2023

data collaboration platforms explainer

iab. australia

Find your section

Thanks to our contributors

Danny Tyrrell
Co-Founder
DataCo Technologies

Rick Knott
ANZ General Manager
InfoSum

Chloe Hunter
Operations Lead, Enterprise Enablement ANZ
LiveRamp

Simon Pereira
Chief Commercial Officer
XPON Technologies

Vishal Shah
Lead Solutions Engineer
MiQ

David Raitt
GM Marketing Solutions
Near

Jonas Jaanimagi
Technology Lead
IAB Australia

Introduction

The IAB Australia Data Council has been diligently working to deliver meaningful guidance, comprehensive definitions, and best practices in the realm of compliant data collaboration for the purposes of digital advertising.

In recent times, the most prominent product within this category has been Data Clean Rooms. The council deemed it more essential to offer insights into the broader spectrum of products that facilitate data collaboration - whilst also presenting localised updates on each of these (including Clean Rooms).

Also, in relation to Clean Rooms, we have drawn upon the exemplary work conducted by the IAB Tech Lab in their recently published **'Data Clean Rooms: Guidance and Recommended Practices'** document (released in February 2023), which we wholeheartedly endorse for all interested parties.

To access this excellent piece of work from IAB Tech Lab, simply click here

Before delving into the diverse array of product types for data collaboration, it is prudent to examine the various methodologies of data matching as a key element of this introduction.

It all starts with a business challenge

Typically, the first step before getting into data collaboration and matching is identifying why you want to match data, and it comes down to the objectives you're trying to achieve, i.e. a clear business challenge. Once this is clear, what can help is a set of areas and competencies that a brand can evaluate for themselves with respect to data:

Availability

This tries to answer the question **"What data do you have and how usable is it?"** It may turn out that you do not possess the data for the problem you wish to solve, and that could be one of the key drivers for data collaboration, either internally or externally.

Architecture

This looks at aspects like data structure, unity, security, and governance. You may have all the data you need, but if it is sitting there silo-ed and unstructured, it is unable to reach its full potential, eventually pointing to the need for collaboration.

Analysis

Advanced analytics and data science techniques help you uncover stories your data is trying to tell you, but knowing which type of technique or match to apply is equally crucial. Equality-based or Deterministic matching may make sense in some cases, whereas you may need to venture into Probabilistic or machine-learning based matching in certain others.

Application

Identifying use cases for applying your models is crucial, and the applications of one single collaboration can be many. Asking the question "Will this collaboration help us make data-driven decisions across multiple levels of the business?" is important.

Activation

Lastly, the strongest data matching exercise is wasted if it is not allowed to be applied in the real world to learn and evolve over time. Activation and automating intelligent media buying is the final step towards measurement of your data collaboration strategy.

Once you conduct an in-depth review of the above areas, you are in a much better position to identify the type of collaboration and matching you want to go for.

Data Matching

Increasingly, companies are turning to leaders in data analytics to help them solve the most common authentication, activation and attribution challenges:

1. **How can you help me collaborate with different data sources and overcome challenges with siloed data?**
2. **With limited access to data analytic teams or with limited skills in my marketing team how can I leverage 1st, 2nd or 3rd party data?**
3. **When faced with data of varying quality, with incomplete records or outdated content, where do I begin?**

Data analytics companies are quickly scaling to solve these challenges, engineered with consumer privacy at the forefront. 'Privacy by Design' practices ensure that any data shared is consent based, encrypted and anonymised with built-in processes to forget and purge user data on requests.

What is the first step in matching data?

In its simplest form a match rate represents the percentage of the records from company A that match to the customer records of company B. They can be used to understand compatible overlap between you and the partner you are collaborating with, for targeting or suppression. In practical day-to-day terms, it is usually a measure of how likely marketers are to reach a consumer with any given vendor. This can be with an identity provider, data enrichment provider, direct with media owners or buying platforms, or any other second-party data use cases.

While the industry average for matching data is 25% in Australia, specialist companies are pulling ahead and securing match rates north of 65% but don't be fooled by huge match rates, the business outcome and value derived from the data match [partnership] should be your primary objective.

Whilst everything starts with match rates, it is worth noting not all match rates are the same. In some ways, match rates are to data collaboration what click-through rates were to the early days of internet advertising.

Whilst meaningful to a degree, they are also an opaque, misunderstood and over-simplified methodology for inferring value. Match rates are the beginning of the story, not the end goal. A high match rate does not automatically mean the campaign performance will be good. A lower match rate does not automatically mean the campaign performance will be weak. Accuracy and the relevance of the match matters. Using high-quality data to generate the match matters. Having a clear data strategy in place and then using the match as a step to enable it matters.

Match rates are frequently used as a north star metric to determine the potential reach of campaigns and often even as a proxy for match 'quality'. However, a match rate percentage alone lacks context about the process or the quality of the data within. Furthermore, it is not the only indicator of the potential performance of a data partnership. It is essential to understand how a match is calculated and with what data, especially when evaluating different opportunities, so that you can compare like for like. To help determine that correlation as you test and learn, it is always worth comparing actual campaign results to the initial match rate and match methodology.

Matching directly or indirectly?

There are two methods for matching with another company's dataset - either directly or indirectly. Clean rooms and data collaboration platforms provide an opportunity for direct identifier matching across single or multiple identifiers. Whereas for those who don't have access to clean room environments, third party vendors can provide indirect data matching solutions.

Direct data matching between companies like brands and media owners, empower marketers to understand their audience and confidently tailor marketing in a relevant and considered way. Typically deterministic identifiers are used to confirm the match in some form of true/false test like private-set intersection. The matching is one-to-one against a known audience, so the accuracy is extremely high.

Indirect data matching through third party vendors are solutions for those who lack the tools (e.g. data clean rooms) to conduct the match for themselves. Two companies can share their first party data and commingle into a third-party location. Here the vendor can access and configure the match against its own identity spine. This centralised approach is typically simple and easy to use for a customer, but does require all participating parties to be comfortable sharing, exposing and releasing control of the data, into the third party. Furthermore, it is worth noting that the match rates are completed through the vendor's own proprietary and private methodology.

This can mean a lack of transparency around some key contextual variables, such as the type of data used, precision level, household composition, lookback window, type of match, or audience expansion used.

If the intention of the data match is to run a campaign programmatically, you must ensure the match rate is reported on IDs that are addressable in the bidstream (and not internal identifiers). Otherwise match rates can be easily and unintentionally inflated, as there can be two or three match rates that need to be calculated:

- **the offline match with the identity provider's offline graph;**
- **the match to their digital graph; and**
- **the actual match with the media owner or bidding platform dataset.**

This final match calculation is the most valuable as it determines true reach and scale so it is essential that it is exposed in the initial match test.

Therefore, high initial match rates aren't always equal to high reach or high potential. Discarding opportunities based on the partial picture that match rates paint can prevent organisations from achieving their goals, not to mention the loss of time and money.

To evaluate any match rate, both direct and indirect, marketers need to understand:

1. **Their strategic goal and the role the match rate plays vs. other metrics, such as reach**
2. **The quality and composition of the datasets being used to match**
3. **The precision level (individual/household etc.) and the freshness of the data**
4. **The difference between match tests and actual campaign performance like ROI, ROAS, CPC, and CPA**
5. **When to apply deterministic or probabilistic audience expansion**

To further complicate things, the advertising industry is currently in flux. Globally, match rates are being impacted by technology changes and a lack of high-quality consented data. Furthermore, the disappearance of third-party identifiers, the rise of privacy legislations and the emergence of new media channels such as retail media, connected TV, audio and more are further challenging match rates through third parties and driving the use of data collaboration platforms. Marketers should therefore:

- Have a clear strategy and business goal in mind with 'privacy' at its core
- Prioritise working directly with partners who have superior consent management
- Focus on using high-quality data vs. high match rates to drive the best results
- Ensure you know exactly how the match is enabled
- Create actionable matched segments
- Increase reach using audience expansion tools and leveraging ID vendors tactically

Deterministic vs. Probabilistic Data Matching

Deterministic data is a single identifier which identifies a user, like an email address or a cookie ID, and has a high likelihood of being true as it is based upon factual variables or declared consumer inputs. However, not all applications and websites require users to login or provide specific information and it can lack scale for smaller publishers.

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 datasets. 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.

Probabilistic data is data based upon more than one data input such as behavioural events, online browsing activity and click-throughs. It is also often the combination of more than one deterministic data point, to create new probabilistic understanding. This data can be analysed and grouped by the likelihood that a user belongs to a certain demographic, socio-economic status or class.

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

Probabilistic matching isn't as true as deterministic matching, but this should not be confused with it not being as accurate. Probabilistic data can often react much faster to a change in location, interest or intent, and can help to identify when there are different users of a digital product, than the declared deterministic identifier would have you believe. 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 can also be safer from a consumer privacy perspective.

Of course though, probabilistic by its very nature, has the potential for false-positives. Furthermore, any defined outputs are often unverifiable as they are usually defined by proprietary technology that lacks transparency in matching methodologies and algorithms. This is especially valid if you are relying solely on probabilistic matching to identify, track, and target users across different devices and applications.

What is Data Collaboration?

Put simply, data collaboration uses technology to combine and analyse data sets within an organisation or with partners to enable a wide range of use cases, from uncovering new consumer insights and enabling accurate cross-screen measurement to expanding reach and creating brand-building media networks.

Data collaboration can be intracompany (connecting data silos), cross-brand, cross-industry, cross-cloud, and more.

In defining in more detail what data collaboration is we have separated the key capabilities into six core topics and will take each in turn.

We simply explain each topic in turn to help provide insights and guidance, without any specific recommendations into the types of data collaboration solutions that are available. Best practices and more detailed recommendations follow in a later section.

1. **Enabling data assets internally**
2. **Integrating and enriching data assets with external partners**
3. **Authentication, quality assurance & validation**
4. **Generating value from data assets via activation**
5. **Measurement & attribution**
6. **Regulatory compliance, privacy & security**

1. Enabling data assets internally

We have broken this section down into four key areas, assuming that the business cases and strategies have been well defined, and the business is fully committed and holistically supportive of those plans.

Data Identification and Assessment

Understanding the different sources and types of data that exist internally is a critical starting point. To effectively assess these data assets, the first step is to conduct an inventory of existing assets and determine it is all currently being utilised. This is crucial because many businesses struggle to enable their data due to a lack of understanding of where all their data is located internally. In some cases, work may be required to update and maintain a registry of all data inventory to uncover everything cleanly. Also you may have to look at breaking down internal data silos that have been unknowingly hoarding useful data assets for some time.

Internal data generally comes in these three forms:

Structured data

Is highly organised and exists in predefined formats such as web logs, CRM data, e-commerce transactions and GPS coordinates.

Unstructured data

Exists in the form it was generated and lacks a well-defined structure. Examples are social media posts, audio/video assets, images and webpage content.

Semi-structured data

Is a mix of structured and unstructured data that has some structure, but it does not conform to a strict data model or schema.

Once all the data has been identified, management should explore and categorise its key attributes. This will help build a better understanding of the nature of the data and how it can be leveraged for the planned business purposes.

Establish Data Governance

A framework for data governance should be considered and planned for early in the process. A good governance framework should establish the processes needed to dictate how the data is collected, stored and used. This will help ensure that any data being assessed is accurate, relevant, and compliant with privacy laws and regulations.

Thereafter this will allow a business to start identifying which data is sensitive, establishing controls to prevent unauthorised access and creating controls to audit those who have access and building systems to enforce governance rules and protocols. These steps can help secure and protect data to ensure regulatory compliance and help the business to trust its data and the related processes.

Construct a Competent Data Infrastructure

Building a data infrastructure will allow you to collect, store, and analyse data more effectively, and may involve investing in data management tools, such as customer relationship management (CRM) software or a data analytics platform.

Alongside the technical infrastructure a cultural effort is required to ensure that you have people with the right skills and experience required to meet the business objectives. You also need to support that talent through clear communication, a sense of purpose and a sense of ownership over the strategy. Often the biggest challenge to becoming a data-driven organisation is related to people and the associated change management requirements, not just the technological needs.

Data Ingestion

Establishing the right type of ingestion process and what technology is required to meet your needs based upon the infrastructure is the next step. Fully review the different types of solutions and associated platforms that can meet your needs. This may also require some ancillary features such as a tag management solution (TMS) for instance to collect all the signals from your owned and operated digital assets.

Generally, the ingestion process involves three steps known as ETL - extraction, transformation and loading.

Data extraction

Data is taken from its originating location.

Data transformation

Data is cleansed and normalised for business use.

Data loading

Data is moved into a database, data warehouse or data lake to be accessed for use.

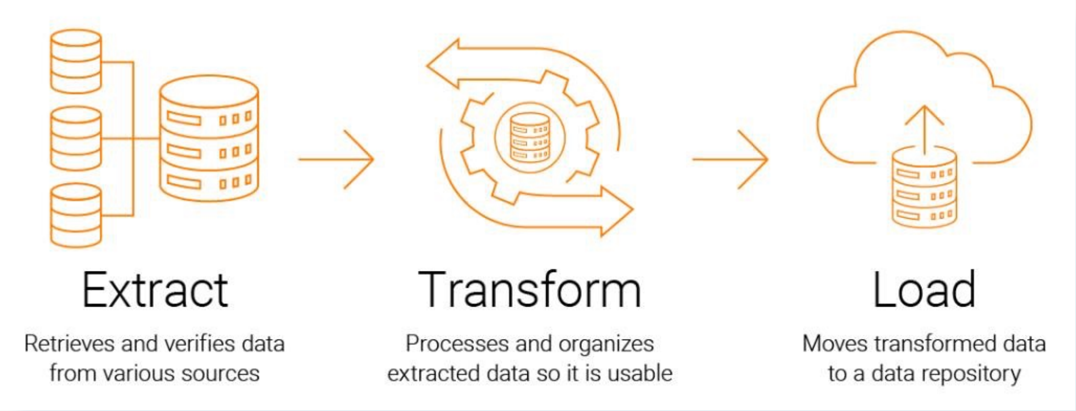

Extract
Retrieves and verifies data from various sources
Transform
Processes and organizes extracted data so it is usable
Load
Moves transformed data to a data repository

Image source: informatica.com

Data management teams will often face additional considerations and requirements at each of these steps, such as how to ensure the data they've identified for use is reliable and how to prepare it for practical use. Be prepared to review the processes of data collection and data governance as use cases emerge and your data program matures, identifying what data sets are missing from any data collection process and what collected data sets hold no value.

Also work to automate the process as much as possible from data ingestion to cataloguing to ensure efficiency and speed as well as adherence to the protocols established by the governance program. This will also help to implement tools that uncover problems in the data collection process, such as any data sets that don't show up as expected.

Develop data-driven Insights

Being able to competently and quickly start analysing your data to develop insights that can inform your strategy is obviously key. This may involve segmenting your audience based on demographics, behaviours, or other factors, or using predictive modelling and analytics to further enhance any actionable insights.

Once you have started to generate insights from the data, you need to communicate those insights in a way that is understandable and actionable. The ability to tell 'meaningful stories' through the data is a powerful tool. This may involve creating data visualisations, such as charts or graphs, or presenting the insights in a report or dashboard.

Finally, it's important to monitor the performance of your data-driven insights and iterate on them as necessary. This may involve updating your data sources, refining your analytics techniques, or adjusting your business objectives.

2. Enriching data assets with external partners

Data enrichment is the process of collaborating and adding either first-party or third-party data to a dataset you're already working with so as to make more informed business decisions. It also allows businesses to make data useful and reliable for end users, empowering them to answer a bigger set of questions, as you have more information available and can also answer questions with deeper insights than with un-enriched data.

When businesses enrich their data assets, they are adding value to it by making it more useful. Businesses also generate a greater understanding of their customers when they enrich data, allowing them to tailor products and services to their customers' needs. This is why marketing is one of the most prevalent applications of data enrichment as it solves key challenges by providing additional attributes about current customers and identifying new audiences that are likely to become high-value customers.
Data enrichment can transform a business's basic customer data into a more complete picture of demographic, geographic, psychographic and purchasing behaviours.

Before getting started with data enrichment, it's important to establish your overall goal in enriching your data. The overall goal is to improve your data's quality and accuracy, but this is often too broad a starting key performance indicator in most cases.
When establishing your data enrichment goal, you should consider what additional information or data attributes your business will need to collect and what data types would be most relevant to achieve your aims.

To accomplish this, a business will look to work with data from a trusted outside source to add raw consumer data to their existing data assets. For optimal results, it's important to work with a partner with a large data repository of clean data devoid of duplicates, or fake profiles, that can provide ongoing assistance and analysis. By working with a strong data enrichment partner, customer profiles can be consistently updated to ensure customer data stays clean, relevant, and useful. Marketers and data analysts can build more complete consumer profiles that evolve as the customer does.

Traditionally the main data assets to be enriched in terms of attributes are:

Demographic

Enriching demographic data allows you to target messaging to specific demographic groups. This aids businesses in ensuring advertisements and messaging are relatable to the consumer.

Behavioural

When you enrich behavioural data, you are adding customer behavioural patterns to their user profile. Adding behavioural patterns to a profile allows you to identify the areas of interest for that customer, as well as their journey leading up to their overall purchase decision. It is important to enrich behavioural data as it aids you in determining the effectiveness of advertising campaigns and justifying marketing budgets.

Geographic

Businesses that enrich geographic data can target messaging to different geographic groups, ensuring that users see content that is more relevant to their location.

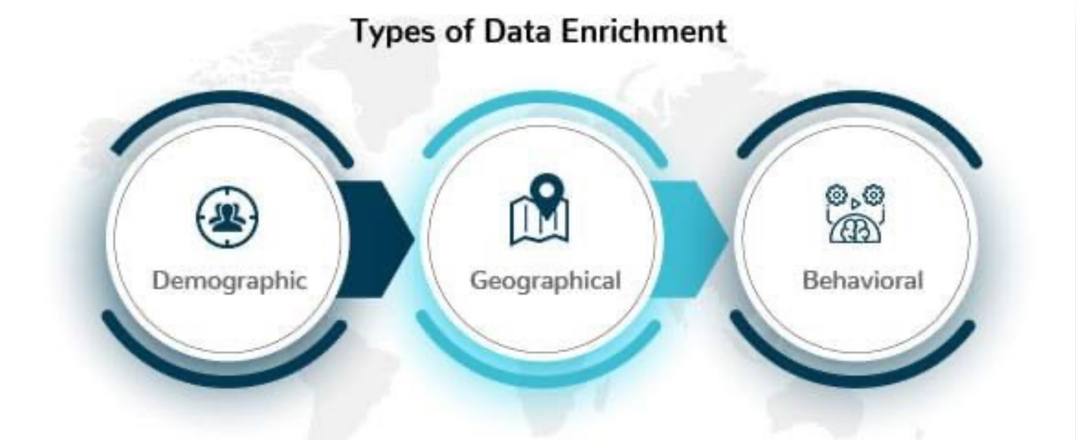


Image source: audienceplay.com

Data enrichment isn't a task that is done once as a set and forget. It's an ongoing process to ensure your business works with data that has high value and it's crucial to ensure data is up to date and continuously enriched. If data is not kept up to date, over time, data will naturally decay and lose its value to the point where it becomes worthless.

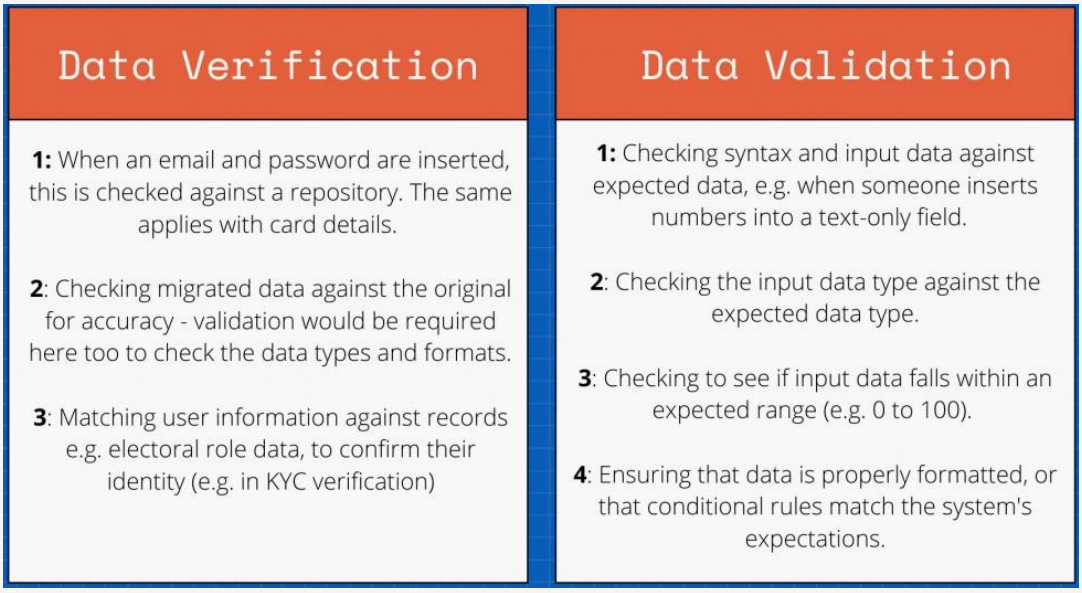
3. Authentication, Quality Assurance & Validation

Having previously established a competent governance framework allows you to have protocols for ongoing quality assurance processes and practices.

Firstly, ensuring that any data collected is done in accordance with the procedures and that the data stored in the registry database meet the requisite standards of quality, which are generally defined based on the intended purposes. Thereafter processes for data validation and verification can be established.

Technically validation is the process of determining whether a particular piece of information falls within the acceptable range of values for a given field.

Data verification, on the other hand, is actually quite different from data validation. Verification performs a check of the current data to ensure that it is accurate, consistent, and reflects its intended purpose.



Data Verification	Data Validation
1: When an email and password are inserted, this is checked against a repository. The same applies with card details.	**1:** Checking syntax and input data against expected data, e.g. when someone inserts numbers into a text-only field.
2: Checking migrated data against the original for accuracy - validation would be required here too to check the data types and formats.	**2**: Checking the input data type against the expected data type.
3: Matching user information against records e.g. electoral role data, to confirm their identity (e.g. in KYC verification)	**3**: Checking to see if input data falls within an expected range (e.g. 0 to 100).
	4: Ensuring that data is properly formatted, or that conditional rules match the system's expectations.

Image source: understandingdata.com

Verification may also happen at any time. In other words, verification may take place as part of a recurring data quality process, whereas validation typically occurs when a record is initially created or updated.

Verification plays an especially critical role when data is migrated or merged from outside data sources. Consider the case of a company that has just acquired a small competitor. They have decided to merge the acquired competitor's customer data into their own billing system. As part of the migration process, it is important to verify that records came over properly from the source system.

Small errors in preparing data for migration can sometimes result in big problems. If a key field in the customer master record is assigned incorrectly (for example, if a range of cells in a spreadsheet was inadvertently shifted up or down when the data was being prepared), it could result in shipping addresses or outstanding invoices being assigned to the wrong customer.

Therefore, it's important to verify the information in the destination system matches the information from the source system. This can be done by sampling data from both the source and destination systems to manually verify accuracy, or it can involve automated processes that perform full verification of the imported data, matching all of the records and flagging exceptions.

Verification is not limited to data migration. It also plays an important role in ensuring the accuracy and consistency of corporate data over time.

4. Generating value from data assets via activation

Ultimately there are a number of ways that businesses can generate value from data such as improving decision making, improving operational efficiencies and unearthing new revenue streams. Activation ultimately is the usage of data to reach target audiences, develop new models, and deepen the understanding of relevant audiences. Some of these opportunities may require integrations with additional services, such as DSPs or SSPs.

We provide a general list of some of these below, with more of a focus specifically on the opportunities related to advertising.

Making better data-driven decisions

Through analysing data, businesses can identify patterns and insights that can inform their decision-making process, leading to better informed and more effective business strategies.

Improving operational efficiency

Data analysis can help businesses identify inefficiencies in their processes and operations, enabling them to streamline operations and reduce costs.

Enhancing customer experience

Leveraging customer data can help businesses to gain better insights into their customers' preferences and behaviours, allowing them to offer more personalised and targeted experiences.

New products and services

Data can enable businesses to identify new product or service opportunities, innovate and bring new offerings to the market.

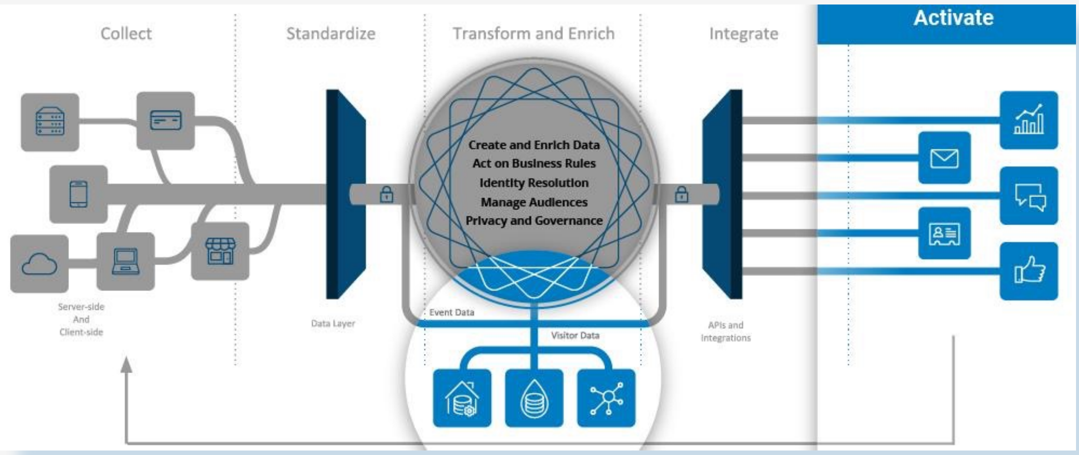


Image source: tealium.com

Audience targeting

By leveraging data such as demographics, psychographics, and behavioural patterns, businesses can target the right audience with their advertising campaigns, leading to better performance and improved returns on investment.

Data enriched supply

Publishers can enrich their supply to meet the needs of marketers for both direct and programmatic campaigns.

Audience extension

Using data assets that are attractive to advertisers, publishers can help brands access their audiences away from that publisher's owned and operated media.

Monetising data as a stand-alone product

Businesses can also look to safely enable their data assets for other parties to enrich their inventory or buying. Selling access to data to one or more partners that seek to drive customer behaviours or gain greater insights in a privacy-centric manner.

Personalisation

Data insights can help businesses understand their customers' preferences and behaviours, allowing them to deliver tailored messages and offers that resonate with their audience and drive higher engagement.

A/B testing

Data can be used to conduct A/B testing, allowing businesses to compare the performance of different ad creatives, headlines, and calls to action, and optimise campaigns accordingly.

Improved RTB

Data can be used to inform real-time bidding strategies, enabling advertisers to bid on ad inventory based on the likelihood of a user converting, leading to more effective and efficient ad spend.

Attribution modelling

By analysing data across multiple touchpoints and channels, businesses can better attribute their advertising efforts to specific outcomes and optimise their campaigns to maximise their impact.

Predictive analytics

Data can be used to predict future trends and behaviours, helping businesses to proactively respond to market changes and improve their forecasting abilities.

5. Measurement & Attribution

Not all data collaboration platforms will necessarily have measurement capabilities. Obviously analytics & audience measurement platforms will and it is common for DMPs, CDPs & increasingly Clean Rooms to have some capabilities.

Most measurement calculations require impression/exposure data from different sources to be joined and matched with conversion/sales data from advertiser's data sources and often with one or more parties requiring personal data to be shared and transferred directly to another party. This risks exposure of personal data as well as advertiser's business information.

Moving forwards there is a growing expectation that data collaboration platforms can provide organisations with matching capabilities to quantify effectiveness and performance of marketing and advertising campaigns without requiring any data to be shared, commingled, or transferred to other parties.

Below are some capabilities that are worth considering, depending upon requirements. It is worth being aware that these platforms may have some discrepancies or different capabilities to any current measurement tools currently in use.

Data integration and consolidation

A core data management capability, enabling you to collect and consolidate data from various sources - including online and offline sources, to provide a holistic view of customer behaviours.

Audience Segmentation

Segment audiences based on various criteria such as demographics, interests, and behaviours. Based on these characteristics advertisers can look to tailor campaigns to specific groups of users and optimise ad targeting.

Lookalike modelling

Lookalike modelling allows you to identify users that share similar characteristics and behaviours to existing customers, allowing advertisers to expand their reach to new audiences.

Tracking customer interactions

Collect data from various sources such as websites, mobile apps, email, social media, and offline channels to track customer interactions across all touchpoints. This data can include clicks, views, purchases, and other behaviours.

Cross-device tracking

Allows you to track user behaviour across multiple devices, including desktops, tablets, and mobile devices - to provide a seamless user experience across devices and better understand user behaviours.

Attribution analysis & modelling

Perform real-time or offline attribution actions to identify which touch point across the various channels and campaigns was responsible for the conversion without exposing user level conversion/engagement data.

Campaign optimisation

Work to provide real-time data on the performance of advertising campaigns, including impressions, clicks, and conversions – enabling advertisers to tweak campaigns based on data-driven insights and improve ROI.

Multi-channel campaign reporting

Work to generate insights into the performance of marketing campaigns across various channels such as email, SMS, push notifications, and social media. This should include key metrics such as open rates, click-through rates, conversion rates, and revenues generated.

A/B testing

Test different versions of marketing campaigns and measure their impact on customer behaviour.

Incremental lift measurement

A solution that calculates the incremental impact of a campaign on metrics such as sales without exposing actual individual customer purchase data.

Reach and frequency

Calculate the number of individuals or audience that saw an advertisement and at what frequency over a period of time. Campaign and audience verification - identify if a campaign reached the right audience on the right channel at the right time.

ROI/ROAS analysis

Apply actual or predictive models of calculating return across the media mix to optimise campaign performance and budget allocation based on aggregated-level insights only without sharing individual conversions, purchases or other activities.

Real-time reporting and alerts

Real-time reporting and alerts on customer behaviour, campaign performance, and other key metrics can enable you to quickly identify and respond to changes in customer behaviour or campaign effectiveness.

Customised insights

Data analysis on customer behaviours and campaigns performance to improve advertising strategies and optimise campaigns.

Custom reporting

Custom reporting capabilities allow businesses to generate reports on specific metrics and KPIs related to advertising measurement and attribution with genuine flexibility.

6. Regulatory compliance, privacy & security

Regulatory compliance is an essential aspect of any data collaboration and all participating entities must ensure that any data sharing and collaboration adhere to the various and ever-evolving global laws, regulations, and industry standards designed to protect data privacy and security. Failure to comply with these regulations can result in legal and financial penalties, as well as damage to an organisation's reputation.

A key requirement is that data governance controls are available in order to ensure the data is protected in all stages of collaboration, based upon the governance framework of those utilising the technology. It's also advisable to train employees on data privacy and security best practices, including how to handle sensitive data, how to identify and report security incidents, and how to comply with relevant laws and regulations.

Data Collaboration Platforms are also expected to integrate a range of privacy and security capabilities that prevent parties from directly accessing an individual's personal data. There are a number of privacy-enhancing technologies (PETs) that are employed to ensure that the privacy of any sensitive personal data in the dataset are protected.

Broadly, PETs are technologies that offer businesses the ability to accelerate safe data collaboration, build customer intelligence, and maximise the value of data without relinquishing control or compromising consumer security & privacy.

One point to note is terms such as privacy, security and encryption are often used interchangeably. Whilst these terms are related, often with large areas of overlap, they are also distinct in that one does not necessarily solve for the other. For example, it is possible to have great security, but poor privacy. This is because privacy can be both technology based and non-technology based.

Encryption is a great example of this as whilst it is both a privacy and security technology, many privacy legislations around the world still consider encrypted PII to be PII. Therefore, encryption changes nothing regarding how you must handle this data from a liability standpoint.

Examples of both security & privacy capabilities are below:

Encryption

Is a technology to convert or scramble plain text data into a format that is unintelligible, which can only be understood by decrypting or reversing the encryption.

Homomorphic Encryption

Is a form of encryption which allows a party to perform transformations on data while it is still encrypted.

Secure Multi-party Computation

Is a technology that enables multiple parties to perform a computation and reveal results and insights, whilst still keeping their data private from one another.

Private Set Intersection

Is a cryptography technology that allows two or more parties with datasets to learn about the common data attributes between those two datasets without exposing the data (other than the common attributes), through a series of tests.

Federated Learning

Is a machine learning technology that allows multiple parties to locally perform the part of a computation relevant to their data, which is then aggregated to infer the overall results and insights. Federation is a product's ability to connect multiple data systems across heterogeneous environments. For example, federated systems can query across two or more data warehouses.

Trusted Execution Environments

Is a hardware technology that isolates a computation from the host system to keep the data and state private from all parties and yet still infer overall results and insights.

Synthetic data

Is a technique that artificially creates data that is similar to the original data to reduce the risk of individual personal data from being exposed or re-engineered.

Pseudonymisation

Is a technique that removes personal data in a dataset by replacing it with an unintelligible number. It is typically done by applying hash functions, salted hash functions, encryption or double blinding

Noise Injection

Is a technique that adds random irrelevant data into a dataset to obfuscate an individual's data or make an individual's data statistically irrelevant

Differential privacy

Is a mathematical technique to rigorously guarantee a specific level of privacy for an operation by using any combination of tools like redaction, rounding or injecting noise. Processes and protocols that protect the underlying individuals or households from re-identification, reconstruction, or tracing.

K-Anonymity

Also known as cohort sizing, is a property of an anonymized data set which makes it much more difficult that an individual can be re-identified. Typically it requires determining a minimum cohort size based on the characteristics of a data set.

With regards to security, the related capabilities of data collaboration technologies are critical to ensure that sensitive data is always protected during collecting, sharing and collaboration. Businesses should carefully evaluate the security features of data collaboration technologies before adopting them to ensure that they meet their expected security requirements, and some additional considerations for this are listed below:

Access and permission controls

Robust access controls that allow administrators to control who has access to what data, for what purpose, at what granularity, and for what duration preventing any unauthorised or preapproved access to sensitive data.

Audit trails

Allow administrators to track who has accessed what data and when, helping to identify potential security incidents and retain a record of data access for compliance and audit purposes.

Data loss prevention

Are capabilities that can detect and prevent the unauthorised transfer of any sensitive data, and would normally include features such as data classification, content filtering, and activity monitoring.

Secure communication tools

Built-in tools, such as secure messaging and file sharing, ensure that sensitive data is not inadvertently shared with unauthorised users when users are working with the technologies.

Another key topic is the testing of the technical robustness of any technology should be a standard part of any comprehensive review of a tech platform or solution.

Typically, this process involves “penetration testing”, which is a method of testing the security of IT systems, networks, and applications by simulating a real-world cyber attack. The objective of a penetration test is to identify vulnerabilities that can be exploited by attackers and provide recommendations for improving the security of the tested systems.

During a penetration test, a team of skilled security professionals, commonly referred to as ethical hackers, will attempt to breach the security of the target system by exploiting known vulnerabilities, misconfigurations, and weaknesses in the system’s defences. The process includes several steps, including reconnaissance,
scanning, exploitation, and reporting.

In the reconnaissance phase, the team gathers information about the target system, such as IP addresses, domain names, and other publicly available information that can be used to identify potential vulnerabilities. The scanning phase involves using various tools and techniques to scan the target system for vulnerabilities and weaknesses. The exploitation phase involves attempting to exploit the identified vulnerabilities to gain unauthorised access to the target system. Finally, the reporting phase involves documenting the findings and providing recommendations for improving the security of the tested systems.

Penetration testing is a vital component of a comprehensive request for proposal (RFP) for technology, as it helps identify weaknesses in IT systems that could be exploited by attackers. It can also assist organisations in meeting compliance requirements and enhancing their overall security posture.

Lastly, we wanted to provide a simple explainer of the difference between encryption and de-identification - as this has become a much discussed topic recently in Australia with the ongoing privacy review. Encryption and de-identification are both methods used to protect sensitive information, but they differ in their approach and the level of security they provide.

Encryption involves converting readable data (e.g., plaintext) into an unreadable form (ciphertext) using a mathematical algorithm and a secret key. The ciphertext can only be decrypted and read by someone who possesses the correct key. Encryption is commonly used to protect data in transit and at rest. For instance, when you use a secure website to enter your credit card information, that data is encrypted before being transmitted over the internet.

In contrast, de-identification is a process of removing or changing identifying information in a dataset so that the remaining data can no longer be linked to an individual. De-identification is used to protect privacy and confidentiality. Various methods of de-identification are available, including removing personal identifiers such as name or address and modifying specific data elements to make it difficult or impossible to re-identify an individual.

The key difference between encryption and de-identification is that encryption protects data by rendering it unreadable, while de-identification protects data by removing or altering identifying information. Although encryption provides a higher level of security as it makes the data unreadable even if an unauthorised user gains access to it, de-identification relies on the assumption that the modified data cannot be re-identified. However, de-identification is often used in situations where it is not feasible or practical to use encryption, such as when sharing data for research purposes.

The Key Benefits of Data Collaboration

Advertisers

Can leverage new and effective strategies to connect with their customers, suppress their customers or use their customer data as a seed for lookalike modelling to acquire new customers.

Privacy

Clean Rooms can help with adherence to the privacy and governance policies of the organisations and local legislation.

Publishers

Can maximise the value of their audience in a post cookie world with loss of traditional web and device identifiers

Data vendors

Can bring their assembled datasets to a marketplace for monetisation through collaboration for data enrichment, expansion and activation

First and Second Party Data collaboration

Clean Rooms enable businesses to extract greater value and operationalise their first-party data without exposing or sharing proprietary data with other parties.

Security

Clean Rooms ensure that data is not altered or manipulated during the analysis process but remains fully protected from internal or outside attacks, misuse or leakage.

Different types of Data Collaboration Platforms

There are a number of types of collaboration solutions available and it is a fast evolving product space. We try to cover the most popular ones in the list below:

Online/Offline Data Matching

As touched upon in the introduction, matching is a key aspect of any data collaboration. Matching services have been around for a while now, utilising information about customers obtained offline (such as through direct mail and surveys) and bringing these online as actionable assets via on-boarding partners, who can integrate this information with robust online profiles.

This gives marketers the ability to build consumer group demographic profiles based on the audience's various online behaviours, which allows for advertising that is more likely to generate engagement due to increased relevance.

Data Management Platforms (DMPs)

DMPs emerged as the first type of platform that could collect and enable data from a wide range of different sources and make them actionable, traditionally as cookie pools. DMPs collect, store and manage online and offline data sets to gain insights before creating actionable segments to target digital campaigns. For internal data, DMPs might pull from CRM software or from company-owned channels like websites or email.

For external data, DMPs might connect to third-party data brokers or corporate partners. DMPs will work primarily with anonymous behavioural data such as cookies, device IDs, and IP addresses generated from pages of websites.

Once a DMP has gathered the data, they will segment it and build profiles of each type of user customer based upon pre-defined behavioural rules for what users are doing when visiting. DMPs can also generate inferred look-alike profiles of other users that share similar behavioural attributes, to increase the scale of potentially relevant users. DMPs then share these actionable insights on audiences with the ad server enabling them to deliver against these segments of users. Be aware that connecting these audiences into the ad server can both limit and/or increase the available audiences for certain campaigns - depending upon the targeting attributes that are selected.

As mentioned, the most prevalent ID used for this type of platform is cookies, which has made these platforms less popular for any future-proof approach as cookies are becoming increasingly limited functionally. DMPs are also less capable of managing PII, making them less robust and future-proof legislatively with the ongoing data privacy changes.

Customer Data Platforms (CDPs)

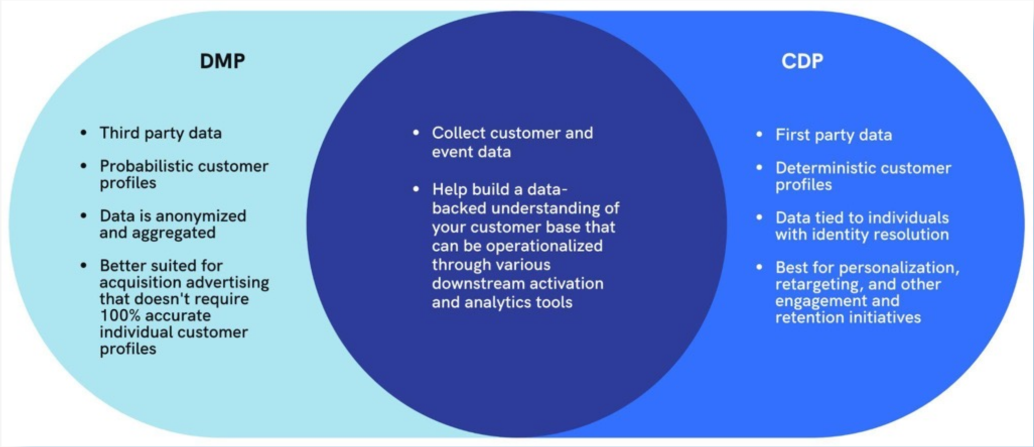
CDPs are similar to DMPs, but focus specifically on first-party data or customer data. They are capable of collecting data from multiple online and offline interactions and matching them to a single customer profile. One main feature is they can profile interactions from anonymous customers and retroactively tie that data to a customer once it is identified. A CDP typically can only manage known 1st party consented data.

CDPs can store the same information as DMPs, but also very detailed and sensitive information on people's profiles and behaviours aggregated from both online and offline sources. These are often generated from purchase transactions, Customer Relationship Management (CRM) database tools or filled forms and can contain data such as purchase transactions, postal addresses, email addresses and phone numbers as well as numerous web behaviours - and very often containing sensitive PII (personally identifiable information).

Any customer data collected by a CDP is attributed to a persistent profile. Profiles can inform future interactions based on that specific customer's history. For the purposes of execution, CDPs will have to attempt to authenticate any users they have access to online and then also (often via DMPs) make those users addressable via cookie-based advertising platforms.

Whereas DMPs tend to retain data for only a short amount of time (max 90 days or so), CDPs can retain it for much longer. Also, certain CDPs merge anonymous and known activity to a single profile once a user converts, allowing publishers to gain a more complete view of the user journey whilst DMP audiences and segments are built for advertising and tend to follow a more pre-determined rigid structure.

For a quick comparison between DMPs and CDPs see below:



DMP

- Third party data
- Probabilistic customer profiles
- Data is anonymized and aggregated
- Better suited for acquisition advertising that doesn't require 100% accurate individual customer profiles

- Collect customer and event data
- Help build a data-backed understanding of your customer base that can be operationalized through various downstream activation and analytics tools

CDP

- First party data
- Deterministic customer profiles
- Data tied to individuals with identity resolution
- Best for personalization, retargeting, and other engagement and retention initiatives

Image source: mparticle.com

In terms of core functionality

CDPs have four primary purposes. The first is to centralise and unify data, the second is to add value to the data through enrichment - which can involve incorporating 2nd/3rd party data or segmentation modelling. The third function is to explore and provide intelligence, allowing for the surfacing of audience insights and high-value segments for targeting. Finally, CDPs enable activation, which can be done either directly via the platform or through pushing audience segments to digital activation platforms such as DSPs or offline sources like direct mail houses or email and short message service providers (EMS/SMS).

The use of statistical, machine learning, and AI models is also important to assist in surfacing audience segments that can be activated for better targeting and engagement, thereby driving customer value. Technical elements such as data portability, data storage, ID resolution, security, and governance frameworks are essential to enable audience discovery and activation, they all ultimately serve to support the primary goal of driving improved marketing performance for brands.

Identity Resolution

Identity resolution is the near-real-time process of connecting hundreds of identifiers used by different channels, platforms, and devices. Through the unification of the various identifiers and appended data points a persistent Individual or household ID can be created, which can be utilised as being a more meaningful and shareable and unified profile for each individual, family unit, or household.

It enables marketers, along with supporting agencies, technology platforms, data owners, and publishers - to tie them back to the same person or household using deterministic, probabilistic, or a hybrid approach for people-based targeting, measurement, and personalisation.

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 pseudonymised profiles to target, campaign manage and measure online advertising.

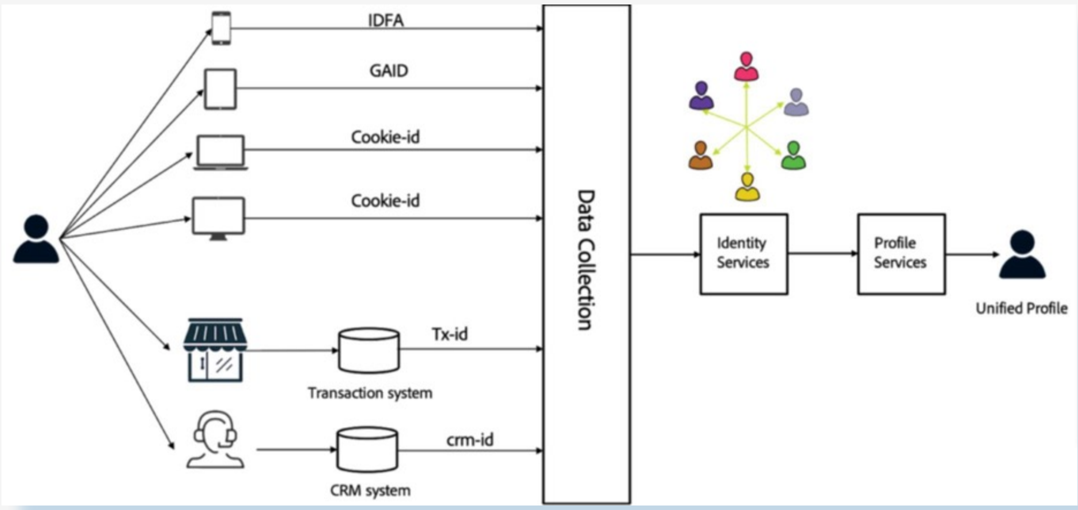


Image source: adobe.com

Data Clean Rooms

Clean rooms are a safe and neutral environment where multiple datasets can be connected for various marketing use cases, including analysis, enrichment, activation, and measurement. A data clean room provides a secure environment where multiple data sources are matched and analysed, in compliance with privacy regulations, while allowing each participant to maintain privacy and control of their data. The safety and security of data combined with the power and intelligence of multi-party computation enable companies to instantly match and analyse data across unlimited datasets in real-time whilst still eliminating the risk of exposure, leakage, or misuse.

A publisher's user-level first-party data can be ingested from CRM systems (including historical data) into this secure environment. Any other data sources including historical and current transaction data can also be made available in the clean room environment for a variety of use cases. Any sensitive data sent to the clean room is encrypted for transmission and as soon as it enters the clean room it is secured, protecting it from any unauthorised access.

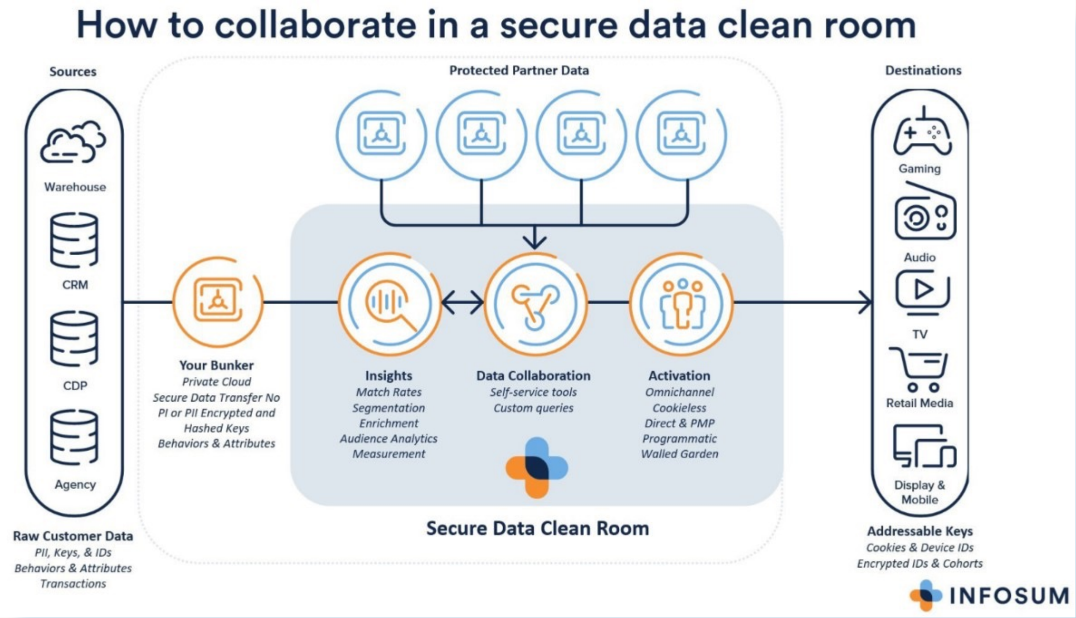


Image source: infosum.com

Publishers retain full control over the environment and approved partners can get a feed with encrypted data out as an output. Publishers and other organisations can then look to connect their data within a clean room to other high-quality data partners to maximise the scale, accuracy, and performance of all data-driven strategies including audience planning, activation, and measurement. The resulting enriched and fully anonymised data can then be shared in a privacy-centric way with approved buyers, and other partners, as required.

Clean Rooms share characteristics common to other data collaboration platforms (such as data preparation, data normalisation, data matching or querying and data outputs), but what differentiates Clean Rooms from other types of data collaboration solutions is the emphasis on security and privacy.

Regarding security, Clean Rooms provide a combination of access control, governance protocols and system designs to maintain the confidentiality, integrity, and availability of data used in any collaborations. Each party contributing data within a clean room has independant and holistic control over what data is being connected, with what parties, at what granularity, for what purpose, and for what duration.

To better enable privacy, Clean Rooms provide the technology and controls that can uphold the privacy requirements of data contributors. This is achieved by not revealing or exposing personal data of a party’s data to other parties involved in the operations through the application of various privacy enhancing technology (PETs) such as differential privacy techniques, noise injection, and k-animity. These baseline protocols ensure total protection and de-identification of the underlying data subjects privacy even if data was to be exposed from within the data clean room.

Consent Management Platforms (CMPs)

A Consent Management Platform (CMP) centralises and manages transparency for the consent and objection of the users of a website. The CMP can read and update the data collection purposes of a company that participates in the delivery of digital advertising (commonly referred to as a vendor) within a publisher's website, app, or other digital content.

Vendors can declare their purposes for accessing a user's device or browser or processing their personal data within a publicly available list - often referred to as a Global Vendor List (GVL).

Typically CMPs perform the following tasks:

- **Provides users with transparency into the vendors that a publisher has chosen to work with.**
- **Provides transparency into the purposes that a vendor wishes to leverage consumer consent.**
- **Stores a user's consent signals (for example a third-party cookie) in the user's browser and makes this consent information available to vendors in different jurisdictions.**
- **Ensures that consent for a purpose applies only to the vendors that have declared (e.g. via a GVL) that they will only use data for that purpose specified.**

Measurement & Analytics

Audience measurement platforms provide advertisers with insights into how their target audience interacts with their ads. They collect data on audience demographics, behaviours, and preferences, which advertisers can use to refine their targeting strategies.

Analytics platforms meanwhile provide advertisers with tools for measuring the effectiveness of their advertising campaigns. They use data analysis and visualisation techniques to help marketers identify trends, insights, and opportunities for optimisation.

Use case and working example

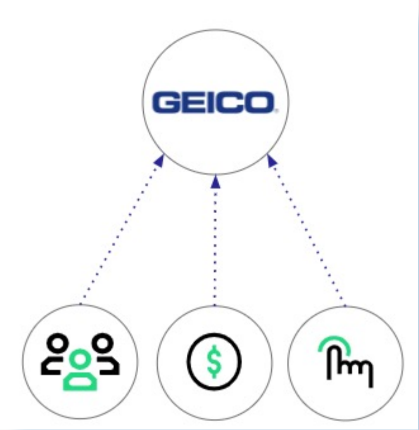


Image source: liveramp.com

Collaboration can look a number of different ways within the advertising industry.

Internal data collaboration

Within a given organisation, there may be different teams with different access or ownership of data regarding their consumers and related activity. They are often tasked to bring all the different data points together for a complete, holistic picture. In the example below, a car insurance brand has three different data points:

1. **A list of lapsed consumers.**
2. **A list of new consumers who have recently purchased insurance from the car insurance brand.**
3. **A list of consumers who have recently visited the car insurance brand's website.**

While all these data points vary in nature and ownership, it might be beneficial for the car insurance brand to access and often join data together for a deeper collaboration into their consumers and non-consumers.

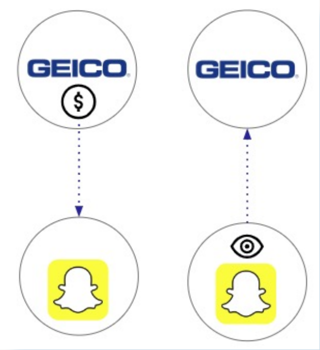


Image source: liveramp.com

External data collaboration

Marketers often don't own the entirety of their data points needed and may require collaboration across a number of partners and their owned data. This could entail one-to-one, one-to-many, or many-to-many collaboration.

One-to-one external data collaboration: Data collaboration between two partners.In the examples below, assume the car insurance brand is advertising with Snap.

1. **One example of one-to-one collaboration could be that the car insurance brands shares their first party insurance sales data with Snap for a complete look at Snap media effectiveness on insurance sales.**

Example business questions that could be addressed with this type of collaboration:

- How does Snap contribute to driving insurance sales?
- Are there portions of the Snap campaign presenting more or less opportunity?

2. **A second example of one-to-one collaboration could be that Snap shares their first party exposure data with the car insurance brand for a deeper look at media delivery metrics like reach and frequency related to their audiences.**

Example business questions that could be addressed with this type of collaboration:

- Are some consumer types more or less responsive to Snap media?
- How does Snap contribute to driving insurance sales?

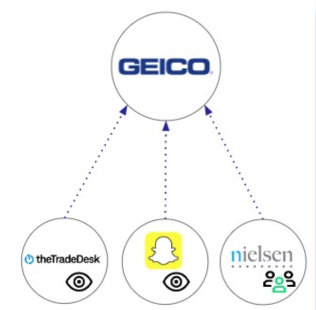


Image source: liveramp.com

One-to-many external data collaboration: Data collaboration between more than one partner. In the example below, a car insurance brand is collaborating with a number of their marketing partners, receiving exposure data from a DSP (such as The Trade Desk) and Snap, and audience data from Nielsen for a deeper look at cross-screen media effectiveness and audience insights.

Example business questions that could be addressed with this type of collaboration:

- What is the combined reach/frequency of my media buy - across partners?
- Which media channel or partner is driving the most incremental reach?
- What is our optimal cross-screen frequency?
- Are some consumer types more or less responsive to media?
- How do The Trade Desk and Snap contribute to driving insurance sales?
- Are there portions of my cross-screen campaign presenting more or less opportunity?

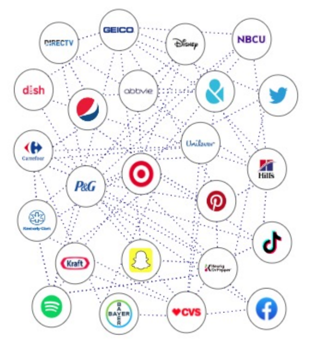


Image source: liveramp.com

Many-to-many external data collaboration: Options are endless but this refers to multi-party data collaboration for media, measurement, audience, and sales insights among other use cases across all the different consumer and marketing touch points.

Example business questions that could be addressed with this type of collaboration:

- What is the combined reach/frequency of my media buy - across partners?
- Which media channel or partner is driving the most incremental reach or conversions?
- What is our optimal cross-screen frequency?
- Are some consumer types more or less responsive to media?
- How do The Trade Desk and Snap contribute to driving insurance sales?
- Are there portions of my cross-screen campaign presenting more or less opportunity?

Recommendations, Considerations and Best Practices

Data collaboration technologies themselves cannot exist alone. They are part of a much broader business and technology landscape that must be considered. There are several considerations that should be factored in when assessing or implementing a new data collaboration technology.

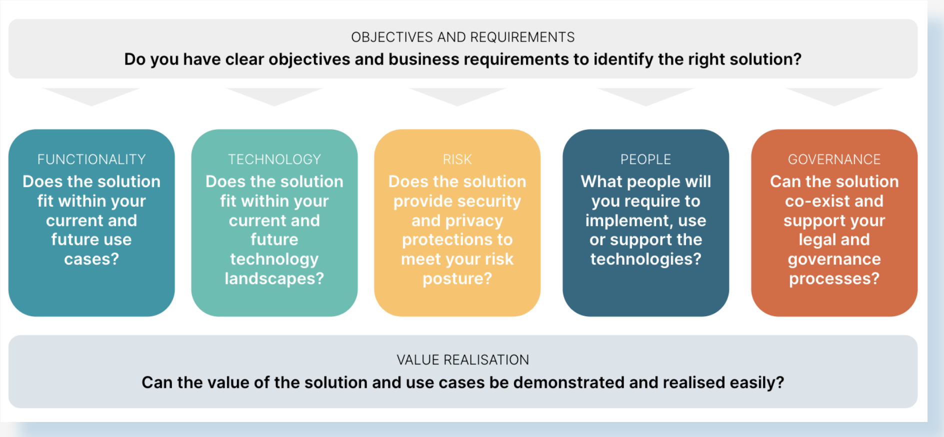

OBJECTIVES AND REQUIREMENTS
Do you have clear objectives and business requirements to identify the right solution?
FUNCTIONALITY
Does the solution fit within your current and future use cases?
TECHNOLOGY
Does the solution fit within your current and future technology landscapes?
RISK
Does the solution provide security and privacy protections to meet your risk posture?
PEOPLE
What people will you require to implement, use or support the technologies?
GOVERNANCE
Can the solution co-exist and support your legal and governance processes?
VALUE REALISATION
Can the value of the solution and use cases be demonstrated and realised easily?

Image source: dataco.ai

Do you have clear objectives and business requirements to identify the right solution?

Before beginning any data collaboration initiative, it is crucial to have clear objectives and requirements in mind. This involves identifying what data will be shared, who it will be shared with, and what type of data collaboration platform is best suited for your needs.

By having clear objectives and requirements, you can ensure that you are using the right technology and resources to achieve your goals. Consider the specific questions you want to answer with the data, the types of data that need to be connected, and who needs access to that data.

Does the solution provide the functionalities required to meet your use cases?

When assessing collaboration solutions, it is important to evaluate the various functionalities and capabilities offered by different platforms. Some platforms may offer tailored, pre-defined solutions for specific use cases such as reporting or audience creation, while others focus on a wider platform approach that gives users the capabilities to create solutions themselves to meet use cases. Understanding your use cases is important to understand what functionality is really required.

Regulation, consumer expectations, and technologies will continue to change. When selecting a solution for data collaboration, it is important to select a solution that is fit for purpose both now and in the future. Consider their ability to support new use cases, handle increasing amounts of data, and adapt to changing regulatory requirements.

Does the solution fit within your current and future technology landscapes?

Data collaboration technologies should also fit seamlessly into your wider technology landscape, as well as your partners' technology landscape. Interoperability is key to ensure that data can be easily shared and accessed across different systems.

Consider factors such as the platform's compatibility with other systems, APIs, and data formats, as well as the ability to integrate with emerging technologies. Additionally, it's important to select a solution that is supported by a vendor with a clear roadmap for future development and support.

Does the solution provide security and privacy protections to meet your risk posture?

Data privacy and security are essential considerations for any data collaboration initiative. Different data collaboration solutions offer varying levels of data protection and privacy-enhancing technologies are available as needed, while others apply privacy and security protections by default with granular controls to further restrict or lessen obfuscation. When selecting a solution, it is important to identify what level of data protection is required for your specific use case. It's important to consider factors such as what type of data is being shared, who it's being shared with, and how it will be used.

Some solutions may offer features such as encryption, data anonymization or pseudonymization techniques, access controls, privacy budgets, and audit trails, while others may require additional layers of security to be implemented. It is important to evaluate these features and determine whether they are sufficient for your specific needs.

What people do you require to implement, use or support the technologies?

Different data collaboration solutions may require varying levels of technical expertise. From low-code to pro-code, it is important to understand the skills that will be required to use different solutions effectively. Consider the skill levels of users who will be accessing the platform and the availability of resources to support training and development. It is important to select a platform that aligns with the skills and experience of your team both now as well as in the future usage as collaboration platforms scale.

Different data collaboration solutions require varying levels of implementation effort. Some solutions may require technology to be installed,maintained, and executed by technical analysts, engineers, or data scientists, whilst others may be low-code,easy to set up, and accessible to users with varying skill sets and backgrounds.

Understanding the implementation effort and time required is crucial in selecting the right solution. Consider factors such as the complexity of the platform, the availability of IT resources to support implementation and ongoing, the cost of certain technical resources, or and the time it will take to get started and prove value.

Can the solution co-exist and support your legal and governance processes?

Legal agreements for data sharing must be aligned with when implementing data collaboration. When selecting a solution, it is important to evaluate how well it can support legal obligations. This includes understanding what types of legal agreements and requirements are necessary, and whether the solution offers features such as comprehensive audit capabilities (reactive measures) or controls to ensure compliance (proactive measures).

Even with protections for data being processed, different regulatory or consent obligations may come into effect when information is revealed about an individual through collaboration platforms. When selecting a solution, it is important to evaluate how well it works with consents to ensure you remain compliant.

This includes understanding what types of consents are necessary, how they will be obtained and documented, and how the solution will support compliance with any applicable regulations or standards.

It is also important to evaluate how data collaboration solutions can help to align with or accelerate other governance processes within an organisation. This includes evaluating how well the solution can support data governance activities, such as data quality management, data lineage, data cataloguing, and data lineage management.

Can the value of the solution and use cases be demonstrated and realised easily?

Finally, for collaboration technologies to be successful for your organisation, they need to make sure they can support strong business cases for value creation. It is essential to consider both the initial and ongoing costs associated with solutions, as well as the time it takes to demonstrate and realise value.

Assessing costs should consider not only the cost of the platform itself but also any additional expenses such as training, support, and maintenance. You should also consider how you can prove the value of collaboration before fully investing in a particular solution.

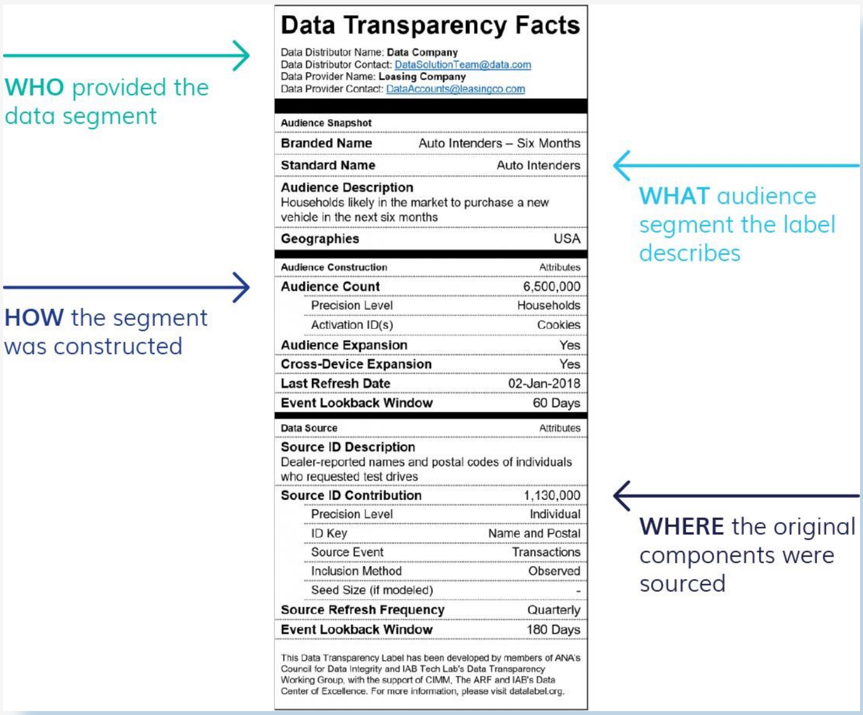
IAB Tech Lab’s Data Label Initiative

Alongside these key considerations for industry is the work undertaken by IAB Tech Lab in and around data transparency. A key project as a part of this is the Data Label.

Marketers are increasingly making media and marketing spend decisions based on a wide range of audience data, however traditionally there are few tools that enable data buyers to understand ‘what’s inside’ the various data segments they buy and few standards that can enable a consistent labelling of any details contained therein.

In 2019 we saw the launch of the Data Label project, the structure of which is based upon global Data Transparency Standards released by IAB Tech Lab. The initiative provides a simple, consistent and easily digested set of standards – allowing sellers to clearly specify where the data comes from, how it was collected and organised, its recency, if it was manipulated or modelled and what rules were used in establishing the data within any particular audience segments.

Just like a Nutrition Label, the industry’s Data Transparency Label is intended to give every marketer, agency, data provider and publisher a clear view of the syndicated audience segments they use:



WHO provided the data segment

HOW the segment was constructed

WHAT audience segment the label describes

WHERE the original components were sourced

Data Transparency Facts

Data Distributor Name: **Data Company**
Data Distributor Contact: DataSolutionTeam@data.com
Data Provider Name: **Leasing Company**
Data Provider Contact: DataAccounts@leasingco.com

Audience Snapshot	
Branded Name	Auto Intenders – Six Months
Standard Name	Auto Intenders
Audience Description Households likely in the market to purchase a new vehicle in the next six months	
Geographies	USA

Audience Construction	Attributes
Audience Count	6,500,000
Precision Level	Households
Activation ID(s)	Cookies
Audience Expansion	Yes
Cross-Device Expansion	Yes
Last Refresh Date	02-Jan-2018
Event Lookback Window	60 Days

Data Source	Attributes
Source ID Description Dealer-reported names and postal codes of individuals who requested test drives	
Source ID Contribution	1,130,000
Precision Level	Individual
ID Key	Name and Postal
Source Event	Transactions
Inclusion Method	Observed
Seed Size (if modeled)	-
Source Refresh Frequency	Quarterly
Event Lookback Window	180 Days

This Data Transparency Label has been developed by members of ANA's Council for Data Integrity and IAB Tech Lab's Data Transparency Working Group, with the support of CIMM, The ARF and IAB's Data Center of Excellence. For more information, please visit datalabel.org.

Image source: iabtechlab.com

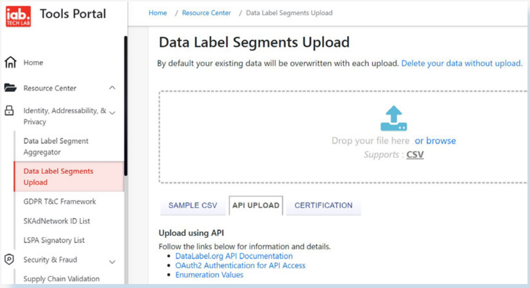
Overall, the benefits of the Data Label can be summarised as:

Educating industry around what is contained within different audience data segments, how they have been constructed and how often they are being populated and/or refreshed. This allows buyers to analyse and compare data sets and providers prior to purchase and activation.

Providing consistency and transparency in terms of the product constituents and naming conventions. This helps by removing barriers and confusing technical terms enabling everyone to better understand audience data in simplified easy to understand language.

Enabling a minimum level of quality assurance in terms of what is being bought and/or utilised. This allows buyers to activate with confidence by choosing the best audience data characteristics that match their preferred target audiences.

An online marketplace which provides access to the Audience Segment Metadata from Data Labels that have been uploaded is now available within the IAB Tech Lab Transparency Center. This provides a one-stop-shop for uploading, maintenance and reporting capabilities – as well as integrations across participating data marketplaces.



Data Label Segments Upload

By default your existing data will be overwritten with each upload. Delete your data without upload.

Upload using API

Follow the links below for information and details.

- DataLabel.org API Documentation
- OAuth2 Authentication for API Access
- Enumeration Values

Image source: iabtechlab.com

IAB Australia recommends that publishers and data suppliers start utilising the Data Label for their most commonly traded segments as a commitment to the highest standard of audience data transparency. In turn, buyers and agencies should start recognising those segments for their consistent and transparent approach and increasingly demanding their usage in any future ongoing transactions.

Related to these standards is the IAB Tech Lab's compliance program – which is available to any organisation that offers data, whether syndicated separately or bundled alongside with media, and is also open to adoption by data marketplaces where data is bought and sold. Those organisations that complete the compliance program affirm their full commitment to the highest standards of audience data transparency.

IAB Tech Lab's Data Clean Room Interoperability Standards

In February 2023, IAB Tech Lab released its 'Open Private Join and Activation' documentation, as a first in a series of Data Clean Room interoperability standards.

The intention being to better describe the specifications for implementing a matching operation between two parties and the supporting mechanisms to use the output of the operation to target matched users for advertising.

The key topics within the technical document relate to:

- Understanding the privacy and security goals in a Data Clean Room specific to any two-party matching process. This is achieved by providing a blueprint architecture to describe the participants, the input and output requirements of the matching system and the activation protocol components.

- Understand how to activate audiences in ways the privacy goals can remain preserved through to the end use of the outputs. This is managed via an 'Activation Protocol' that focuses on the design goals, encryption characteristics, format, and intended usage. The activation protocol also enables a matching system to pass confidential information in ad requests, encrypted for an intended ad system such as an SSP or DSP. Consequently, the participating SSP and DSP ad systems can perform private targeting of OPJA-matched user ad impressions

- How to structure and format the inputs and read the outputs for a matching operation. The Matching Systems provide high-level descriptions of open matching system component designs, and also presents some reference designs for both private set intersection (PSI) and trusted execution environment (TEE) based matching systems.

- Understand potential threat vectors and collusion scenarios from malicious actors that can result in failing to preserve privacy goals. These scenarios and threat vectors that must be considered by any component designs adhering to the OPJA specifications.

Below is a blueprint of the architecture, depicting principal data flows, with some further information by participant type further below.

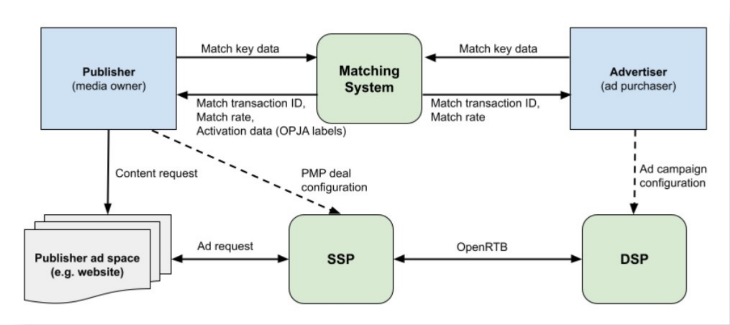


Image source: iabtechlab.com

Advertiser

The advertiser is the entity that wants to display an advertisement to a list of users identified by PII records (e.g. email addresses, phone numbers). The list of users may be, for example, the advertiser's existing customers or loyalty members, and the PII records may have been obtained through either online or offline means. The advertiser may be the advertiser organisation itself, or a delegated organisation acting on behalf of the advertiser, such as a technology vendor. Possible types of vendors here may include data collaboration systems, Data Management Platforms (DMPs), Customer Data Platforms (CDPs), etc.

For the purposes of this proposal, IAB Tech LAb did not distinguish between various types of delegated vendors, since they are trusted by and are at the discretion of the advertiser. Also, IAB Tech Lab considered the specific scenario where the advertiser wants to display an advertisement to a list of identified users when the users are interacting with digital media properties controlled by a publisher.

Publisher

The publisher is the entity that has an identified user audience. The publisher controls digital media properties (e.g. websites, applications) with advertisements and is able to associate identified users (e.g. email addresses, phone numbers) with users on their media properties. The publisher wants to enable an advertiser to display an advertisement to the list of identified users overlapping with the advertiser's list.

The publisher may be the publisher organisation itself, or a delegated organisation acting on behalf of the publisher, such as a technology vendor. Types of possible vendors are similar to those that an advertiser may use, though we assume that the advertiser and publisher, if they are delegating the process described here, could be using different vendors, such that no single organisation has access to PII records from both the advertiser and publisher involved in the OPJA. For the purposes of this proposal, we shall not distinguish between various types of delegated vendors, since they are trusted by and are at the discretion of the publisher.

Matching System Operator

Some architectures enabling the described process, such as the one depicted in the image above could require or benefit from the help of a third party matching system. In other cases it could be feasible to enable an OPJA match to happen between the advertiser and publisher using a peer-to-peer protocol. In the first case where a third-party matching system is involved, we must consider the third party entity operating the matching system and its relationship with the other participants involved in enabling OPJA. Solutions designers must also consider the privacy and security design goals as they relate to a third-party matching system operator.

Supply-Side Platforms (SSPs)

In order to enable ad delivery targeting identified users matched using OPJA, the publisher sends ad requests to an SSP.

When the SSP receives ad requests from the publisher's media properties, and depending on the proposed activation mechanism, the SSP may configure targeting of Private Marketplace (PMP) deals to matched users prior to forwarding OpenRTB requests to DSPs, or may simply forward ad requests to DSPs containing activation data that the advertiser's DSP can access.

Demand-Side Platforms (DSPs)

In order to target identified users matched using OPJA, the advertiser configures advertisement campaigns in a DSP. The advertiser may either target an advertising campaign to PMP deals pre-resolved by the publisher's SSP, or may target the matched users from the DSP based on information available in the OpenRTB requests processed by the DSP.

End User

While not pictured in the blueprint architecture shown in the image above, the end user is the entity that owns the PII record (e.g. email address) that it has voluntarily and separately shared directly, with both the advertiser and the publisher, and that accesses the publisher's controlled media properties where advertisements are displayed.

Future Evolution of Data Collaboration In an Era of Consumer Privacy

The future evolution of data collaboration will be heavily influenced by the advancement of technology and development of global regulations, particularly in the areas of privacy, control, and artificial intelligence. As privacy concerns continue to rise, there will be an increased demand for privacy-enhancing technologies that enable secure data sharing while preserving data privacy, allowing businesses to collaborate with greater confidence.

Furthermore, end-user control over their data has the potential to increase with the growing maturity and adoption of distributed ledger technologies and digital wallets, which can provide individuals with greater traceability and control over their data. This could lead to new business models, such shared data monetisation, where users can benefit more directly from sharing their data with businesses.

Another key area that may impact the evolution of data collaboration is the growth of artificial intelligence. Recent growth in popularity and use of different AI models, including large language models such as GPT-4, demonstrates how important they may become within our society. However, as their use and success is heavily dependent upon the data available, the drive for access to unique data to drive differentiated outcomes will become stronger as well. This will lead to increased collaboration among businesses to access and share unique datasets to feed these models.

Regulatory requirements will also play a significant role in shaping the future of data collaboration. Regulatory environments have already begun to shift, with the likes of EU's General Data Protection Regulation (GDPR) or the proposed Privacy Act Review changes in Australia putting businesses under increasing pressure to provide value propositions back to consumers for the use of their data. Shifts in regulation or expectations will lead to new forms of data collaboration, where businesses collaborate to provide value-added services to consumers in exchange for responsible use of their data.

Businesses that are able to adapt to these changes and collaborate effectively with others will be better positioned to thrive in a data-driven world.

Conclusion

In this explainer we've looked to provide some background and guidance on data collaboration for marketing purposes by looking at this product area as a category, rather than just considering the more popular or in vogue solutions.

Any business looking to invest in this space should make the effort to carefully evaluate various factors to ensure data privacy compliance, accurate analysis, and that core requirements are met. So in conclusion we wanted to provide some key considerations and questions for you when reviewing your various options.

Scalability & Features

Ensure that you choose a solution that can scale with your business needs, handle large data volumes and accommodate growth in sources, users, partners and analytical requirements.

- ***How many datasets, of what size and what level of complexity can the solution connect with and run computation against?***

- ***Does the solution allow for real-time activation via integrations with activation channels, and if so - which activation channels are supported?***

- ***Does the solution provide advanced analytics and machine learning capabilities, enabling you to extract valuable insights, build audience segments, and easily visualise the results?***

- ***Is there a limit on the number of Data Contributors or other participants that can interact in the same available solution?***

- ***Can the solution handle various data types, formats, and sources, including first-party, second-party, and third-party data?***

Compatibility & Speed

The solution should be compatible with your existing data platforms, marketing tools, and technologies.

- *Are there seamless integration capabilities and any potential limitations that could hinder requirements?*
- *How quickly can you configure and set up a solution?*
- *How long does it take to grant permissions and join datasets?*
- *How long do computations take to run, and can this be dialled up and down dynamically?*
- *How fast can insights be gleaned and executed?*
- *Can insights be gleaned and activated from connected datasets?*
- *Does the solution support datasets that are continuously changing?*
- *Does it support low-latency queries?*
- *Can data be joined and queries get executed over the joined dataset without the need to upload it into the platform?*
- *Will it be easy to onboard data collaboration with any other existing customers and can I expect to find new data collaborators who are already onboarded?*

Controls & Functionality

Ensure that you will have access to the tools you need to extract valuable insights, build audience segments, and meet your specific business requirements. The opportunity to create custom reports, different data models, and easy-to-use data visualisation options should be discussed. Also look for a provider that offers adequate support and training resources to help your team adopt and use the platform effectively. This should include documentation, webinars, and ongoing customer support services.

- *Does the solution provide functionality that maximises simplicity and ease of use for both data scientists and non-technical users?*
- *Does the solution facilitate an ad sales process?*
- *Does the solution incorporate data and/or analytics assets, such as pre-defined query templates, to aid in implementation and usage?*
- *Does the solution provide each data contributor the ability to define their own scoped access and permissions control and at what granularity?*
- *Which programming languages are supported (e.g. ANSI92 SQL, Python, C++, Java, Scala)?*
- *Can you import/use public and/or private libraries?*
- *Can you leverage machine learning within the operational workflows?*
- *What level and latency of logging and auditing is provided to the participants of other parties' usage of their data?*
- *Can the solution span across multiple regions, multiple clouds, and/or multiple data platforms?*

Data Privacy & Security

It's always critical to ensure that any potential provider complies with data protection regulations (e.g., GDPR, CCPA) and follows industry best practices for data security.

- *Does the provider have a proven track record in safeguarding sensitive customer information?*
- *Does the service comply with data protection regulations and follow industry best practices for data security?*
- *Does the provider have a good reputation in the industry?*
- *Can they provide customer reviews, case studies, and testimonials?*
- *Is there a proven track record of success and innovation in the data privacy and marketing analytics space?*
- *Are there publicly referenceable customers using the solution with similar use cases, who have already approved the solution from a privacy, legal, and scalability perspective?*
- *Can data remain in the region of origin without copying or transferring the data to perform cross-regional collaboration or computation?*
- *Is data required to be copied from one region to another in order to facilitate a collaboration or computation, if so would this require a transfer of data?*
- *How many regions does the service operate in and are there residing data/storage infrastructure?*
- *Where does computation take place? Does the data need to be moved in order to be used?*
- *Does the solution provide or support one or more PETs (Privacy Enhancing Technologies) to execute computations that preserve individual privacy with technical protections (e.g. Homomorphic encryption)?*
- *Does the solution provide or support technologies that mathematically guarantee a required level of privacy (e.g. Differential privacy)?*

Pricing & Licensing

Review all the various associated pricing models, including any up-front costs, ongoing fees, and any potential hidden costs. Make sure all the costs are considered for all your future use cases and fully align with your budget.

- *How much is a basic licence, what exactly does it cover and are there different pricing tiers?*
- *Do any participants or Data Contributors also need to pay for licences?*
- *How much does data storage cost?*
- *Are there additional fees for data preparation?*
- *Are there additional fees for connecting to specific other databases or other parties?*
- *Is there a per-query/operation cost?*
- *Are there costs for maintaining privacy technologies for encryption etc.?*
- *Is there a fee for computational usage, and if so is it separate from any fees for queries?*

We hope that you have found this explainer useful and if there is any feedback (both good and bad!) please do feel free to get in touch.

Further Reading

IAB Tech Lab's Data Clean Room Interoperability Standards

https://iabtechlab.com/wp-content/uploads/2023/02/FINAL-DRAFT-PUBLIC-COMMENT-Open-Private-Join-Activation-IAB-Tech-Lab.pdf

Data Clean Rooms: IAB Tech Lab Guidance and Recommended Practices

https://iabtechlab.com/wp-content/uploads/2023/02/FINAL-DRAFT-PUBLIC-COMMENT-Data-Clean-Room-Guidance-IAB-Tech-Lab.pdf

IAB Tech Lab's Data Transparency Standards

https://iabtechlab.com/standards/data-transparency/

IAB Tech Lab Transparency Center

https://tools.iabtechlab.com/transparencycenter/explorer/supplyChain/datalabel

Clean Room Primer (LiveRamp co-published)

https://liveramp.com/lp/eb/clean-room-primer/