Ten questions you should ask before sharing data about your customers

October 9 2017, by Christine O'keefe

Credit: cottonbro studio from Pexels

In 2016, a group of University of Melbourne researchers managed to decrypt some data that should have been anonymous.

Using publicly available information, the team pulled service provider numbers out of a sample of Pharmaceutical Benefits and Medicare Benefits Schedule data published online by the Australian government.

Needless to say, people were worried. But while the official response was swift, the exercise showed the potential vulnerability of some datasets that have ostensibly been anonymised to protect privacy.

Still, there are many reasons why it might be useful to share or release data.

A government health department may choose to make data available for medical research. A supermarket may share customer data with a local petrol station to launch a loyalty scheme.

When data is shared, de-identification can provide one way to do it while protecting privacy. That is, transforming data so that the risk of re-identifying an individual or revealing personal information about someone is low.

But de-identification is a complex process. Along with the Office of the Australian Information Commissioner, CSIRO Data61 has developed a De-Identification Decision-Making Framework to help data holders identify, evaluate and manage the relevant risks.

One potential solution

Any government, business or organisation that handles information about people – whether purchases or preferences, location, phone numbers, social media activity, or health services access, for example – needs to think about de-identification.

The technical heart of de-identification typically involves selecting an

appropriate data sharing mechanism (such as open data or secure transfer to a single partner). It usually also involves modifying the data so there is a lower risk of re-identification.

Modifications could include removing names, addresses and other identifiers. It could also include removing or reducing detail in sensitive variables, or adding a small amount of random "noise" to obscure the true values.

How should de-identification be carried out?

De-identification is about risk management, because producing safe, *useful* data means that zero privacy risk is not realistic. Instead, a balance should be found.

Our guide provides a comprehensive look at the issue, but the following ten questions are a place to begin.

1. **What do you know?** Understand the nature of your data, as well as the other data, people, infrastructure and governance associated with your data.
2. **What are your legal responsibilities?** Know which laws apply to your dataset and what obligations they impose. These may include the Privacy Act among others.
3. **What is your data like?** Focus on the data type, features and properties. This involves the data subjects, variables, quality and age. This is important in assessing the re-identification risk.
4. **What is the use case?** Know why you want to share your data, which groups will access them, and how those groups might want to use them. This is important in selecting the appropriate data sharing mechanism and modifications like adding a small amount of random "noise".
5. **What are your ethical obligations?** Consider, for example,

consent, transparency, stakeholder engagement and governance.

6. **What processes will you need to go through to assess disclosure risk?** Establish plausible attack scenarios using risk assessment methods. For example, someone trying to re-identify their neighbour in a local council dataset using characteristics they can easily observe, such as size of family, number of cars, and whether the home has reverse-cycle air-conditioning.
7. **What are the relevant disclosure control processes?** This includes selecting the appropriate data sharing mechanism (such as open data or secure transfer to a single partner) and appropriate data modification methods, including possibly reducing the amount of data under consideration.
8. **Who are your stakeholders and how will you communicate with them?** Stakeholders could include data subjects, the general public, partner organisations, the media, funders and special interest groups. Trust and credibility must be built.
9. **What happens next, once you have shared or released the data?** This includes keeping a register of all the data you have shared or released. It's being aware of developments such as new data-sharing technologies, changes in the law (like the Notifiable Data Breaches scheme coming into effect in 2018) and keeping track of future related data releases.
10. **What will you do if things go wrong?** Have a plan to respond to a disclosure in the event one were to occur. Such measures include having a robust audit trail, a crisis management policy and adequately trained staff.

The De-Identification Decision-Making Framework is not intended to eliminate the need to "call in the experts". Indeed, expert advice - particularly on the more technical aspects of de-identification - may be crucial.

However, these ten questions will help to start the conversation about

what is involved in the de-identification process, and how to begin identifying, evaluating and managing the risks.

This article was originally published on The Conversation. Read the original article.

Provided by The Conversation

Citation: Ten questions you should ask before sharing data about your customers (2017, October 9) retrieved 18 April 2024 from https://phys.org/news/2017-10-ten-customers.html