Reference: adapted from https://app.xtensio.com/.

Improved privacy for aggregated data sets

The problem

Top 3 problems

As a government agency wanting to meet their obligations by sharing data with either researchers and the public I need to know whether it is safe, in terms of privacy, to do so.

As a government agency that already has data available, I want to know why it is safe and if not, how do I make it safe, and about the right balance between providing what the users want and the safety that is necessary.

As a business owner, I'm not sure about what kind of metrics I should be using or how I should be measuring them on my data e.g. how to measure privacy, risk, utility etc. Hence I need privacy assurance and integration of data.

As a user, I need data to be available for research.

A perceived dilemma is that privacy safety and utility of data cannot be achieved at the same time and the traditional process of assurance is manual and hence slow.

Existing alternatives

Statistics Netherlands developed their own variant of this called tau-ARGUS.

Our solution

We will determine what sections of datasets are sensitive and need protection from what sections don't. We then anonymise the data by simplifying the sensitive sections so that individuals don't stand out and are not identifiable. This whole process is executed using algorithms in R (an open source software). The solution will initially only be used for social and demographic environments, however it is easily scalable to business scenarios and for reaching to a wider audience. After several rounds of validation the IP will be used to develop an "R Shiny" app.

Key metrics

Safety

Utility

Number of organisations that release datasets usir

Number of datasets released with no personally id

Value proposition

We will be addressing the pain point of safe data sharing & usage with the flexible, automated and quantified solution that is ready to be used. Hence, in doing so, we will be capitalising on Stats NZ's IP, connections and solid reputation.

High level concept

We have a method to maintain confidentiality and value when bringing together identifiable data sets.

Advantage

This solution is the only one in the market/NZ and the IP is ready for usage and has been internationally reviewed.

We have connections/stewardship with other agencies.

Stats NZ is a big data owner and already has a reputation built surrounding privacy. Stats NZ would like to recover some costs based on the build of the IP. Promoting privacy preserving data sharing is the first step for Stats NZ to promote the capacity building for organisations.

Also from aggregation, the rate of drop in data quality is not as high as existing methods.

Channels

Association meet-up groups

Events/conferences or training around privacy/open data

Connections with other agencies

Stats NZ (stakeholders)

Open Government Data Programme

Social Media

Customer segments

Agencies wanting to know more about how to fulfil their government obligations.

Researchers & academia wanting to know that when they release their findings their output is completely confidential.

Business wanting to share data within units and create synergy with other organisations to better profit on the strategic asset.

Data owners wanting to open up the possbility of creating different use cases or business models based on what they own.

Early adopters

Agencies

Businesses

Scoring

Complexity: 3 - Based on the understanding of what we know in the IP, there's still an element of understanding the definitions of the risks behind certain fields/data.

Risk: 4 - There are ways to mitigate the risk around this venture. Risk metrics are calculated using a well-defined, automated algorithm that has been tested internationally. The anonymising algorithms have been used and tested, however there are risks around being held accountable for anonymization. Hence, for the time being we will focus on indicating the existence of identifiable information in data.

Effort: 3 - There are examples of this type of product/UI around data handling being done before. We would look to productise the solution using R Shiny.

Acquisition: 2 - Stats NZ already has the product in use and working with other government agencies.

Value: 7 - People will see the industry and government doing something around protecting their privacy. This will ideally reduce the amount of risky data being released by improving practises.

Revenue streams

In the early stages, we will provide consulting services to generate quick revenue, and with sufficient use cases, we will encapsulate it into a generic product to sell.