identifiers explainer guide.

this document provides a very simple explainer of the various identifiers in the Australia market, including an ID Matrix of the 20 most popular solutions

iab. australia
data advertising

overview.

purpose:

This document provides a very simple explainer of the various identifiers in the Australia market, including a visual ID Matrix of the 20 most popular solutions

The IAB Australia Data Council is keen to provide this as a starting point of research into this fast-evolving space, for both buyers and sellers.

This explainer has been supplemented by a fuller table of the 20 individual providers available in Australia with more detailed information therein.

extra information

extra

We have also included a simple intro, a glossary of terms and some suggested next steps for both buyers and sellers.

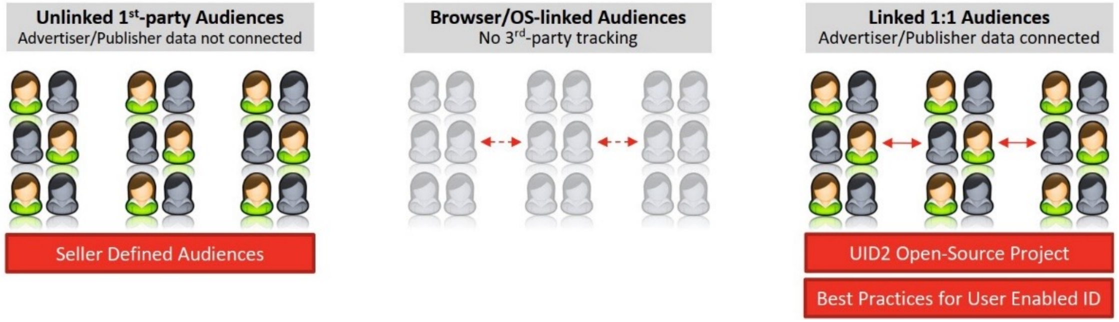
contents.

introduction.

As a result of the ongoing deprecation of third-party cookies, recent privacy features in Apple's iOS environments reducing IDFA volumes and the resulting negative impact on the mechanics of digital advertising, in March 2021 IAB Tech Lab released a set of four standards which draw from existing standards in-market and included Best Practices for User-Enabled Identity Tokens Guidelines.

These specifications have drawn a clearer focus on how one of the (three) future-proof approaches will be the use of identity service to competently link 1:1 audiences between publishers and advertisers.

These services have traditionally kept the open internet free and to consumers through advertising and moving forwards will require either an explicitly opted-in device-based ID or secure, user-enabled ID from a login, email, etc. potentially connected to a clean room approach.



Unlinked 1st-party Audiences
Advertiser/Publisher data not connected
Seller Defined Audiences
Browser/OS-linked Audiences
No 3rd-party tracking
Linked 1:1 Audiences
Advertiser/Publisher data connected
UID2 Open-Source Project
Best Practices for User Enabled ID

image source: IAB Tech Lab

They will also have to be secured and have highly transparent uses offering consumers a full suite of privacy-focused controls.

starting point.

As a starting point for this explainer guide, we should define the meaning and details around 'identity' and 'identifiers', as these terms are often used liberally to refer to similar things.

what are identifiers?

We'll start with identifiers, which are prolific and consumers can easily have hundreds or thousands of identifiers across all of their browsers and devices.

We align to the IAB Tech Lab view that identifiers come in 3 types - **consumer, creative assets, and the business entities** involved in the supply chain. These identifiers are the core building blocks that help fight fraud, improve brand safety, deliver a better experience to consumers, and support measurement and attribution.

IDENTIFIERS FOR EFFECTIVE ADVERTISING

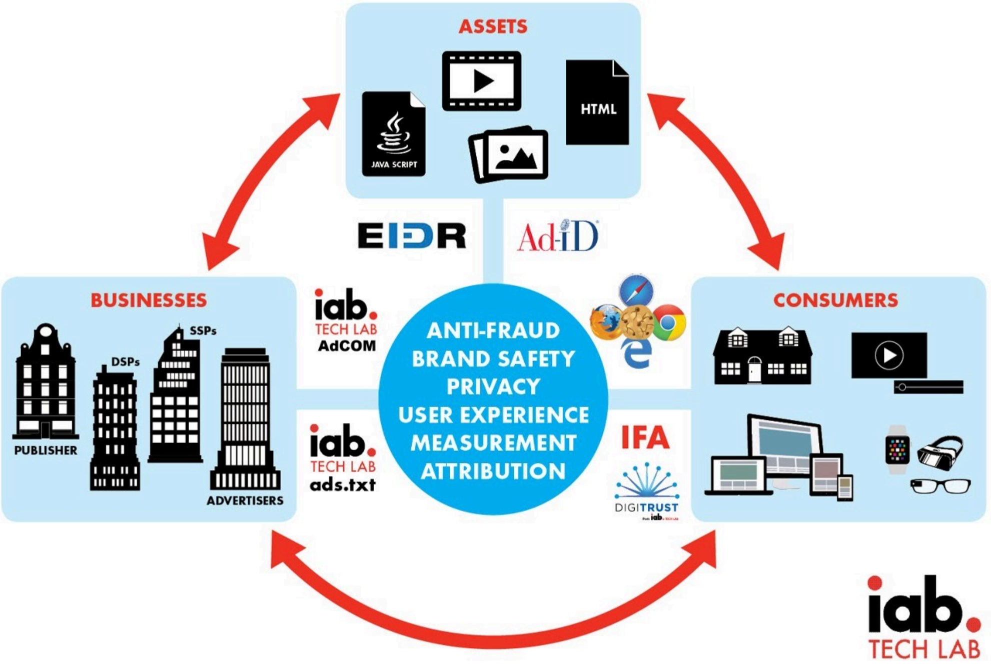


image source: IAB Tech Lab

types of identifiers.

We align to the IAB Tech Lab view that identifiers come in 3 types - **consumer, creative assets, and the business entities** involved in the supply chain.

identifiers by type.

Consumer IDs

Identify individual users or a group of people within a household (all generally anonymously), but may ultimately be tied to devices or browsers, depending on the available data (such as logins) on various platforms. Examples are cookies, DeviceIDs or IFAs (Identifier For Advertising) on mobile and/or OTT (Over-The-Top video) devices. These are utilised for the purpose of understanding user behaviours and interests for targeting and personalisation, assessing where/when a person saw an ad (for measurement and attribution), and applying known privacy preferences consistently across sites, apps, and devices. This allows platforms to develop insights into users' needs and deliver a better experience by providing more relevant ads.

Asset IDs

Identify creative assets as they go through the advertising supply chain, to make it easier to understand what was or will be shown to a consumer, ensuring that the right content is delivered to the right individual (separating ads from competitors, age appropriate etc.), and enabling accurate measurement/tracking of which creatives were displayed where and who they were presented to. Asset IDs are also important to help with brand safety by tying the ID to metadata about the creatives.

Business IDs

Identify the various companies such as publishers, advertisers, and vendors that provide content and ads to consumers, and execute a range of other functions across the advertising supply chain. These IDs are used mainly to manage trust, reduce fraud, and improve transparency.

identity management.

As consumers can have so many different identifiers across all their various browsers and devices, the real effort is in being able to manage and leverage these into a genuinely useful version for marketers and publishers.

Cookies, for instance, have been a very useful short-term identifier as they can be used (for now) and shared fairly widely, but often for the purposes of true identity management they are simply too inferred and degrade too quickly.

However, through the unification of the various identifiers and appended data points a persistent Individual ID can be created, which can be utilised as being a more meaningful and shareable and unified profile for each individual.

This process is often referred to as Identity Resolution and the tools used to align the various identifiers and store them is known as an Identity Graph.

These persistent customer identities can also be supplemented with other data, including offline to build out a fuller and more accurate anonymised profiles to target, campaign manage and measure online advertising.

Ultimately, all the various device-level identifiers are merely an enabler of advanced identity resolution solutions.

examples of different identifiers.

IDFA
GAID
Cookies
Hashed email addresses
Phone numbers
CRM Data
Financial Transactions

different identifier examples.

IDFA

Apple's Identifier for Advertisers is a device identifier assigned to users of Apple devices. It is used for anonymous tracking and identification of consumers.

GAID

Google's advertising ID is a device identifier assigned to users of Android devices. It is used for anonymous tracking and identification of consumers.

MAIDs

General term for a mobile advertiser ID, which is a unique string of digits identifying a specific mobile device.

Cookies

Also called HTTP cookies, internet cookies, or browser cookies, are files stored on the hard drive of computers designed to hold a small, specific amount of data about a particular website or client. Their primary purpose is to identify the user so his or her web experience can be customised and to streamline the online surfing process by saving certain information such as email, home address, shipping information, username or password, or interests. We define 1st, 2nd and 3rd party further on.

Hashed email addresses

As most people keep their personal email addresses forever and are ubiquitous, making emails a key future identifier in the future of digital marketing.

Hashing is a method of encrypting data, such as email addresses, into a hexadecimal string. Each email has its own unique hexadecimal string, made up of 32-character codes, that remains consistent no matter where the email is used as a login and is unique to each email address. This code cannot be reversed, making it completely anonymous. See also the definition of tokenization further on – including a comparison of tokenization vs. encryption.

Phone numbers

Mobile phone numbers, in particular, have more recently been commonly used as a consumer identifier for two-factor authentication as these phone numbers have increasingly become more reliably attached to individuals over the long term. As a sensitive piece of personally identifiable information (PII) data, phone numbers should also be securely hashed.

different identifier examples.

CRM Data

A customer relationship management (CRM) database is all of the data about consumers that businesses capture and store within a CRM system. The different types of data stored inside a CRM database include data points such as: contact name, title, email address, social profiles, contact history, lead scoring, order history and personality traits. Most of this is personally identifiable information (PII) data and therefore very sensitive.

Financial Transactions

Similar to standard CRM data, however this also includes personally identifiable information (PII) data such as age, job, nationality, social status, place of residence - as well as income and consumption data, size of incomes, size of debts, different sources of income, credit/debit card purchases, standard inter-bank payments, loans and deposits.

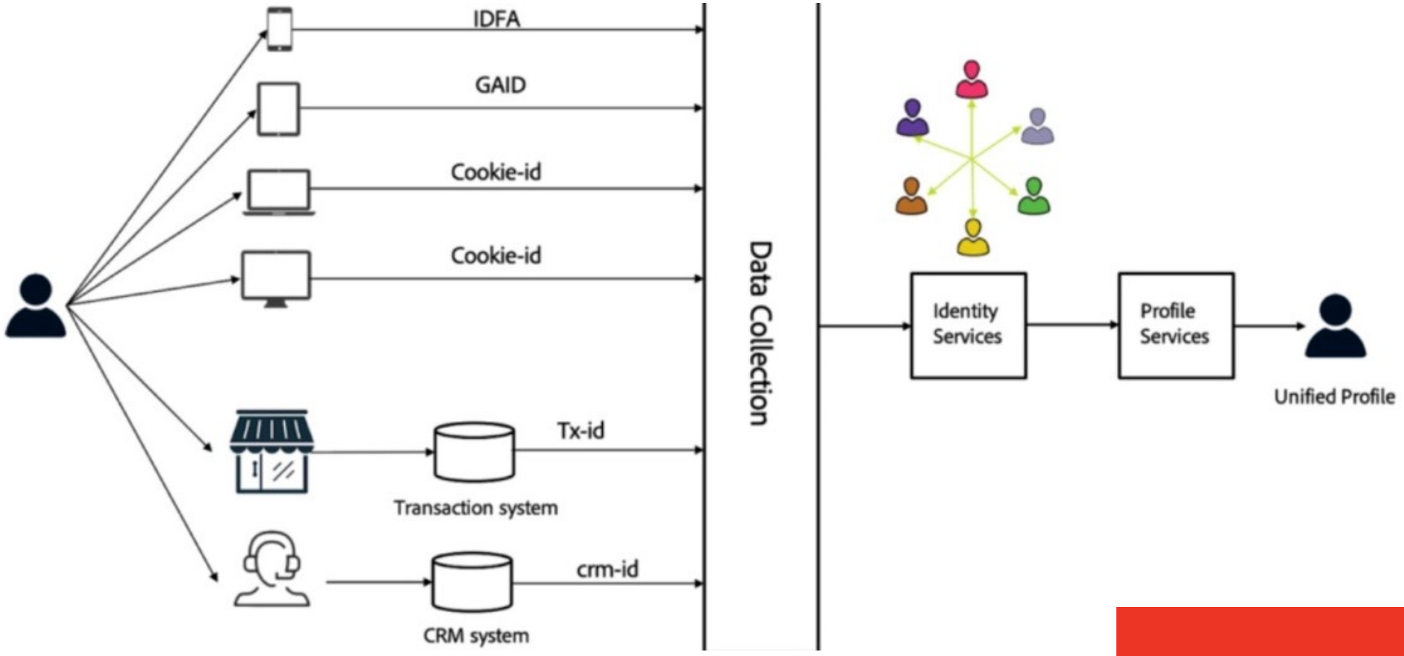


image source: IAB Australia Data Handbook

identity resolution.

An identity resolution solution can permanently keep track of any preferences and consent signals from identifier to identifier, as well at each data touchpoint - thereby enabling effective and actionable insights for marketing and personalisation.

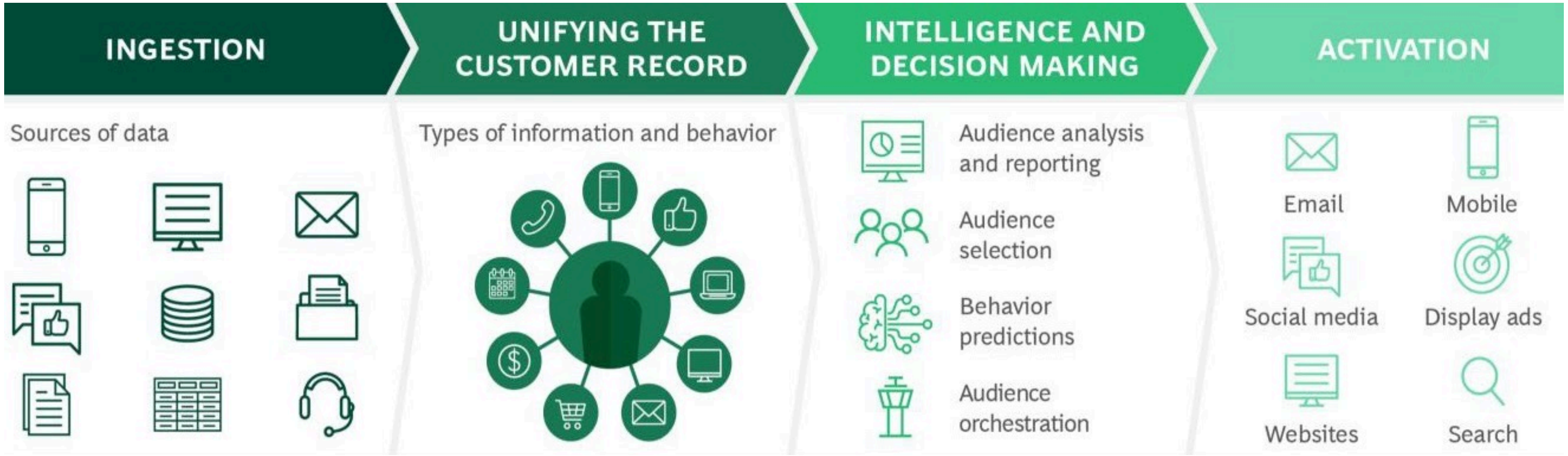


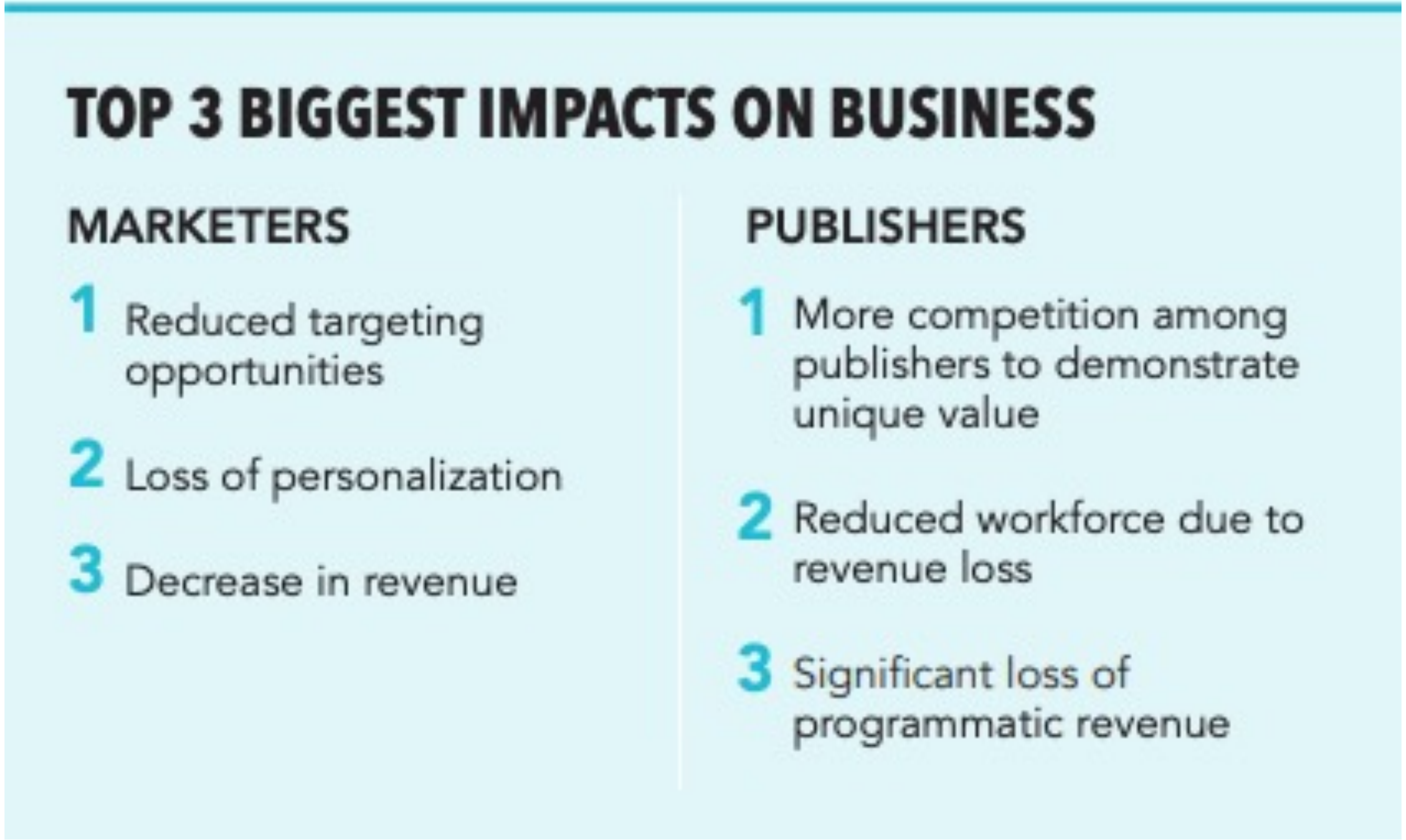
image source: CDP Institute

This enables management, via a single customer view, a consistent consumer experience as the updates and privacy controls can be made more seamless and persistent. With privacy reviews underway here in Australia this is increasingly a critical capability.

The unified approach is also the most effective technically for consistently and cleanly managing a consumer's digital privacy.

deprecation of third-party cookies.

Google will phase out third-party cookies in its Chrome browser sometime in 2023. Since 2017, other popular browsers such as Apple's Safari and Mozilla's Firefox have already implemented some default blocking against third-party tracking cookies.



TOP 3 BIGGEST IMPACTS ON BUSINESS

MARKETERS

1 Reduced targeting opportunities

2 Loss of personalization

3 Decrease in revenue

PUBLISHERS

1 More competition among publishers to demonstrate unique value

2 Reduced workforce due to revenue loss

3 Significant loss of programmatic revenue

source: Lotame global survey report "Beyond the Cookie Part 2" (October 2021)

deterministic & probabilistic matching.

deterministic data.

Deterministic data is linked to something which identifies a user, like an email address or a cookie ID, and has a high likelihood of being authentic.

Deterministic data provides a reliable foundation for marketing operations as it is based on factual variables or declared consumer inputs, and a key advantage of deterministic matching is the resulting high degree of accuracy.

However, not all applications and websites require users to login or provide specific information and it can lack scale for smaller publishers.

probabilistic data.

Probabilistic data is data based on behavioural events such as online browsing behaviours and click-throughs.

Probabilistic data can be analysed and grouped by the likelihood that a user belongs to a certain demographic, socio-economic status or class.

Algorithms are often leveraged to help in these processes and interrogate these behavioural patterns, device types and touchpoints to generate inferred interests or help to determine the probability of the user’s age, gender or socio-economic status.

deterministic & probabilistic matching.

deterministic data.

The most prevalent method to deterministically match users is via email addresses.

Emails tend to be unique to consumers and can be identified and matched across a very wide range of data sets. Large data owners such as popular platforms (Facebook, Twitter, LinkedIn etc.) can deterministically match with ease, as they regularly require users to sign in with an email address and authenticate to access their services via various devices.

Most vendors will regularly provide their overall match rates - and some may provide customisable match algorithms or confidence scores highlighting how likely the matches are accurate, based on their specific first-party customer data and data quality profiles.

probabilistic data.

Probabilistic matching isn't as accurate as deterministic matching but **can utilise deterministic data sets to help refine the algorithmic efforts and improve the accuracy of insights.**

One of the core advantages of using probabilistic matching over deterministic matching is scale, as you don't need to collect authenticated email addresses or other pieces of personal data to be able to identify them across different devices. It's also safer from a consumer privacy perspective.

However, there are some disadvantages such as the accuracy of any resulting outputs and the lack of transparency in matching methodologies, as any algorithms used are often proprietary. This is especially valid if you are relying solely on probabilistic matching to identify, track, and target users across different devices and applications.

7 considerations for buyers and sellers.

one

Thoroughly review your strategic requirements, both in the short and medium-long term.

two

Fully review and understand the data assets that you have access to.

Consider whether you will need to leverage 2nd & 3rd party audiences from elsewhere - or if you have enough scale and quality to leverage only your own 1st party assets.

three

Collaboratively agree upon the product & technical capabilities required to achieve your aims & work with your current technical vendors to understand their capabilities.

four

Ensure that you consider both the technical capabilities required as well as scale, as both will be important in practice for the future.

five

Consider fully reviewing any other options in-market that could meet your defined needs, benchmark your current solutions and have competent experts asses the options.

six

Effective RFPs (request for proposal) only request relevant information and provide ample information about your brand and any potential identity resolution needs. It should always reflect and align to your high-level strategic goals and KPIs.

seven

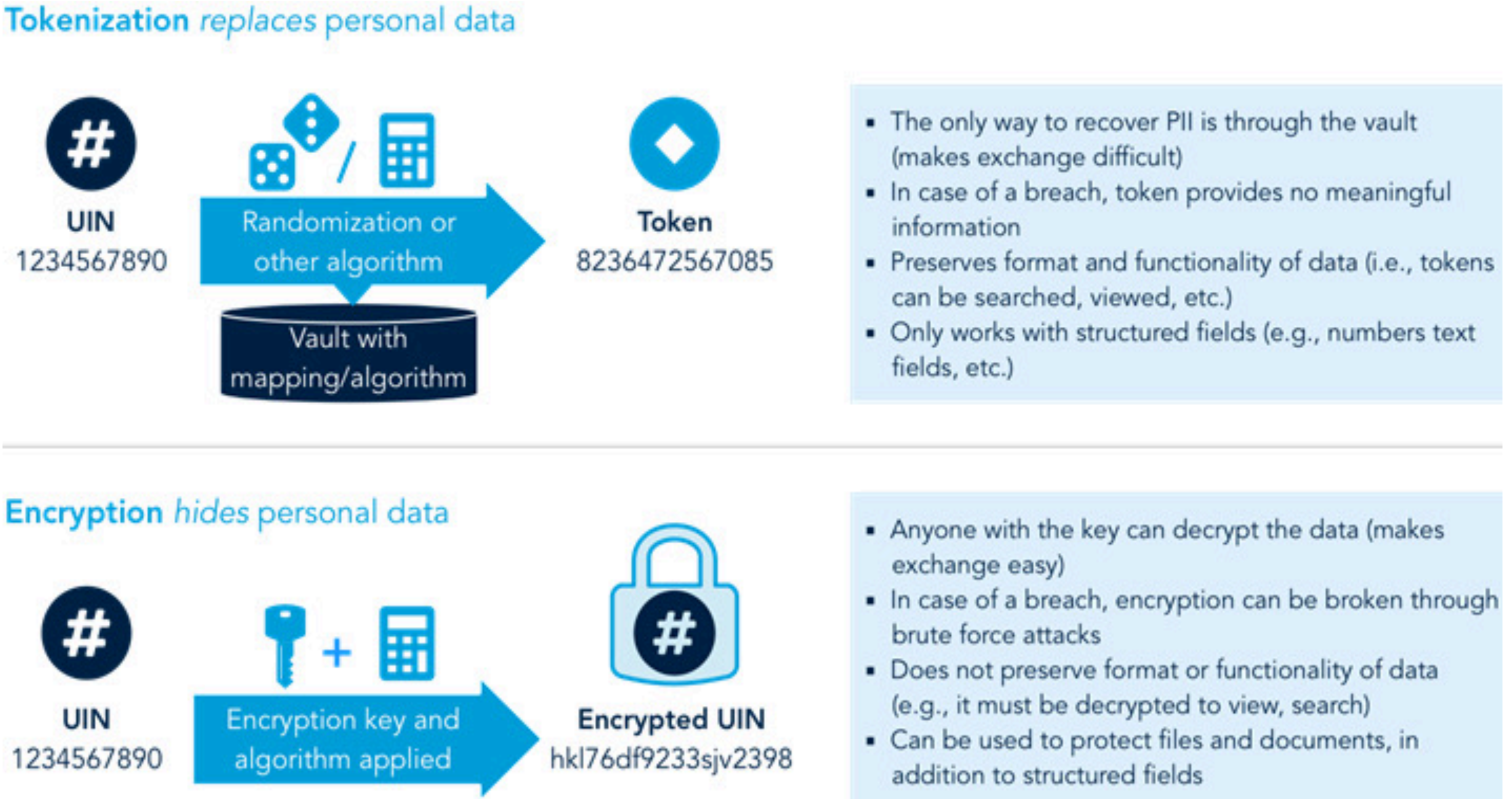
Consult fully on all of the related privacy requirements, future risks and considerations, across all the markets you are active in - both now and on an ongoing basis moving forwards.

tokenization vs. encryption.

tokenization

increasingly it is becoming clear that legislators are taking very seriously the potential risks related to both the exposure and breach of any PII in the ad tech eco-system. As a result, the Tokenization of ID's is being seen as more future proof approach.

Both tokenization and encryption obscure personal data – however, tokenization is potentially operationally simpler and cheaper to implement than encryption, as decrypting the data isn't required in order to use it. Platforms such as Infosum and Karlsgate are examples of vendors enabling this approach and Unified ID 2.0 is preferring this approach as an ID solution.


Tokenization replaces personal data
UIN
1234567890
Randomization or other algorithm
Vault with mapping/algorithm
Token
8236472567085
Encryption hides personal data
UIN
1234567890
Encryption key and algorithm applied
Encrypted UIN
hkl76df9233sjv2398

- The only way to recover PII is through the vault (makes exchange difficult)
- In case of a breach, token provides no meaningful information
- Preserves format and functionality of data (i.e., tokens can be searched, viewed, etc.)
- Only works with structured fields (e.g., numbers text fields, etc.)

- Anyone with the key can decrypt the data (makes exchange easy)
- In case of a breach, encryption can be broken through brute force attacks
- Does not preserve format or functionality of data (e.g., it must be decrypted to view, search)
- Can be used to protect files and documents, in addition to structured fields

image source: The World Bank Group

glossary of terms.

first party data

this data is collected directly from consumers interacting with your assets, both online and offline. For example, in-store purchasing habits stored in a CRM or behaviours people exhibit on your website.

second party data

is first-party data that you're getting directly from another source and isn't unique to your brand. For example, an airline (like Qantas) might team up with a global accommodation supplier (like Airbnb) to share their first-party data sets with one another.

third party data

is collected from an external source that doesn't have a direct relationship with the people it's collecting data about. For example, a third-party data company might pay publishers to put their pixel on their site and then use that information to piece together online profiles.

zero party data

is an increasingly used term for first-party data which consumers are intentionally and openly opting into sharing with business entities. Technically and legislatively, this remains first-party data and any distinction is effectively redundant, other than for product marketing purposes.

data clean rooms

these solutions enable businesses with legitimate access to large amounts of first-party data to safely and securely keep and update that information without directly accessing the raw data, negating the risk of ever exposing the consumer data. Thereafter, other companies (second parties) can bring their own data and campaign insights from one clean room and compare it with another, again without directly exposing customer data.

Most companies offering clean rooms can also run machine learning modelling over data sets, giving advertisers actionable insights into their audiences without the personalised data ever being revealed.

glossary of terms.

cookies

also known as an HTTP cookie, web cookie, or browser cookie, this is a string of text sent from a web server to a user's browser that the browser is expected to send back to the web server in subsequent interactions.

A cookie has a few core attributes: the cookie value, the domain and path within which it is valid, and the cookie expiry. There are other attributes as well that limit the cookie to https-only transactions or hide it from JavaScript.

PII

this is personally identifiable information and refers to information used or intended to be used to identify a particular individual, including name, address, telephone number, email address, financial account number, and government-issued identifier.

CDPs

with all of the efforts in the advance of third-party cookie deprecation for digital marketing the evolution of solutions required have been shifting from traditional DMPs over to the fuller suite of capabilities that we see in Customer Data Platforms (CDPs). CDPs differ from DMPs as they will usually use persistent 1st party data rather than 2nd and 3rd party data which is not retained for as long.

Additionally, and critically CDPs will collect and manage sensitive personal consumer data in the form of personally identifiable information (PII) – whilst DMPs hold anonymous data most typically in the form of cookies.

Due to the sensitivity of the data being handled it is critical to ensure that you have access to the right levels of resources in terms of technical skills, analytical skills and legal advice. The related complexity as a result of the increased capabilities and responsibilities are significant versus traditional CRM systems and DMP platforms.

identity providers.

Provider	ID Solution	Data Sources	Base Identifiers	Consent Type	Availability and Addressability	Interoperability	Prerequisites
Criteo	Criteo Graph	Global publishers, advertisers & data suppliers	Probabilistic & deterministic data	1st, 2nd & 3rd party	Yes, via Criteo Media Platform	Yes via RampID & Unified ID 2.0	Participation in Criteo's First-Party Data Collective
Equifax	IXI	Financial partners & data suppliers	Hashed emails, financial transactions, phone numbers & postal addresses	1st, 2nd & 3rd party	Yes, via all major DSPs & SSPs	Interoperable with most global identifiers	Strict prevetting process
Experian	MarketingConnect	Financial partners & data suppliers	Hashed emails, financial transactions, phone numbers & postal addresses	Authenticated & consentual 1st party	Yes, via all major DSPs & SSPs	Interoperable with most global identifiers	Strict prevetting process
Eyeota	Eyeota	Global publishers & data suppliers	Probabilistic & deterministic data	1st, 2nd & 3rd party	Yes, via Eyeota Translate	Yes, via Eyeota Translate	Must have a common identifier within any datasets
Google	Customer Match	Owned & Operated	Hashed emails, phone numbers & postal addresses	Authenticated & consensual 1st party	Search, the Shopping tab, YouTube, Gmail and Display	*TBC*	All Customer Match customers are vetted with very clear requirements
ID5	ID5 Universal ID	Global publishers	Probabilistic & deterministic data	1st, 2nd & 3rd party	Yes, via all major DSPs & SSPs	Via Unified ID 2.0	Can meet GDPR compliance requirements
InMobi	UnifID	Global publishers & data suppliers	Probabilistic data	2nd & 3rd party	Yes, via all major DSPs & SSPs	Via Unified ID 2.0	Ability to sync
Lifesight	Lifesight CIP & Life ID	Global publishers, financial partners & data suppliers	Probabilistic & deterministic data	2nd & 3rd party	Yes, via all major DSPs & SSPs	Via Unified ID 2.0	Ability to sync

identity providers.

Provider	ID Solution	Data Sources	Base Identifiers	Consent Type	Availability and Addressability	Interoperability	Prerequisites
LiveRamp	ATS & RampID	Global publishers & data suppliers	Hashed emails	1st, 2nd & 3rd party	Yes, via all major DSPs & SSPs	Interoperable with most global identifiers	Publishers must have access to user authentications
Lotame	Panorama	Global publishers & data suppliers	Probabilistic & deterministic data	1st, 2nd & 3rd party	Yes	Interoperable with most global identifiers	Ability to sync
Meta	Facebook Custom Audiences	Owned & Operated	Hashed emails, phone numbers & postal addresses	Authenticated & consensual 1st party	Only across owned & operated	*TBC*	All Custom Audiences customers are vetted with very clear requirements
Near	Proxima	Global publishers + online & offline data partners	Hashed emails, phone numbers and home address	1st, 2nd & 3rd party	Yes, via Near Allspark	Yes, via Near Allspark	Must have a common identifier within any datasets
Oracle Data Cloud	Oracle ID Graph	Global publishers & data suppliers	Probabilistic & deterministic data	1st, 2nd & 3rd party	Yes, via all major DSPs & SSPs	Via Unified ID 2.0	Ability to sync
Unified ID 2.0	Unified ID 2.0	Global publishers	Hashed emails, which are encrypted via a tokenization solution	Authenticated and consensual 1st party	Yes	Interoperable with most global identifiers	Must agree to abide by UID2 ecosystem terms. Source code donated by The Trade Desk
Yahoo	ConnectID	Owned & Operated	Hashed emails, tokenized	1st, 2nd & 3rd party	Yes, via Yahoo Preferred Network (prev 'Gemini') + Yahoo DSP & SSP	Interoperable with most global identifiers	Publishers or brand must have mechanis.m for gathering user emails

descriptions for the identity providers matrix.

This **matrix** of ID providers has been collated and completed by the IAB Australia Data Council.
An explanation of each of the columns and the information they contain is below:

Provider

which entity owns the ID solution?

ID Solution

what is the product name of the ID solution, offered by the provider?

Consent Type

what is the relationship to the entity that has gathered consumer consent for the utilisation of these signals (i.e. is it a first, second and/or third party relationship)?

Data Sources

from where the data is sourced and what is the status of ownership?

Availability and Addressability

how, or within which types of activation platforms, is the solution made available for the purposes of addressability for the purposes of digital marketing?

Base Identifiers

from which types of identifiers are these solutions built upon?

Interoperability

is this solution interoperable with other identity solutions in-market?

Prerequisites

are there any requirements, limitations or considerations for potential customers that are reviewing this solution?

media agency solutions.

All the major media agency holding groups are looking to provide end-to-end data and identity solutions to their clients, and ultimately want to ensure that they can also enable privacy-safe data integrations with other platforms and ad-tech companies.

These solutions are either as a result of in-house development or acquisitions – and some examples of these are below. Please contact the relevant media agencies for more information on the related capabilities and how they may be able to help.

Provider	ID Solution
Dentsu	Merkle M1
GroupM	Choreograph ID
IPG	Kinesso (based upon Acxiom)
Omnicom Group	Omni ID
Publicis / Epsilon	Epsilon People Cloud / CORE ID

other global providers.

Adobe

Adara

Adstra

AlikeAudience

Amperity

Audience Project

BiGDBM

Britepool

Crimtan

Datonics

DigiCenter

FullContact

Infutor

IRI

mParticle

Media Wallah

Neustar

OneData

Retargetly

Salesforce

Semcasting

ShareThis

SirData

TailTarget

The ADEX

Throtle

TrasUnion

Treasure Data

TrueData

Valassis

Weborama

Zeotap

iab tech lab transparency standards.

IAB Tech Lab has recently released draft specs to enable yet further transparency to the supply chain by enabling publishers, agencies and brands to declare the identity services they work with.

The approach aligns with the other transparency standards, in the .json file being stored at the root domain for easy manual access by humans and readable by machines & crawlers (as you can with *ads.txt, sellers.json* and *buyers.json*).

By knowing exactly which identifiers are used by which publishers, *id-sources.json* should help provide greater clarity into ID adoption.

Brands and agencies could map out which publishers work with which identifiers and match these against their own audiences, which will themselves be associated with different identifiers. This will help them figure out where in the publishing landscape they can locate their addressable audiences.

the *id-sources.json* standard aims to:

Provide a standard way for companies to declare which user identity sources they use.

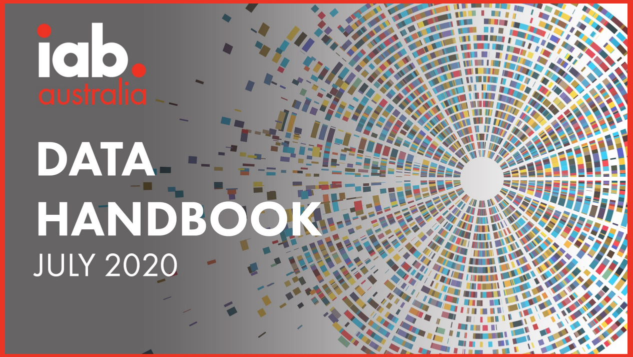
Work like the other supply-chain transparency standards as a participant hosted, structured declaration that machines can read.

Ease ad campaign execution between advertisers, publishers, and their chosen technology providers by making it explicitly clear who supports what.

For the full set of specifications simply click here

further reading.

data handbook



iab. australia

DATA HANDBOOK

JULY 2020

project rearc

IAB Tech Lab:
Standards for Responsible Addressability and Predictable Privacy (Project Rearc)

beyond the cookie

iab. australia

BEYOND THE COOKIE:
MAPPING THE FUTURE OF MARKETING MEASUREMENT

thankyou.

iab.
australia
data advertising