



Review and Analysis of Social Media Usage for Custom Audiences

November 2024

Contents

- Executive Summary..... 4
- 1.0 Overview..... 9
 - 1.1 Scope of the review 9
- 2.0 Understanding the use of custom audiences lists for social media advertising ... 10
 - 2.1 Overview 10
 - 2.2 What are custom audience lists?..... 10
 - 2.3 Uploading data - what does 'hashed data/hashing' mean? 11
 - 2.4 How hashing is used in custom audience lists..... 11
 - 2.5 Conclusions 12
- 3.0 Hashing 12
 - 3.1 Overview 12
 - 3.2 Common Use Cases 12
 - 3.3 Hashing Algorithms..... 13
 - 3.4 Illustration of Hashing 13
 - 3.5 Encoding and Hashing 13
 - 3.6 Known Risks and Controversies 14
 - 3.7 Reducing risks associated with hashing 15
 - 3.8 Published information about hashing usage and risks 15
 - Article 1: Introduction to the Hash Function as a Personal Data Pseudonymisation Technique (EDPS)..... 16
 - Article 2: No, hashing still doesn't make your data anonymous (FTC) 16
 - 3.9 Conclusions 17
- 4.0 Process Flow – How targeted social media advertising happens..... 17
 - 4.1 Process flows 17
 - 4.2 Meta 18
 - 4.3 LinkedIn 18
 - 4.4 Google 19
- 5.0 Additional Security Controls 19
 - 5.1 Overview 19
 - 5.2 Conclusions 20
- 6.0 Information Security Review 20
- 7.0 Assurance from Social Media Platforms 23
 - 7.1 Assurance Information 23
- 8.0 Key findings..... 24
 - 8.1 Use of custom audience lists and privacy 25
 - 8.1.2 Historic use of custom audience lists on additional platforms..... 26
 - 8.2 Providing custom audience lists to social media platforms 26

8.3 Data security and retention..... 26

8.4 Using data from custom audience lists to enhance their own user profiles.. 27

8.5 Conclusions 27

8.6 Recommendations 27

9.0 Unintended disclosures 27

9.1 LinkedIn 27

9.2 Meta 29

9.3 Recommendation 31

Appendix 1: Key Public Communications 32

Appendix 2: Glossary..... 33

Appendix 3: Terms and Conditions w/ Social Media 34

Appendix 4: Statistics on targeted advertising 36

Executive Summary

This report outlines the findings of Inland Revenue's internal review into the use of taxpayer information for targeted advertising on social media platforms.

Background

In the course of its work, Inland Revenue is required to contact customers for a variety of reasons which supports the integrity of the tax system. Under the Tax Administration Act the Commissioner is charged with the care and management of the tax system and in particular, the Commissioner is required to have regard to the importance of promoting compliance including voluntary compliance. To support this, Inland Revenue undertakes a wide range of marketing activities helping customers know about available support, new products or when they may have a return or debt due. This helps to ensure as many taxpayers as possible can meet their obligations or claim their entitlements. Inland Revenue uses a variety of channels for marketing including billboards, digital advertising, videos, radio and social media.

In an increasingly digital world, social media has grown in relevance as a major advertising channel. Inland Revenue has been actively using social media channels for targeted and non-targeted advertising, including custom audience lists, for over ten years. It has been an effective way to help customers meet their obligations and access their entitlements.

Advertising campaigns on social media are mostly carried out through Meta (specifically Facebook), Google (including YouTube) and LinkedIn (Inland Revenue has two LinkedIn accounts – one for Tax Professionals, the other general Inland Revenue followers). Campaigns can also be managed by an external advertising agency.

Inland Revenue opened its first social media account in 2013, and first trialled targeted social media advertising using custom audience lists in 2014. These are lists of specific customers for who the information Inland Revenue wants to share is relevant, for example they may have entitlements they can access or debt they need to pay. Custom audience lists include a range of data that will help identify those relevant people that may have an account with the social media platform – for example first name, last name, date of birth and email. However, the specific information on each list will depend on the platform and how best to target advertisements to individuals using the platform. No financial or tax specific information is used.

The typical process for sharing custom audience lists with platforms is for the list to be securely uploaded and stored by the platform after a procedure called hashing takes place, which deidentifies customer information. This means an individual is not able to be identified by the platform from the hashed data. The platform then carries out matching against its own data which is also in hashed form. This is carried out in an automated process within the platform. The matched data becomes the audience lists which the advertising will reach.

Context for the review

On Monday 9 September 2024, RNZ published an article about Inland Revenue's use of taxpayer information for targeted advertising on social media platforms. This generated public concern and media attention about the privacy practices Inland Revenue uses to generate custom audience lists and share them for targeted social media advertising. Coverage was specifically on the use of deidentification tools (called hashing) and the implications of this for protecting customers' personal information.

Concerns fell into three main categories:

1. Taxpayers are required to provide personal data for tax and social administration purposes and were concerned that they had no control over how their information might be used.
2. Taxpayers being unable to opt out of having their details provided to social media companies.
3. The security controls used (hashing) do not sufficiently de-identify people. This concern was highlighted by reference to a press release from the United States Federal Commission and European Regulators, sharing this concern.

Review topic and scope

In response to public concerns, Inland Revenue paused all social media advertising use of custom audience lists and undertook an internal review to consider whether any of the concerns were valid and to ensure that practices used are compliant with the Privacy Act.

The review set out to understand:

- whether Inland Revenue's use of custom audience lists for targeted advertising complies with the Privacy Act 2020.
- the process for providing custom audience lists to social media platforms.
- data security and retention in the platforms.
- whether social media platforms used data from custom audience lists to enhance their own user profiles.

Key findings

The key findings of the review are as follows:

Use of custom audience lists and privacy

The Privacy Act describes personal information as information about an identifiable individual. When hashed data is uploaded to the platforms it is not considered personal information.

This is because it is in hashed form and does not identify anyone. In addition, the hashing of the data is one layer of security, and the data is uploaded into a secure platform where it is matched with the platform's hashed data in an automated, machine-to-machine process. The data is deleted after the matching process is completed and not used in any other way by the platforms.

Inland Revenue's Privacy Policy notifies people that we will use email addresses and mobile phone numbers to send customers reminders about their tax affairs, and we send hashed information to third parties.

The custom audience list information provided is deleted by the platform after a period of time (which varies depending on the platform).

In 2016 Inland Revenue completed a privacy impact assessment regarding the use of custom audience lists on Facebook. The privacy impact for the project was assessed as Medium; several privacy risks were identified but could be adequately mitigated. This assessment was updated in 2024 and the privacy impact of using custom audience lists remains the same.

Providing custom audience lists to social media platforms

The information provided to the platforms is securely uploaded through an Inland Revenue browser. Where data is hashed, this is automatically performed using an algorithm within the browser of the Inland Revenue device uploading the custom audience list. Both the hashing algorithm and transmission are compliant with NZISM specifications. All data is securely stored on the platform.

The Office of the Privacy Commissioner is assessing whether or not Inland Revenue had the appropriate safeguards in place to minimise the privacy risk of the use of the hashing technology as an information sharing tool.

Data security and retention

Inland Revenue requested security assurance reports from the social media platforms used for targeted advertising with custom audience lists. It is satisfied that the information received shows appropriate standards e.g. international standards and the New Zealand Information Security Manual (NZISM), for managing information security were met.

Social media platforms using data from custom audience lists for their own purposes

The social media platforms indicated that custom audience list information is not used to enhance or build profiles of their users. This means that any data that has not been provided to the platforms directly from individuals is not retained or used by the platforms in the future.

Outcome

Having undertaken the review, we believe that the process taken in using custom audience lists in targeted social media marketing is recognised as legitimate both in New Zealand and internationally.

There continues to be ongoing public concerns about the practice of using custom audience lists for social media advertising. We recognise the importance of building and maintaining public trust as a cornerstone of an effective tax and social policy system.

For these reasons, Inland Revenue will be ceasing the use of custom audience lists for the foreseeable future.

The review has highlighted issues can arise when sharing information with a third party. It is essential that information sharing with any external party, when there is a need for troubleshooting, must be performed securely. Any attachments or other forms of data upload must be encrypted prior to leaving Inland Revenue's network. Once the troubleshooting process is completed, it should be ensured the suppliers remove or delete the information. This should be documented as part of the standard operation procedures.

Recommendations

It is recommended to set review dates for Privacy Impact Assessment's (PIA) to see if anything has changed over time. This includes whether Inland Revenue still has the same view of the activity being undertaken and if the privacy impact rating is still the same.

It is recommended that Inland Revenue undertake a full review of the use of social media for marketing and advertising purposes including the use of specific capabilities and products as well as the information sharing required to enable them.

It is recommended that, in the event Inland Revenue determines that social media marketing or advertising capabilities which target individuals is required, the appropriateness of the implementation of an opt in/opt out process within the core tax system is considered.

Unintended Disclosures

During the course of review two unintended disclosures were identified.

Unintended disclosure - LinkedIn

The review identified that there was an exception with some data provided to LinkedIn, where not all uploaded data was hashed. In addition to hashed email address, first name, last name and country was uploaded to the LinkedIn secure platform without being hashed. Company information was also provided. This was uploaded within the secure LinkedIn platform to be matched in an automated process.

While this data was uploaded to be matched in a machine-to-machine environment and there was no human interaction, disclosing a person's name (regardless of how it is processed) to LinkedIn does not comply with the Privacy Act. This is not a notifiable privacy breach as the information disclosed was minor (name and country), not sensitive and not likely to be misused and is not likely to cause serious harm to affected individuals.

It does not need to be reported to the Privacy Commissioner or affected individuals. However, Inland Revenue notified the Privacy Commissioner of this disclosure when it was discovered during the review.

Unintended disclosure - Meta

It was identified that there was one instance of an unhashed custom audience list being shared with Meta support. This took place in early 2024 when Meta experienced problems with matching an uploaded hashed custom audience list, containing the data of 268,000 customers.

A Meta support person requested a raw (unhashed) file to try and solve the issue. They were provided with this file to try and solve the issues relating to matching the data. Meta confirmed that the file would have been deleted once the issue was resolved.

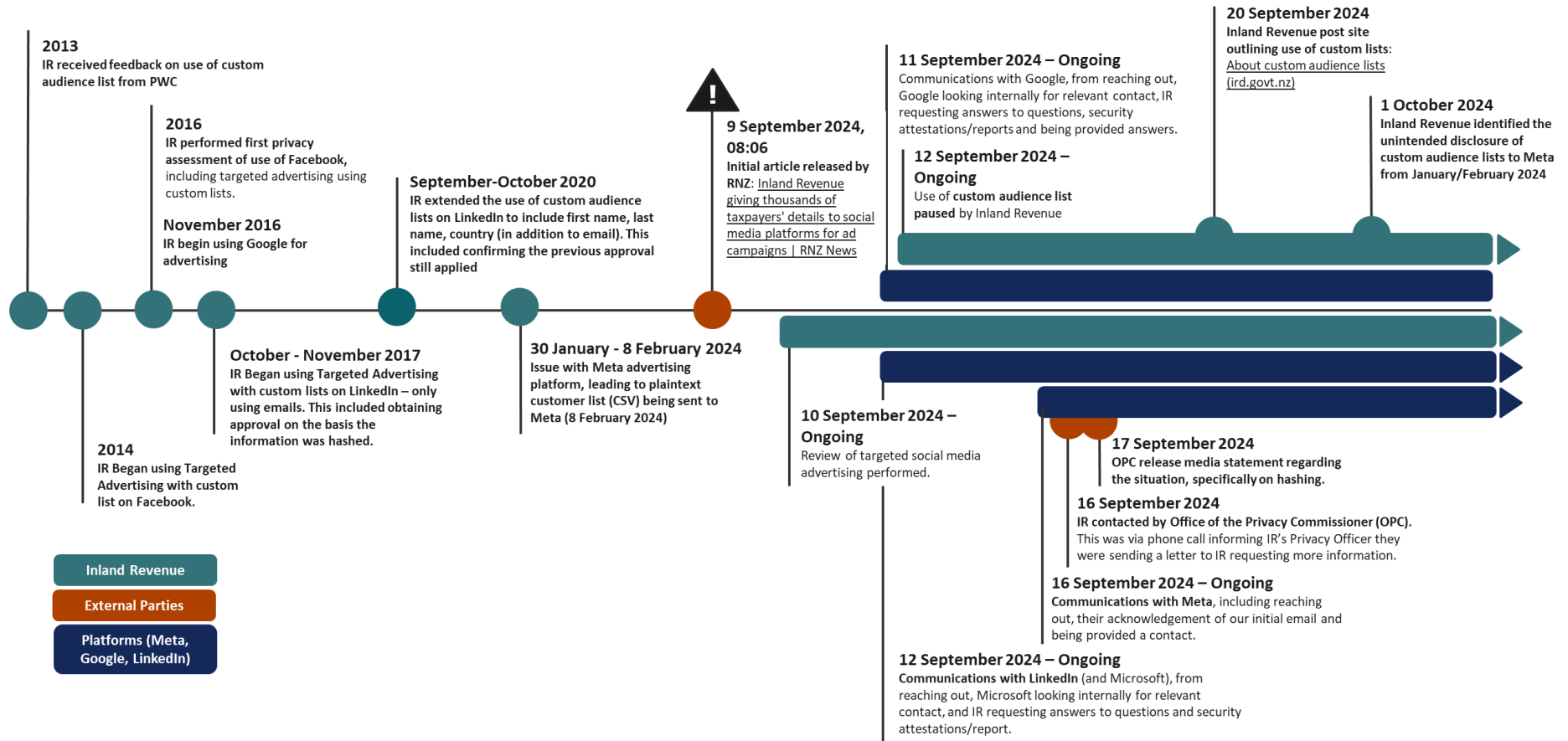
The file was viewed by Meta support for the purposes of trouble shooting, however sharing this file via email was not part of the approved approach for custom audience lists. This is considered a non-notifiable privacy breach due to no serious harm has or is likely to occur as a result of the list being shared with a Meta support person.

Although the breach is not notifiable, when this activity was discovered, Inland Revenue contacted the Office of the Privacy Commissioner to let them know that this unapproved sharing of personal information had occurred. Inland Revenue has chosen to contact everyone who was included in the custom audience list for transparency and to rebuild trust.

Recommendations

The review recommends an internal assurance process be undertaken by Inland Revenue to understand how these breaches happened.

Event Timeline



1.0 Overview

This section outlines the public concerns that prompted an internal review and the scope of the review.

What Happened?

On Monday 9 September 2024, RNZ published an article about Inland Revenue's use of taxpayer information for targeted advertising on social media platforms. This generated public concern and media attention about the privacy practices Inland Revenue uses to generate custom audience lists and share them for targeted social media advertising. Coverage was specifically on the use of deidentification tools (called hashing) and the implications of this for protecting customers' personal information.

Concerns fell into three main categories:

1. Taxpayers are required to provide personal data for tax and social administration purposes and were concerned that they had no control over how their information might be used.
2. Taxpayers being unable to opt out of having their details provided to social media companies.
3. The security controls used (hashing) do not sufficiently de-identify people. This concern was supported by reference to a press release from the United States Federal Commission and European Regulators, sharing this concern.

Since the article was released there has been ongoing media interest, as well as public interest. Inland Revenue has received more than 8,000 Privacy Act requests from individuals seeking to understand if they have been included within the data shared to social media platforms. In addition, 23 OIA requests for information about custom audience lists have been received.

In response to public concerns, Inland Revenue paused all social media advertising using custom audience lists and undertook a review.

1.1 Scope of the review

This review sets out to understand:

- whether Inland Revenue's use of custom audience lists for targeted advertising complies with the Privacy Act 2020.
- how Inland Revenue provides custom audience lists to social media platforms.
- data security and retention in the platforms.
- whether social media platforms are using data from custom audience lists for their own purposes.

In order to address these areas, it is important to understand why custom audience lists are used and how the data is shared from Inland Revenue to social media platforms and in what form.

2.0 Understanding the use of custom audiences lists for social media advertising

This section outlines why custom audience lists have been used at Inland Revenue for targeted advertising on social media platforms and the role of hashing in uploading these lists.

2.1 Overview

In the course of its work, Inland Revenue is required to contact customers for a variety of reasons which supports the integrity of the tax system. Under the Tax Administration Act the Commissioner is charged with the care and management of the tax system and in particular, the Commissioner is required to have regard to the importance of promoting compliance including voluntary compliance. To support this, Inland Revenue undertakes a wide range of marketing activities helping customers know about available support, new products or when they may have a return or debt due. This helps to ensure as many taxpayers as possible can meet their obligations or claim their entitlements. Inland Revenue uses a variety of channels for marketing including billboards, digital advertising, videos, radio and social media.

In an increasingly digital world, social media has grown in relevance as a major advertising channel. Inland Revenue has been actively using social media channels for targeted and non-targeted advertising, including custom audience lists, for over ten years. It has been an effective way to help customers meet their obligations and access their entitlements.

Inland Revenue opened its first social media account in 2013, and first trialled targeted social media advertising using custom audience lists in 2014. This was performed with the support of a New Zealand based marketing agency.

Advertising campaigns using custom audience lists are carried out by the marketing team directly through the following platforms:

- Meta (specifically Facebook),
- Google (including YouTube),
- and LinkedIn (Inland Revenue has two LinkedIn accounts, one for Tax Professionals, the other for general Inland Revenue followers).

The adverts are relevant to the target audience in terms of the subject matter (e.g. student loans, tax refunds) and are general reminders about things like payments. They do not, for example, specify that someone has a student loan.

Inland Revenue also runs campaigns through external advertising agencies. The external agencies are not involved in all campaigns. Their role includes for example, setting up the advertising campaign, designing posts and linking these to matched audience lists (not the unhashed custom audience list information) that Inland Revenue has set up. This approach is typically used for larger campaigns that use a range of channels to reach customers. The process of how this information is shared with them can be seen in Section 6.0 Information Security Review, Sharing with Inland Revenue Advertiser.

2.2 What are custom audience lists?

Custom audience lists are used in targeted advertising to improve accuracy of reaching the appropriate audiences with the relevant advertisements. Custom audience lists include a range of data that will help identify those relevant people – for example first name, last name, email and date of birth (however these values depend on the platform and how

Inland Revenue use it). No financial or tax information about these individuals is used in creating custom audience lists.

These lists are uploaded to the platform after a procedure called hashing, which means that an individual is not able to be identified from it. This data is then matched with data the platform holds.

Further details about how the matching process is performed, the data included, and data security applied is outlined below in the report.

There are two scenarios where custom audience lists are created for advertising campaigns at Inland Revenue:

1. Monthly advertising for reoccurring activity (e.g. Student Loan campaigns, GST customers, Working for Families):
 - The Centre of Enterprise Data and Analytics (CEDA) team provide the file outlining the target audience monthly. There are a range of user segments covered to target the different campaign audiences.
 - These occur monthly as the audience experiences lots of change (e.g. have overdue debt, moving overseas, self-employed).
2. Direct advertising campaign:
 - These are for advertising which are one off, or non-consistent campaigns (e.g. Brightline property tax rule).
 - The Compliance Strategy and Innovation (CSI) and CEDA teams are involved in extracting the custom audience lists that will be used to establish target audiences.

The methodology for these targeted advertising campaigns differs depending on the platform being used, and whether the content is targeting individuals or businesses.

2.3 Uploading data - what does 'hashed data/hashing' mean?

When information is uploaded to social media platforms it goes through a process called hashing, which de-identifies the data. It is important to understand the hashing process to assess whether sharing information in this way for advertising purposes breaches privacy. The process and requirements are set up by the platform and anyone using this form of targeted advertising agrees to the process. It is up to each organisation to assess the suitability of the process.

2.4 How hashing is used in custom audience lists

A targeted advertising campaign is a marketing strategy to display adverts to a specific group of people. A set of information is required before any agency can pinpoint this specific group e.g. demographic, geographic, and interests.

Inland Revenue may provide information such as first name, surname, date of birth, email address, phone number, city, postal code and country to our advertising platforms. This information is hashed locally within an Inland Revenue managed internet browser, before being sent to the platforms through a secured channel. Respective platforms will then match it with their internal database, to create a custom audience list. The uploaded hashed data from Inland Revenue gets deleted once this process is complete. Any unmatched data is also deleted.

2.5 Conclusions

Inland Revenue has been actively using social media channels for targeted and non-targeted advertising for over ten years.

Targeted advertising on social media platforms is carried out by using custom audience lists to improve accuracy of reaching the appropriate audiences with the relevant advertisements. These lists include a range of data that will help identify those relevant people – for example first name, last name, email and date of birth. This information is hashed locally within an Inland Revenue managed internet browser, before being sent to the platforms through a secured channel.

The Privacy Act describes personal information as information about an identifiable individual. When hashed data is uploaded to the platforms it is not considered personal information.

3.0 Hashing

This section outlines in detail the process of hashing to de-identify personal information and known risks and controversies.

3.1 Overview

Hashing is a process of scrambling raw data into a fixed-size string of seemingly random characters – called a hash. Key characteristics of hashing include:

- Consistent – Hashing the same data with the same hashing algorithm will produce the same output, regardless of who performed the hashing.
- Fixed output length – Regardless of input size, the output is always the same fixed length to enable effective storage.
- Fast – Hash function is quick.
- Collision resistance – The more advanced the algorithm, the less likely to find two inputs that produce the same hash output.
- Basically irreversible – Modern hashing algorithms (SHA-256) are essentially irreversible with existing technology. This may change in the future with the emergence of quantum computing.

3.2 Common Use Cases

Hashing is used for a range of use cases and key examples include:

- Password storage: Passwords are commonly stored as hashed. During authentication, the user input password is hashed and matched against the stored hash. Usually incorporated with salt technique (explained below) to mitigate a rainbow attack, making it more secured.
- Data integrity: Hashes can be used to validate the file integrity; a user can compare a downloaded file's hash with the original source hash.
- Digital signatures: digital signatures leverage hashing to verify data integrity during transit. Messages are encrypted, then hashed, and both the encrypted message and the hash are sent to the destination. The encrypted message is then rehashed and then compared to the original hash to validate the integrity of the message.

3.3 Hashing Algorithms

Commonly used hash algorithms (set of rules) include but are not limited to – MD5, SHA1, SHA2 (SHA-256,384,512), and SHA-3. Each of the algorithms carries different levels of complexity and security level.

In the context of social media usage, the platforms used by Inland Revenue implement the SHA-256 hashing algorithm for custom audience list hashing.

SHA-256 is an approved cryptographic algorithm (as defined in the NZ Information Security Manual (NZISM) v3.8) for hashing purposes on information classified In Confidence and below. The NZISM is the New Zealand Government's manual on information assurance and systems security. It is designed to meet the needs of agency information security executives as well as vendors, contractors and consultants who provide services to agencies.

Government Communications Security Bureau (GCSB) sets the standards and guidance that agencies are then required to consider as part of their risk management processes. Government agencies are to assess the security risks and implement the security controls relevant to their environment.

With the current technology and computation process capability, to reverse the hash, it would take multiple lifetimes of a person to try all the possible combinations for SHA256 (2^{256}). This may change with the emerging quantum computers as they may have the potential to break many of the current cryptographic systems. As of now, SHA-256 remains secure against quantum attacks.

3.4 Illustration of Hashing



3.5 Encoding and Hashing

The fact the same input always generates the same hash output is true only when the same character encoding is used. Character encoding is like assigning a letter or symbol a special number, so a computer can interpret and remember it. For example, a computer may see the character "A" as number 65, "B" as 66. The numbers assigned can be different depending on the encoding used.

Commonly used encodings are ASCII, UTF-8, UTF-16, and ISO-8859. Different encodings are used to address different requirements e.g. ASCII were designed to be compact hence used in the early computers with limited memory.

The differences of how encoding affects the output is shown below:



This shows that using different encodings creates different results and combinations, which makes it more challenging for a rainbow attack to be successful (see below for explanation on rainbow attacks).

3.6 Known Risks and Controversies

Hashing does not anonymise data

It is possible to reverse-engineer or brute-force hashes under certain conditions, especially if the input (names, email addresses) is short or from a limited set. Hashing is not immune to a rainbow table attack. In a rainbow table attack, the attacker uses a pre-computed table of hashes to look up the plaintext version of a hashed info. This means that if there is a match spotted between the pre-prepared table of hashes and something in the hashed data, then it would be possible to recognise the plain text version of the hashed data.

See below for a simplified rainbow attack illustration:

A malicious actor identifies Aaron is in the compromised list by finding a match between the pre-prepared table and the hashed list.

Compromised Data

First name (hashed)

```
39fdbdb8ddf75a006ffec2a3ba95c3a04ce5517c608a786ef9a042af9843bd8c
300648cea31d54fdec1ce29f8771bcae5107e97e6a9d3b98567a28cf85306b64
7e8c729e4e4ecc320cb411c4d1419bf5fbad733212d4e9491b7630aaef0b8b1c
458556af7d03d652f06f2fc28727390e6860989a70246cbb93699995f1766798
```

Rainbow Table (common First Name)

| First name | Hash |
|--------------|---|
| Aaron | 39fdbdb8ddf75a006ffec2a3ba95c3a04ce5517c608a786ef9a042af9843bd8c |
| Adam | f7f376a1fcd0d0e11a10ed1b6577c99784d3a6bbe669b1d13fae43eb64634f6e |
| Alex | 4135aa9dc1b842a653dea846903ddb95bfb8c5a10c504a7fa16e10bc31d1fdf0 |
| Ben | 6900dfb584a9e7b72109b1b72518ff62af7a81b7b5a74066a56e5edba6dcf973 |

Re-identification risk

A malicious actor with access to a large dataset or matching inputs (like names and email addresses) could match the hashed values to real-world identities. For instance, if the same hash appears in different datasets, they can cross-reference and de-anonymise individuals by correlating the data. The rainbow table illustration above is applicable here as well.

Lack of salting

Salting is a technique used to further anonymise data during the hashing process. Random data is added to the information to be hashed. This provides increased protection against rainbow table lookups. SHA256 itself does not include any mechanism for salting. Without this technique, identical inputs will always produce identical hashes, making the system vulnerable to attacks like rainbow tables or hash collisions across systems.

Man-in-the-middle attack (MITM)

A MITM attack allows a malicious actor to intercept communication between a client and server (in this context: Inland Revenue and advertising platform). There are a few methods a malicious actor could use to execute this attack:

1. Intercept – Reads all data being transmitted e.g. Attacker sits on the same Wi-Fi network and captures the data using tools like Wireshark (packet sniffing tool).
2. DNS spoofing – Attacker can spoof DNS responses and redirect a user to malicious websites, in this context they pretend to be a social media platform and receive the customer lists.
3. Man-in-the-browser attack – variant for MITM, attacks and modifies a website and injects a malicious script, then harvests the uploaded data.

3.7 Reducing risks associated with hashing

For any form of hashing attack, the data itself must first be accessed.

Accessing hashed data is challenging given the layers of security that are part of uploading and storing hashed data within the platforms that use it.

The risks are reduced significantly by transmitting the data through an encrypted channel like Transport Layer Security (TLS). This can be observed by the HTTPS URL prefix at the internet browser address bar. By encrypting the data, even if an attacker intercepts the data, they cannot read it without the keys used for encryption – these keys are exchanged securely during a secure handshake process. This process ensures both parties use the same method of communication e.g. algorithms and ways to authenticate each other. If the handshake fails, the connection is terminated preventing any insecure communication.

3.8 Published information about hashing usage and risks

In the media articles, two sources were referenced which discussed the risks of using hashing to anonymise personal information. These are considered below:

Article 1: Introduction to the Hash Function as a Personal Data Pseudonymisation Technique (EDPS)

Context

This is a paper written by the European Data Protection Supervisor (EDPS) in 2019 - highlighting hash re-identification. This provided a view of key concerns around using hashing for data anonymisation, and techniques to protect against re-identification risks.

Risk 1, Speed to de-anonymise hashed data

The paper visualised this risk through a scenario on linking a seemingly anonymised hash value back to a potential 20 million phone numbers within 20 seconds (1 million hashes generated per second). To replay this scenario in New Zealand, there is approximately 5.8 million mobile phone connections and it would only take ~6 seconds to turn those phone numbers into hashes. The more attributes (names, email, phone numbers) a malicious actor has access to, the easier it is to relink the hash back to a person.

Relevance to Inland Revenue

Multiple attributes are used for the custom audience list; the process to match the hash back to a person takes extra effort, but not impossible. The re-identification risk remains.

Risk 2, Strength of algorithm used

From the same paper, it mentioned the older MD5 and SHA-1 algorithm should not be used as they have known vulnerabilities like collisions attack. Instead, cryptographically resistant hash functions should be used, including SHA-2 and SHA-3.

Relevance to Inland Revenue

Inland Revenue uses SHA-2 (SHA-256) for this custom audience list function SHA-256 is an approved cryptographic algorithm (as defined in the NZ Information Security Manual (NZISM) v3.8) for hashing purposes on information classified In Confidence and below.

Techniques to protect against reidentification

Several techniques were mentioned in the paper to hinder or reduce the re-identification risks, including encrypt before hash, various salt models (increases randomisation of the hash), and differential models.

Relevance to Inland Revenue

In this context for Inland Revenue, none of the techniques can be used as the social media platforms only accept hashes without the above additional security techniques.

Article 2: No, hashing still doesn't make your data anonymous (FTC)

Context

This blog from the US Federal Trade Commission (FTC) highlights hashing does not anonymise data. While it transforms data into a seemingly anonymous string, hashes can still be reverse engineered (this is demonstrated through the rainbow table attack shown above). Hashes alone do not offer true privacy protection as attackers can exploit predictable data like email addresses.

The FTC concludes that to truly protect privacy, additional measures such as salting and other security techniques should complement hashing. Simply hashing sensitive data is insufficient to prevent re-identification, particularly when the inputs are known or guessable.

Relevance to Inland Revenue

This blog highlights while hashing can de-identify data, hashes can be re-identified without additional security techniques. In Inland Revenue's case we are limited by the processes around hashing provided by the social media platforms, which do not provide salting capabilities (for example). However, there are additional security controls around the use of custom audience lists including uploading data through encrypted channels which significantly reduces the likelihood of data being accessed, even if it were to be intercepted.

3.9 Conclusions

Hashing is a commonly used way of sharing de-identified data.

In the context of social media usage, the platforms used by Inland Revenue implement the SHA-256 hashing algorithm for custom audience list hashing. This meets the standards and guidance set by Government Communications Security Bureau (GCSB). As of now, SHA-256 remains secure against quantum attacks.

While it is possible to apply a rainbow attack to hashed data under certain limited conditions, the hashed data must first be accessed.

The additional layers of security provided, in addition to hashing, by transmitting data through encrypted channels, means the data is safe and not at risk from such attacks.

Inland Revenue uploads hashed data with additional transport layer security using recognised encryption channels and there is no evidence that this process has ever been compromised.

Additionally, the data is uploaded in a process that connects machine to machine with no human intervention.

4.0 Process Flow – How targeted social media advertising happens

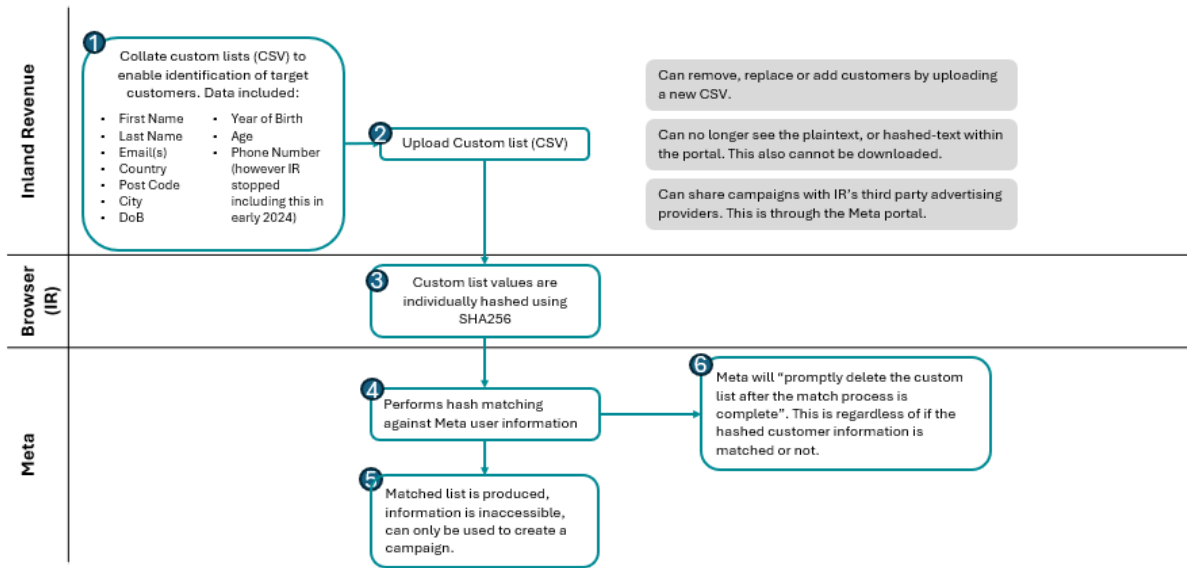
This section shows process flows to show how targeted social media advertising is performed at Inland Revenue for each social media platform used, where the data goes and what form it is in during different stages.

4.1 Process flows

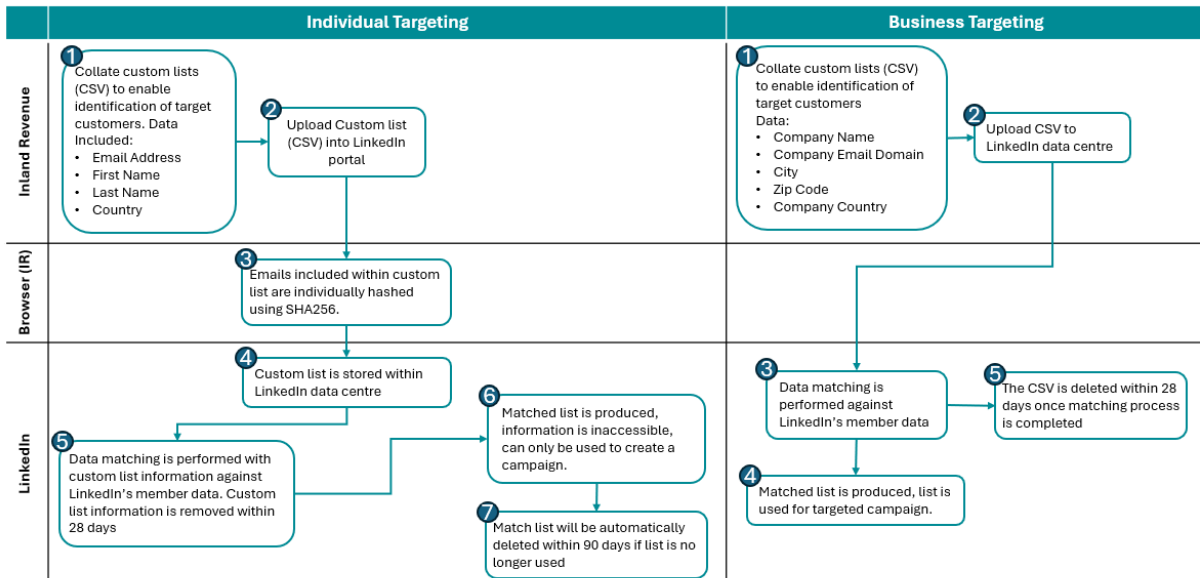
The following processes are built and maintained by the respective social media platform. In this situation, Inland Revenue is a user and not able to define any specific requirements.

Hashing and these data related controls are system security controls – which means these are automated processes where human intervention is not required and information is not disclosed or viewed by a human.

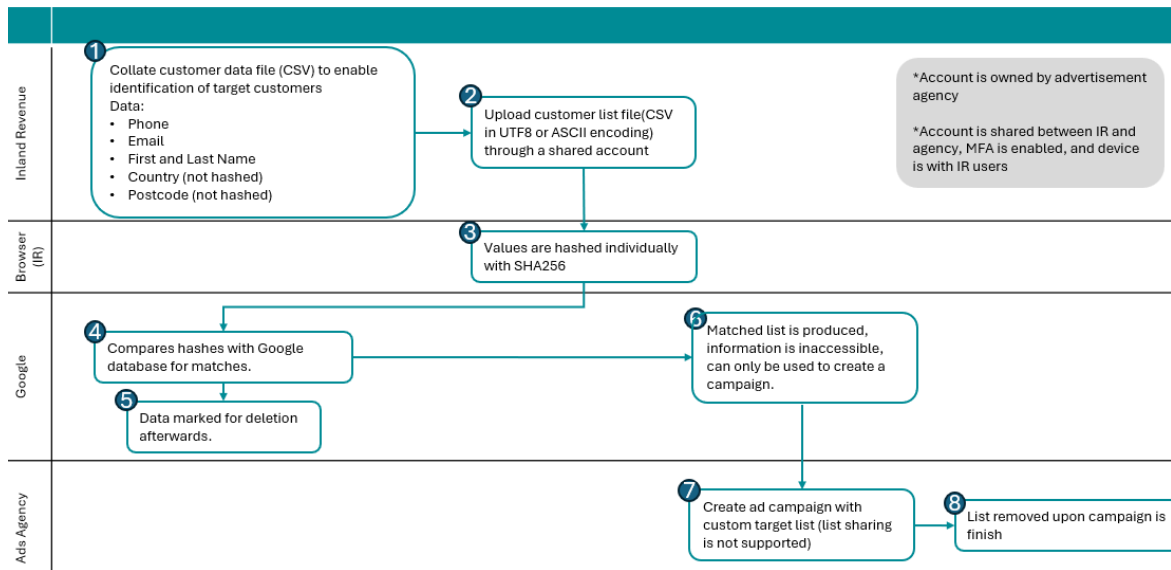
4.2 Meta



4.3 LinkedIn



4.4 Google



5.0 Additional Security Controls

This section outlines the security controls that apply to data when it is uploaded to social media platforms.

5.1 Overview

Hashing is primarily used for de-identification and should be used in conjunction with wider controls available to ensure data security. Hashing and these data related controls are system security controls – which means these are automated processes where human intervention is not required and information is not disclosed nor viewed by a human.

When determining the security/safety of this data when uploaded to the social media platform we have looked at:

- Is the data stored safely (i.e. is the platform secure)?
 - This information is provided through security assurance reports provided by the platforms (seen in *Section 7.0. Assurance from Supplier Platforms*)
- Is the data secure in transit (when moving from an Inland Revenue machine to the platform)?
 - Data transmission between Inland Revenue devices (client) and social media platforms (server) happens through a secured channel with Transport Layer Security (TLS). It encrypts data that is transmitted across the web. This means that anyone who tries to intercept this data will only see a garbled mix of characters that cannot be read without the keys used for encryption
- Is the data retention appropriate?
 - Data retention for the custom audience list (CSV) that is uploaded depends per platform:

- Meta: “promptly delete customer list after the match process is completed”
- Google: Will not retain custom list for any longer than necessary once the matching processes are completed.
- LinkedIn: The customer audience list will be automatically deleted after 28 days of upload.

This information is included within Section 6.0. Information Security Review, Data Retention.

- Is the data used only for its intended purpose, and not used to build/enhance user profiles?
 - All platforms have attested to not using the data provided to build new profiles, or append existing profiles.

This information is included within Section 6.0. Information Security Review, User Profile Enhancing.

5.2 Conclusions

The additional security controls outlined, together with the processes being automated where human intervention is not required and information is not disclosed nor viewed by a human, gives Inland Revenue confidence that adequate security controls are in place.

6.0 Information Security Review

The table below outlines the data provided, login process, data security, retention, sharing with Inland Revenue Advertiser, User Profile Enhancing and Second Order Information Assumptions for each platform.

| | Meta (Facebook) | Google | LinkedIn | |
|---------------|---|--|---|--|
| | Individual | Individual | Individual | Company |
| Data Provided | <ul style="list-style-type: none"> ▪ First Name ▪ Last Name ▪ Email(s) ▪ Country ▪ Post Code ▪ City ▪ Date of Birth ▪ Year of Birth ▪ Age ▪ Phone Number(s) – NOTE: Inland Revenue stopped including phone number in early 2024 | <ul style="list-style-type: none"> ▪ First Name ▪ Last Name ▪ Email(s) ▪ Country (plaintext) ▪ Post Code (plaintext) ▪ Phone Number(s) | <ul style="list-style-type: none"> ▪ First Name (plaintext) ▪ Last Name (plaintext) ▪ Email Address ▪ Country (plaintext) | <ul style="list-style-type: none"> ▪ Company Name ▪ Company Email Domain ▪ Company Country ▪ City ▪ Zip Code <p>All these fields are in plaintext</p> |
| Ad dit | The following data fields can be provided to the relevant platform for data matching, however IR does not currently use these. | | | |

| | Meta (Facebook) | Google | LinkedIn | |
|---|--|---|---|--|
| | Individual | Individual | Individual | Company |
| | <ul style="list-style-type: none"> ▪ Mobile Advertiser ID (Madid) (hashing not required) ▪ Facebook App User ID (UID) (hashing not required) ▪ Facebook Page User ID (PageUID) (hashing not required) ▪ State/Province ▪ Gender ▪ Value | | <ul style="list-style-type: none"> ▪ Job Title ▪ Employee Company ▪ Googleaid ▪ Googleuid ▪ Appleidfa <p>All these fields are in plaintext</p> | <ul style="list-style-type: none"> ▪ Company Website ▪ LinkedIn Company Page URL ▪ Stock Symbol ▪ Industry ▪ State ▪ Company Code <p>All these fields are in plaintext</p> |
| Data Security | <p>The customer list (provided as CSV) is hashed using SHA-256 within the internet browser automatically (on the Inland Revenue device), then this hashed information is uploaded and stored within the platform.</p> <p>This information is hashed individually (e.g. each piece of data provided about an entity is hashed, instead of all data per entity being hashed at once).</p> <p>Inland Revenue staff cannot access the plaintext or hashes within the platform.</p> | <p>The customer list (provided as CSV) is hashed using SHA-256 within the internet browser automatically (on the Inland Revenue device). Hashing is performed on both individual attributes (name, email, phone number) and combined attributes (e.g. mailing address which is a combination of countries and zip codes, and names). This hashed information is uploaded and stored within the platform.</p> <p>Inland Revenue staff cannot access the plaintext or hashes within the platform.</p> | <p>Emails within the customer list (provided as CSV) are hashed within the internet browser automatically (on the Inland Revenue device) using SHA-256. The other values provided (first name, last name and country) are not hashed. The CSV is then uploaded and stored within the LinkedIn data centre.</p> <p>Inland Revenue staff cannot access the plaintext or hashes within the platform.</p> | <p>The customer list (provided as CSV) is uploaded into the Inland Revenue internet browser and NOT hashed. This information is stored within the LinkedIn data centre.</p> <p>This information remains visible to Inland Revenue staff with access to the Inland Revenue LinkedIn tenant.</p> <p>This data will stop being visible within the solution after 90 days of inactivity with the company list.</p> |
| <p>Data in transit security: The transmission of the hashed CSV file from Inland Revenue device to the respective social media platforms (server) happens through a secured channel with Transport Layer Security (TLS). It encrypts data that is transmitted across the web. This means that anyone who tries to intercept this data will only see a garbled mix of characters that is nearly impossible to decrypt. This internet browser also authenticates the identity of the websites through SSL certificate and prompts the user if it is a fraudulent website.</p> | | | | |

| | Meta (Facebook) | Google | LinkedIn | |
|----------------------------|---|--|--|---|
| | Individual | Individual | Individual | Company |
| Data Retention | Meta says they “promptly delete customer list after the match process is completed” – this is regardless of if the hashed customer information is matched or not. | Google says they will not retain customer list for any longer than necessary once the matching processes are completed, Google promptly delete the data files IR uploaded via the internet browser. The matched lists are removed following a campaign’s completion. | LinkedIn says that the customer list will automatically be stored for 28 days and deleted. The matched list created based on the custom list will be stored for 90 days and deleted, if it is not edited or being used in any active campaign. | The customer list is stored for 28 days and deleted. The company list is the only thing that remains, which is deleted after 2 years. However, this cannot be used for targeting after deletion. Company list data is no longer visible after 90 days of inactivity |
| Log on Process | Inland Revenue staff log on to Meta through their personal user accounts (which are expected to have MFA enabled). From here they can access the Inland Revenue Meta Business Suite and perform updates. | The Inland Revenue marketing team to log on to our advertising agencies Inland Revenue specific workspace. Password is shared in plaintext, MFA is enabled - registered MFA device is held by Inland Revenue staff. | Inland Revenue staff log on to LinkedIn through their personal user accounts (which are expected to have MFA enabled). From here they can access the Inland Revenue LinkedIn and perform updates. | |
| Sharing with IR Advertiser | When in the Inland Revenue portal, once de-identified matched lists are uploaded these can be shared with IR’s advertising companies as required. The list within the portal is shared through the portal to the advertising agency. The advertising agency will have their own portal. | As Google does not support matched list sharing, a shared account is used between the advertising agencies and Inland Revenue. Inland Revenue staff upload the custom lists into the shared tenant and do not share the plaintext CSV with third party advertising partners. | When in the Inland Revenue portal, once matched lists are uploaded these can be shared with IR’s advertising companies as required. The list is shared through the portal to the advertising agency. The advertising agency will have their own portal. <i>Note: Inland Revenue has no recorded examples of where company lists have been shared with an advertising agency. If this was shared with an advertising agency, they would be able to see the plaintext information for company lists.</i> | |
| User Profile Enhancing | Meta will not use the data to build or append interest-based profile | Google will not use the data files to build or enhance customer profiles. | LinkedIn will not modify, reverse engineer, decompile, create other works from, or disassemble any Confidential information. LinkedIn does not profile non-members, and we also do not create or enhance behavioural profiles of members with off-LinkedIn data. | |

| | Meta (Facebook) | Google | LinkedIn | |
|--------------------------------------|--|------------|------------|---------|
| | Individual | Individual | Individual | Company |
| Second Order Information Assumptions | <p>Interaction between a user and a targeted advertisement generates statistics to measure how effective a campaign is, this includes: impressions, reach, frequency, link clicks, click through rate, cost per click, reactions, shares. These are accessible by Inland Revenue.</p> <p>Aside from marketing related stats, each social media platform may be able to record all user’s activities/behaviour whenever they interact with the platform (regardless of whether a customer list campaign is used), including but not limited to Likes and Shares, mouse hover tracking, page/ad view duration, user device information (operating system, location, language, browser version), interests based on a user’s search query.</p> <p>Inland Revenue digital advertising includes a ‘call to action’ e.g. please visit this Inland Revenue webpage for more information on student loan repayment process. By interacting with the advertisement, it is possible for a social media platform to infer that a user may be interested in certain topics like student loan (this is different from User Profile Enhancing). However, simply clicking or viewing the ads will not give a social media platform conclusive data but rather increases the likelihood of a user being interested in a certain topic.</p> | | | |

7.0 Assurance from Social Media Platforms

Inland Revenue requested security assurance reports from the Social Media platforms used for targeted advertising with custom audience lists.

7.1 Assurance Information

The assurance information provided, is detailed in the table below.

Description of the different types of assurance requested are outlined below.

| Platform | ISO27001 | SOC2 Report | Security Attestation |
|----------|----------|-------------|----------------------|
| Google | Y | | |
| Meta | | Y* | Y |
| LinkedIn | Y | Y | |

*Meta policy assures relevant audit report such as SOC 2 Type II report is available upon request. Inland Revenue has made the request and is awaiting receipt of the report.

ISO27001

Is an internationally recognised standard for managing information security. The standard outlines requirements for establishing, implementing, maintaining and continually improving an information security management system (ISMS) – the aim of which is to help organisations make the information assets they hold more secure. To receive the certification, an accredited certification body performs an audit against the ISO27001 areas for the relevant party. Once obtained, the certification is valid for 3 years with an annual surveillance audit process to maintain compliance.

System and Organisation Controls (SOC) Report

Is produced through an independent audit of a company's information security systems. This includes the controls the organisation has in place to safeguard those systems, and the information stored, processed, and/or transmitted by them. It comprises five Trust Services Criteria: Security, Availability, Processing Integrity, Confidentiality and Privacy.

There are two SOC reports an organisation can seek:

- SOC 1 Reports – assurance on financial reporting and accuracy.
- SOC 2 Reports – cloud and data centre security controls, this focuses on: Security, Availability, Processing Integrity, Confidentiality and Privacy.

For both SOC reports, there are two types of reports:

- Type I Report – performs point in time assessment of controls.
- Type II Report – performs assessment of controls over a period of time. This includes testing of operational effectiveness over time.

Only a Certified Public Accountant (CPA) or an organisation accredited by the American Institute of Certified Public Accountants (AICPA) can conduct a SOC audit. These are generally performed on an annual basis, however they can be conducted more frequently.

Security Attestation

Inland Revenue received a copy of an assertion report prepared by PWC in 2013 for Facebook which assessed Facebook's Custom Audiences security controls. The conclusion and the opinion over Facebook's management assertion is fairly stated as below:

- Information provided to Facebook is not shared with other advertisers and third parties.
- Facebook implemented safeguards and controls against accidental or unauthorised access, use, alteration or disclosure of data.
- Information provided by advertisers for Custom Audience list generation is only retained for as long as needed for matching process. Facebook disposes of the data once it's no longer required.

8.0 Key findings

The following section outlines the key findings from the review, in relation to what the review set out to understand:

- whether Inland Revenue's use of custom audience lists for targeted advertising complies with the Privacy Act 2020.
- how Inland Revenue provides custom audience lists to social media platforms.
- data security and retention in the platforms.
- whether social media platforms used data from custom audience lists to enhance their own user profiles.

8.1 Use of custom audience lists and privacy

In 2016 as part of usual due diligence when using customer information, Inland Revenue completed a privacy impact assessment regarding the use of custom audience lists on Facebook. A privacy assessment helps identify potential effects that a proposal may have on people and their information. The assessment considered how custom audiences worked, what personal information would be used, security and retention of that information. The privacy impact for the project was assessed as Medium; several privacy risks were identified but could be adequately mitigated.

Privacy assessments were not done for Google or LinkedIn as the information, processes and terms and conditions were the same or similar to what was already assessed.

Inland Revenue updated the privacy assessment into the use of custom audience lists, considering all platforms used, in 2024 following media commentary. The privacy impact remains rated at Medium.

The Privacy Act describes personal information as information about an identifiable individual. When hashed data is uploaded to the platforms it is not considered personal information.

This is because it is in hashed form and does not identify anyone. In addition, the hashing of the data is one layer of security and the data is uploaded into a secure platform where it is matched with the platform's hashed data in an automated, machine-to-machine process. The data is deleted after the matching process is completed and not used in any other way by the platforms.

Inland Revenue's Privacy Policy notifies people how we use their information and that we send hashed information to third parties.

To comply with other Privacy Act obligations, Inland Revenue informs customers how it uses information via its website. Inland Revenue's Privacy Policy states that if a customer gives Inland Revenue their email address or mobile phone number, we may use these to send them reminders about their tax affairs or information about our products and services. Along with other information, the email address, and sometimes phone number, is used in compiling the custom audience list.

The Policy also informs the public that Inland Revenue uses information to target advertising to them indirectly via a third party:

Why you might see a certain advertisement on social media

We may also use or disclose your information to third parties to assist us to communicate or market our services to you.

*To reach groups of people with information that is relevant to them while protecting their privacy, we sometimes provide **hashed** and fully anonymised information to social media channels when placing advertisements. In this process, your personal information is treated with the utmost integrity by us. The social media channel is not given any identifiable information. We fully comply with our obligations under the Tax Administration Act and the Privacy Act to protect your personal information.*

This statement, should it remain in the Privacy Policy, will be amended as hashed information is not anonymised but de-identified (pseudo-anonymised).

The Privacy Act does not require an individual to consent to the use of their information. The Act is purpose-focused rather than consent-focused and allows personal information

collected for one purpose to be used for related purposes. Informing people of their tax obligations or entitlements is directly related to why Inland Revenue holds customer information.

The source information used to collate a custom audience list is personal information for the purposes of the Privacy Act. 'Personal information' is defined as information about an identifiable individual. The source information is also sensitive revenue information (SRI) under the Tax Administration Act 1994 (TAA). SRI is revenue information that identifies, or is reasonably capable of being used to identify, a person or entity, whether directly or indirectly.

SRI is to be kept confidential under the TAA and not to be disclosed unless the disclosure is specifically permitted under the TAA.

Inland Revenue takes its TAA obligations in relation to taxpayer confidentiality seriously and has internal processes for signing off the disclosure of information. Written approval has been located for each of the three platforms used but, given organisational and personnel changes over the last 10 years, there are gaps in timing for that documentation. However, we have been able to identify that certain campaigns in those time gaps involved someone with the delegated authority for disclosing the information and who was very familiar with IR's confidentiality obligations

8.1.2 Historic use of custom audience lists on additional platforms

In the course of the review, it was found that there had been historic use of custom audience lists for targeted adverts on TradeMe in 2017 as part of the Business Transformation Programme. There is no evidence that this platform has been used for targeted advertising with custom audience lists since this instance.

8.2 Providing custom audience lists to social media platforms

In order to advertise with social media platforms, Inland Revenue is required to upload information in the way each platform requires.

The information provided to the platforms is securely uploaded through an Inland Revenue browser. Where data is hashed, this is automatically performed using a standard algorithm within the browser of the Inland Revenue device uploading the custom audience list. Both the hashing algorithm and transmission are compliant with NZISM specifications. All data is securely stored on the platform.

8.3 Data security and retention

Inland Revenue requested security assurance reports from the social media platforms used for targeted advertising with custom audience lists.

After assessing these, Inland Revenue is satisfied that the information provided shows appropriate standards for managing information security, as there are no exceptions noted for the in-scope security controls.

8.4 Using data from custom audience lists to enhance their own user profiles

The social media platforms indicated that custom audience list information is not used to enhance or build profiles of their own users.

8.5 Conclusions

Having undertaken the review, we believe that the process taken in using custom audience lists in targeted social media marketing is recognised as legitimate both in New Zealand and internationally.

However, there continues to be ongoing public concerns about the practice of using custom audience lists for social media advertising.

Inland Revenue recognises the importance of building and maintaining public trust as a cornerstone of an effective tax and social policy system.

The review has highlighted issues can arise when sharing information with a third party.

It is essential that information sharing with any external party when there is a need for troubleshooting, must be performed securely. Attachment or data upload must be encrypted prior to leaving Inland Revenue's network. Once the troubleshooting process is completed, ensure the suppliers remove or delete the information. This should be documented as part of the standard operation procedures.

8.6 Recommendations

It is recommended in response to customer concerns that Inland Revenue cease the use of custom audience lists for the foreseeable future.

It is recommended to set review dates for Privacy Impact Assessment's (PIA) where relevant in the PIA register to see if anything has changed and if Inland Revenue still has the same view of the activity being undertaken and if the impact rating is still the same.

It is recommended that Inland Revenue undertake a full review of the use of social media for marketing and advertising purposes including the use of specific capabilities and the information sharing required to enable them.

It is recommended that, in the event Inland Revenue determines that social media marketing or advertising capabilities which target individuals is required, it considers the appropriateness of the implementation of a Opt in/Opt Out process within the core tax system.

9.0 Unintended disclosures

During the course of review two unintended disclosures were identified.

9.1 LinkedIn

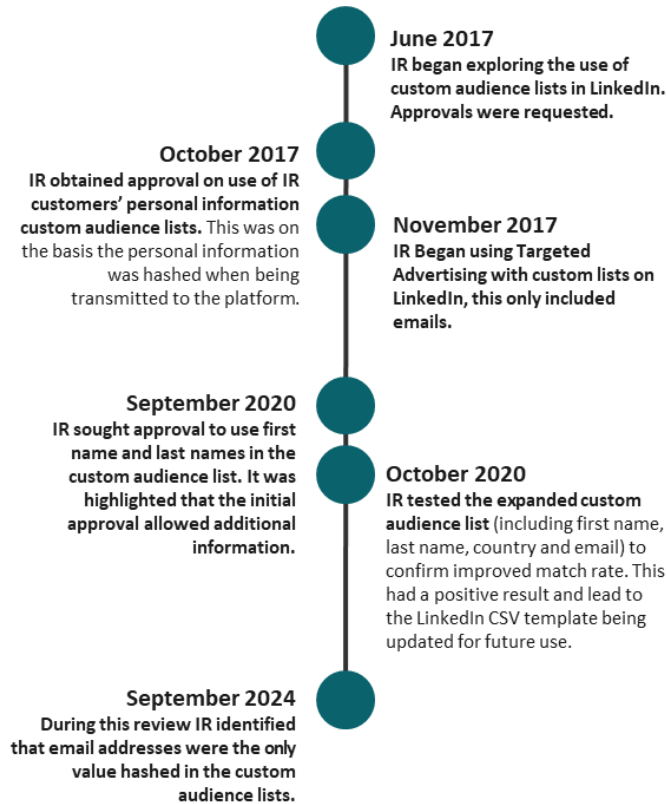
During the review, it was identified that not all data being provided to LinkedIn through custom audience lists was hashed.

What happened?

Inland Revenue has been using custom audience lists with email, first name, last name and country for LinkedIn targeted advertising since 2020. Before this, only emails were provided. Approval was sought to expand the custom audience lists to include first name, last name and country. Approval was given on the assumption that all data provided was hashed and no individual was identified.

During this review it was identified that only emails within custom audience lists uploaded to LinkedIn are hashed.

A timeline is shown below.



Why is this a problem?

The unhashed information provided to LinkedIn is both Personally Identifiable information and SRI. Providing this information unhashed does not align with the marketing guidelines or the approvals obtained.

While this data was uploaded to be matched in a machine-to-machine environment, which does not enable LinkedIn to identify these customers, disclosing a person's name does not comply with the Privacy Act, is not strictly necessary, and is considered a non-notifiable breach.

A non-notifiable breach is a breach where no serious harm has or is likely to occur and does not need to be reported to the Privacy Commissioner or affected individuals.

9.2 Meta

What happened?

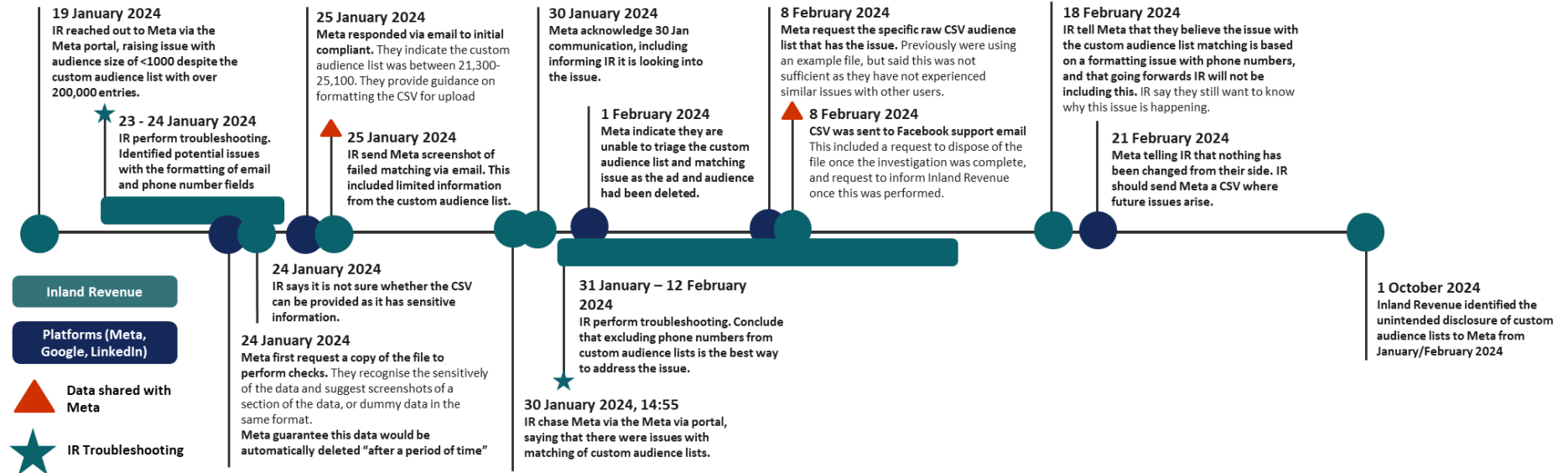
In the course of the review, it was identified that there was one instance of an unhashed custom audience list being shared with Meta support.

In early 2024, Inland Revenue had an issue with the matching of a custom audience list in Meta. This issue saw a file with over 200,000 entries having an unexpectedly low match rate of under 1,000. Based on this, Inland Revenue started to perform troubleshooting internally and reached out to Meta for support. This activity can be found in the timeline below.

Key information from this incident is:

- Inland Revenue provided a cleartext CSV to Meta Support for troubleshooting purposes (following Meta's request). This had 268,068 entries. Each entry included: phone number(s), first name, last name, city, country, zip code, date of birth, email(s), age, year of birth.
- Sharing this file via email was not part of the approved approach for custom audience lists.
- Inland Revenue performed its own troubleshooting and identified that the phone number field in this list was causing the issue. The phone number data was removed from the custom audience lists which resolved the issue.
- Meta stated that data provided would be "automatically deleted after a period of time".
- When the CSV was provided to Meta support, they were told to dispose of the file once completing the investigation and inform Inland Revenue. Meta assured information used for the troubleshooting is kept confidential and secure and they do not retain any records of personal information.

Timeline – Meta troubleshooting incident



Why is this a problem?

Sharing this file via email was not part of the approved approach for custom audience lists. This is not a notifiable privacy breach as it's reasonable to believe sharing the list with a Meta support person has not caused, and is unlikely to cause, serious harm to affected individuals.

Although the breach is not notifiable, when this activity was discovered, Inland Revenue contacted the Office of the Privacy Commissioner to let them know that this unapproved sharing of personal information had occurred. Inland Revenue has chosen to contact everyone who was included in the custom audience list for transparency and to rebuild trust.

9.3 Recommendation

It is recommended an internal assurance process be undertaken by Inland Revenue to understand how this breach happened.

Appendix 1: Key Public Communications


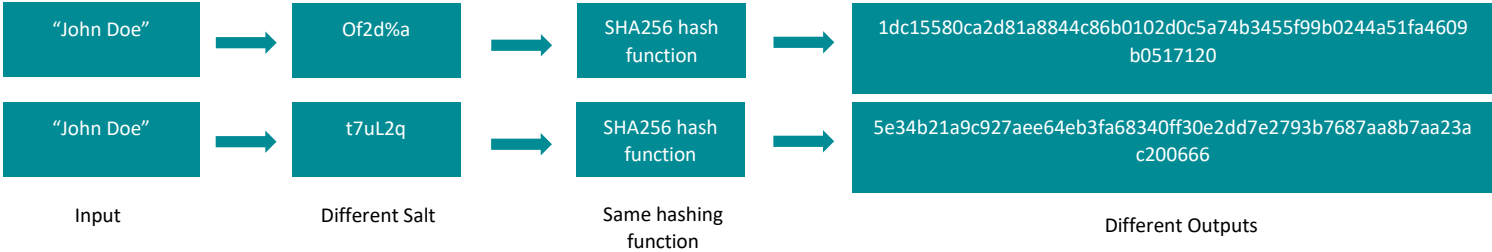
Key websites related to the public concern about Inland Revenue's use of custom audience lists:

- Initial article posted by RNZ: [Inland Revenue giving thousands of taxpayers' details to social media platforms for ad campaigns | RNZ News](#)
- Second article posted by RNZ: [IRD data sharing: Safety of anonymising detail to be examined | RNZ News](#)
- Additional RNZ Article: [Making a hash of it: The lowdown on Inland Revenue and your data | RNZ News](#)
- Inland Revenue's site outlining use of custom lists: [About custom audience lists \(ird.govt.nz\)](#)
- Additional PressReader Article [PressReader.com - Digital Newspaper & Magazine Subscriptions](#)

Sources referenced from the initial RNZ article:

- European regulator on hashing and problem on re-identification - [19-10-30_aepd-edps_paper_hash_final_en.pdf \(europa.eu\)](#)
- US Federal Trade Commission - [No, hashing still doesn't make your data anonymous | Federal Trade Commission \(ftc.gov\)](#)

Appendix 2: Glossary

| Term | Definition |
|---------------------------------------|--|
| Plaintext / Cleartext | Unhashed/unencrypted human readable text. E.g. "John Doe" is an example of plaintext. |
| Hashing | <p>Hashing is a process of scrambling raw data into a fixed-size string of seemingly random characters – called a hash.</p>  |
| Rainbow table | A rainbow table is a precomputed table for caching the outputs of a cryptographic hash function. |
| SRI | Sensitive Revenue Information |
| Salting | <p>Salting is a technique used to further anonymise data during the hashing process. Random data is added to the information to be hashed. This provides increased protection against rainbow table lookups.</p>  |
| Transport Layer Security (TLS) | Transport Layer Security (TLS) is a secure lock that protects your data online. IT encrypts the info, ensure the user is interacting with the right website (authentication, and makes sure nothing gets changed (integrity)). When there is a https in a web address prefix, it means TLS is implemented and keeping the connection safe. |

Appendix 3: Terms and Conditions w/ Social Media

The table below sets out examples of the types of confidentiality (including use of Inland Revenue's information), privacy, security measures and deletion of information clauses contained in the terms and conditions. **"Yes"** indicates if a service provider's terms and conditions contains a clause using that type of wording. A **"no"** just indicates that a service provider does not use that particular wording - the underlying obligation is expressed elsewhere in a different way.

| Confidentiality (including use of information) and privacy | Google | Meta | LinkedIn |
|--|---------------|-------------|-----------------|
| Limits use of IRD information for the purposes set out in the relevant agreement | yes | yes | yes |
| Service provider will comply with IRDs "instructions" on how information may be used | yes | no | yes |
| Service provider will not use information to: append to build or enhance interest based profiles; or modify, reverse engineer, decompile, create other works from, or disassemble any confidential information | yes | yes | no |
| Does not share information (subject to usual exceptions e.g. required by law) | yes | yes | yes |
| Does not sell information that identifies customers | yes | yes | yes |
| Will comply with data protection terms set out in a processor agreement | yes | yes | yes |
| Will comply with obligations applicable to it under applicable data protection legislation | yes | yes | yes |
| "Confidential Information" is defined in a way that includes the information IRD discloses in a campaign | yes | not defined | yes |
| Contains standard confidentiality clause limiting use and disclosure of IRDs information | yes | yes | yes |

| Confidentiality (including use of information) and privacy | Google | Meta | LinkedIn |
|---|--|----------------------------------|--|
| Implement processes and procedures to protect confidentiality and security of information | yes | yes | yes |
| Has appropriate content in privacy policy | yes | yes | yes |
| Security measures | Google | Meta | LinkedIn |
| Use/maintain security measures in connection with its provision of the services | yes | yes | yes |
| Has technical/physical safeguards to protect security of information and guard against accidental or unauthorised access, use, alteration or disclosure | yes | yes | yes |
| Sets out technical and organisational security measures | yes | yes | yes |
| Sets out access controls | yes | yes | yes |
| Security documentation/certification made available | yes (ASO 27001) | yes | yes (SSAE 18, ASAE 3402) |
| Deletion | Google | Meta | LinkedIn |
| Terms continue until all data deleted | yes | no | yes |
| Requirement to delete data within a certain period | yes (as soon as reasonably practicable or a maximum of 180 days) | yes (following matching process) | yes (on termination of the data processing services) |
| Requirement to delete data on IRDs request or IRD can delete itself | yes | yes | yes |

Appendix 4: Statistics on targeted advertising

Use of digital advertising and targeted campaign is observed internationally. According to a research report from Deloitte, there is a strong shift towards people-centric marketing and communication strategies. Public sector entities adopted commercial strategies such as integrating data and behavioral insights into their digital campaigns. Governments aim to influence behaviour, build trust, and support crucial initiatives like public health and economic recovery.

UK government has increasingly turned to data-driven advertising, utilising social media and targeted adverts to engage specific demographic groups with tailored content. Targeted advertising campaigns have also been used to address various social and public health issues, uses a combination of PR, community engagement, and targeted multilingual digital ads to deter illegal migration and educate migrants about legal routes and the dangers of illegal crossing.

Canada has a similar practice on using digital advertising to reach various demographic groups – Government of Canada Advertising for 2022-2023 shows 71% of their advertising budget is focused on digital media, including programmatic ads, social media, and search engine marketing. Adverts are aimed at informing citizens about government programmes, public health initiatives, and social services.