe sharing of public sector data and transform the way government uses data so that better policies and services are delivered for the benefit of all Victorians.
* VCDI can assist Victorian public sector organisations to leverage insights from public sector data in several ways:
  + Strategic services, through which it helps these organisations to set out a vision and strategy to become more data-driven.
  + Provide training and support around best practice data management, governance and ethics.
  + Analytics services, through which it provides data analysis support on projects that improve lives or use of resources. It is also important to note that VCDI takes steps to ensure data is de-identified properly before any it is used for analytics.
* VCDI data scientists analyse government data in a closed and controlled lab environment following strict rules around data sharing, security and privacy under the Victorian Data Sharing Act. This then allows the sharing of the results of that analysis to improve and inform service outcomes.
* The VCDI is also accountable for its activities through reporting annually to Victorian privacy regulators on its operations, functions and potential privacy law breaches.

Transparency

* Communication is key when collecting information from individuals in any instance. When public sector organisations are proposing to share public sector data, it is important that they are transparent in their communication with individuals about whom the data relates.
* They should advise individuals about when their data will be shared, including details the purpose of sharing that data, how their personal information will be protected in the process (i.e., what systems/processes have been put in place to protect them).
* It Is also just as critical that public sector organisations are transparent and honest when something does go wrong. Where a data breach or risk of re-identification is likely to cause harm to individuals, it is important to communicate this to those individuals and, in some instances, to the general public.
* It’s also good practice to report data breaches to OVIC and we can provide guidance as required. Victorian public sector organisations also have obligations under the Incident Notification Scheme - established under the Victorian Protective Data Security Standards - to report certain information security incidents to OVIC.

Strong privacy and security frameworks

* Finally, strong privacy and security frameworks are essential to successfully assuring the public that their information will be protected, particularly where it is to be involved in part of a publicly available dataset.
* This means public sector organisations should:
  + Have privacy and security protections entrenched in the design of data sharing systems;
  + Conduct independent privacy and security impact assessments;
  + Provide clear and transparent privacy policies to the public; and
  + Put in place strict parameters around the collection, use and sharing of their information.
* An example of an internationally recognised privacy and security framework for managing the risks of data sharing is the Five Safes Framework. It is a multi-dimensional approach to managing disclosure risk that focuses on the concepts of safe people, safe projects, safe settings, safe data and save outputs.
* OVIC has a number of frameworks and resources designed to enable Victorian public sector organisations to uphold the privacy of individuals and encourage accountability and trust in government.
  + For example, OVIC’s Privacy Management Framework provides guidance on the policies and procedures that promote good privacy practices within public sector organisations.
  + OVIC’s privacy impact assessment template and guide support public sector organisations to identify, assess and address any potential privacy impacts of their products and services.
  + The Victorian Protective Data Security Standards (VPDSS) establish 12 high level mandatory requirements to protect public sector information across all security areas including governance, information, personnel, Information Communications Technology (ICT) and physical security.
  + And the Victorian Protective Data Security Framework (VPDSF) provides direction to Victorian public sector (VPS) agencies or bodies on their data security obligations and build security risk management capability and maturity through the use of existing risk management principles and guidelines.

Thank you.