Modern identity resolution

Time to see what you're missing

PUBLISHED BY:

CUSTOMER DATA PLATFORM INSTITUTE

celebrus

With changes to privacy regulations, advertising policies, and browser restrictions, now is the time to take a fresh look at your approach to identity.

Introduction

Identity resolution has never been easy, but the challenges have changed over time. Before the Internet, most marketers had very little data about individual customers. Digital technology has turned the data drought into a flood, but required new tools to stitch together cookies, device IDs, account numbers, and other identifiers into unified profiles. Today, the ground is shifting yet again as new privacy regulations, changes in advertising, and new policies from browser vendors restrict which data is available and how it can be used.

In the face of these changes, now is a good time for marketers to take a fresh look at their approach to customer identity.

Data comes first

Identity management is a complex process, but one simple truth is clear: identity results are only as good as the data they're built from. That's why privacy changes and browser restrictions that threaten data flow cause such great concern. Before we look at the specific issues, let's take quick look at the underlying processes.

With considerable simplification, an identity management system works like this:

- The system maintains a list of entities (typically individuals, but could be devices, companies, etc.). Each entity has a system-assigned master ID and a set of identifiers associated with the entity (name, address, email, account number, etc.). The master ID never changes, while any other identifier might be replaced over time.
- When the system is presented with a new identifier, it checks for a match against its current data. Then:
  - If the new identifier is already in the system, it makes a match.
  - If the new identifier isn't present but is associated with an existing identifier, it adds the new identifier. (For example: an email address and device ID are presented together, and the email is present but the device ID is new. The system adds the device ID with a link to same master ID as the email).
  - If the new identifier isn't present and can't be linked to a known identifier, it creates a new master ID with a link to the new identifier. This might later be merged with an existing master ID if we are able to make a connection.

There are many nuances to this process, especially in how a "match" is found. But even from this very brief description, it's clear that the more identifiers the system ingests, and the more accurate the collection is, the more matches it can find. These identifiers fall into two main groups:

Direct identifiers, such as email address or account ID. These generally apply to a single individual. They may be addressable (i.e., could be used to find the individual in the real world, such as an address or phone number) or anonymous (cannot be used by itself to find an individual, such as an account number without associated contact information). Direct identifiers often require an exact match, although some, such as postal address, facial recognition, or device "fingerprints", are matched based on similarity.

Derived identifiers, such as behavior patterns and combinations of non-specific data such as zip code, birthday, and gender. These require statistical analysis to estimate the probability that two sets of data relate to the same entity. This sort of "probabilistic" matching is primarily used to determine whether two devices belong to the same person, based on the devices being frequently used in the same places at the same time. Derived identities are inherently anonymous, although they may be connected to a known identity through a shared master ID.

Threats to data access

Access to new types of identifiers or to more samples of existing identifiers both increase the number of potential matches available to an identity management system. Privacy-related changes that threaten to reduce access include:

Third-party cookie blocking

Cookies are small files placed on a device by tags or pixels embedded in a web site, mobile site, or apps. These tags can send the cookie ID and other data to external systems.

This enables the tag owner to track behavior and interactions over time. First party cookies are cookies that are set within the domain that an individual is browsing; third party cookies are set by a different domain that is different from the domain being browsed. Major browsers including Apple Safari, Google Chrome, and Mozilla Firefox have begun to implement policies (which Apple terms Intelligent Tracking Prevention) which restrict third party cookies and, in some cases, block them completely.

Although some firms have built workarounds that disguise third party cookies as first party cookies, the browser makers have tightened their rules to detect and block these.

The primary use of third-party cookies is to track individuals across Web sites so ads can be targeted at them and subsequent purchases can be linked to ad impressions. This data is generally not shared outside of the ad networks. However, many cloud-based Customer Data Platforms (CDPs), Web analytics and data collection tools that work on a company's own website also rely on third party cookies or workarounds to send data to the vendor. These vendors need to develop alternatives to ensuring continued access to client information.

Mobile device ID blocking

Every mobile device has its own ID, used to identify the device to the mobile network. But advertisers and app developers are generally not allowed to use these. Instead, the operating systems maintain advertising IDs, including IDFA for Apple and AAID (sometimes called GAID) for Google Android. Apple recently started to require that applications get explicit permission ("opt-in") from users to access IDFA and Google will likely follow suit. Because opt-in rates are expected to be low, the flow of data connected with a usable mobile device identifier is expected to fall to a trickle.

GDPR, CPRA, and other privacy regulations

Details vary, but many new privacy regulations prevent sharing of some information or impose

new requirements for consent to different uses. GDPR in particular treats cookies and IP addresses as personal identifiers. These changes are expected to most heavily impact access to "third party" data, which is collected by one company and sold through another. It also means that organizations need to shift more towards true, first party data capture in a fully compliant manner to ensure they are being a good steward of customer information.

Growing the data supply

Marketers are responding to these changes in two main ways. The first is attempting to preserve as much as possible of the old data sources:

Push back

Companies and trade groups are lobbying regulators to loosen existing rules, avoid adopting more stringent new rules, and interpret rules more favorably. Enforcement of new privacy regulations has been lax in many regions, although it's likely to stiffen as compliance grows and regulators feel that remaining violations are more willful. In the U.S., many data collectors are pressing for a federal privacy rule to override state rules, with the assumption that federal rules will be weaker than strong regulations such as California's. Some firms are also pressing Apple, Google, and other private companies to reduce or stop expanding their own privacy-based restrictions.

Refined compliance methods

The advent of GDPR made acquiring consent from customers a requirement. Today, leading enterprises employ refined consent mechanisms that can increase the amount of data that's collected with permission for broader use. This will continue as companies learn how to present consent requests in the most appealing ways. Companies are also improving internal systems that govern data use, enabling them to implement sophisticated rules that allow some uses while blocking those that aren't compliant.

Technical workarounds

Some companies that used third-party cookies for data collection have developed techniques that make them appear to be first-party cookies, such as CNAME techniques in which data being sent to third party domains are 'cloaked' to give them the appearance of the company's own domain. Apple, Firefox, and Google browsers have been swift to restrict these methods, while inventive vendors continue to introduce new work arounds. And so, the game of cat-and-mouse continues, much to the frustration of enterprises who simply wish to gain visibility of their customers' interactions to optimize their digital experiences. Others have adopted methods such as "fingerprinting" that constructs a derived identifier using publicly shared device data. This raises compliance issues and is also increasingly obstructed by the browser makers.

The second way is to prioritize the collection of true, first-party data, which is data the company has gathered directly from an individual. This can be done by embedding digital capture within the company's own infrastructure. Not only does this ensure that the company can set and persist first party cookies in the most acceptable way, but it also gives companies full control of their data instead of sharing potential sensitive personal information with third parties.

On-premise data collection

It's entirely possible to capture, build, and maintain an identity system in a first-party way. Systems that do this capture the data directly into the company systems and set a true first -party cookie. The data can then be processed and shared without constraints imposed by browser makers, although it's still necessary to comply with privacy regulations and align with a customer's preferences of opt-in or out.

First-party identifiers

Ad industry proposals to enable cross-site tracking of individuals rely largely on collecting their email address. These are "hashed" with an algorithm that

yields a unique value which cannot be decrypted to reveal the actual email address. This lets companies that have independently captured the same email identify individuals who have visited both their Web sites. This method may still be rejected by privacy regulators and it is limited to the minority of site visitors who provide their email address. But to the extent that it does work, companies with more complete first-party data will benefit.

Derived identifiers

Rethinking data capture on your owned channels can also ensure your ability to capture cross-device and channel as well. Although device fingerprinting is controversial, other methods that create a derived identifier by combining multiple elements of customer data cause less concern. Examples include matches that compare several elements of a postal address and matches that compare typing styles. New derived identifiers will become possible as systems capture new types of data about their customers' actions.

Making it happen

Few people would question whether gathering more data can improve identity resolution. But they might ask what they can do to get it. There's plenty of basic advice that we need not repeat here: build data capture programs, put in place governance rules, monitor quality, and measure the value created. Here are three less obvious technical features to look for when assessing your existing data management capabilities and evaluating new systems.

On-premises operation

As we've seen, restrictions on sending data to third parties have made it increasingly difficult for these solutions to collect customer data directly from Web sites, mobile sites, and mobile apps. Workarounds that make a third-party cookie appear to be first-party are not likely to survive determined opposition from the browser makers. Look for a solution that captures and sends all data from your websites, mobile sites, and apps directly to your company's owned and controlled environment, whether it's on-premises, in a datacenter, in a private cloud, or a hybrid. If the system does the rest of its processing on those servers, you will have the simplest possible configuration and avoid concerns about sending sensitive company data outside your organization. Otherwise, be careful to examine the security and privacy practices of your external processors to ensure they – and you – comply with all requirements.

Flexible data capture

One of the biggest challenges in data collection is capturing unexpected items. Most solutions use JavaScript tags, which require users to specify in advance what will be collected from a website, mobile site, or apps. This means a new piece of data won't become available until after someone requests it, tags for it, and rolls it out to production. There's often a further lag waiting to collect enough of the new item to be useful.

These issues impact identity resolution in particular when the new elements could be useful in building derived identifiers. Look for a system that can capture and store all interactions, behaviors and experiences, regardless of the channel your customers chose to use, and can contextualize that fully for your business.

Flexible systems which let users leverage new aspects of customer interaction data without setting up new collection parameters provide considerable advantages in these times of disruption and transition. Why only capture some of the data if you can capture all of it and be ready for your future requirements and use cases?

Multiple identity definitions

Identity resolution systems are often limited to a single definition per individual. In our experience, your company may need multiple identity levels, so it can resolve queries about households, companies, and other entities in addition to individuals. In addition, there is often value in having different matching

rules within the same identity type. Many companies apply strict matching rules for reports to customers – ensuring that data from different people is not mistakenly combined – but use much looser rules for marketing analysis, where failing to link related data may obscure the most important customers.

Clear match rules

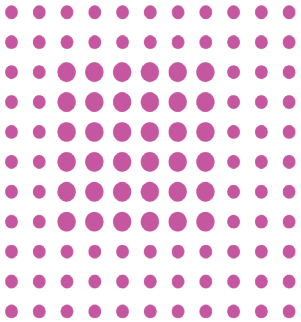
Most identity resolution systems apply a complex combination of rules, templates, and statistical methods to decide matches. While this complexity is often necessary for good results, it can prevent users from understanding what the system is actually doing. Look for a system that makes it easy to review and change rules and lets you see the net impact of any change before it goes live. This is even more important when there are multiple identity definitions, each with a rule-set of its own. Also, you should be able to change these rules without upsetting the underlying data you've captured.

Real time execution

Identity systems vary widely in the degree of real-time processing they support. Most can find an exact match between a presented identifier and what they already have on file, such as looking up an account ID. But it's harder to find a system that can add new identifiers to an existing record in real time, and quite rare to find one that will rebuild all identity relationships to incorporate new information in milliseconds. Which real-time features you need will be based on your business, so be sure to identify your own requirements and then ensure you find a system that meets them.

The road ahead

Identity resolution has never been more challenging. But hidden inside the challenge is an opportunity: to build a system that lets your company react quickly to changes in customer data and to deliver superior customer experiences as a result. Now is the time to compare your current identity management and data capture capabilities with industry-leading alternatives, to find out what you're missing, and calculate what you'll gain when you close the gaps.



About Celebrus CDP

Celebrus CDP captures the market's most complete picture of customer behavior and experience, creating events and profiles in real-time for one-to-one personalization and streaming analytics.

Featuring the industry's only first-party Identity Graph, Celebrus CDP delivers unrivalled identity solutions to a range of enterprise clients. The Celebrus CDP Identity Graph provides an out-of-the-box, yet fully customizable solution, resolving the identity of visitors to digital channels via a flexible range of identifiers.

As a true first-party data capture solution, Celebrus CDP is completely unaffected by the browser restrictions (such as Intelligent Tracking Prevention) described in this paper, meaning it's the only solution of its kind capable of genuine omnichannel data capture and the creation and persistence of detailed yet compliant customer profiles.

Celebrus CDP is quick and easy to deploy and easily connects to industry-standard data applications for customer insight and engagement. Celebrus gives clients complete control by enabling best-in-class privacy compliance and flexible options for hosting data securely in the cloud, using a hybrid solution, or on-premises.

Global businesses in banking, insurance, retail, travel, automotive, and communications industries rely on Celebrus CDP in collaboration with leading industry partners to drive rapid and ambitions transformations of customer experience and engagement.

About CDP Institute

The Customer Data Platform Institute educates marketers and marketing technologists about customer data management. The mission of the Institute is to provide vendor-neutral information about issues, methods, and technologies for creating unified, persistent customer databases. Activities include publishing of educational materials, news about industry developments, creation of best practice guides and benchmarks, a directory of industry vendors, and consulting on related issues.

The Institute is focused on Customer Data Platforms, defined as "a marketer-controlled system that maintains a unified, persistent customer database which is accessible to external systems."

The Institute is managed by Raab Associates Inc., a consultancy specializing in marketing technology and analysis. Raab Associates defined Customer Data Platforms as a category by Raab Associates in 2013. Funding is provided by a consortium of CDP vendors.

CDP Institute

231 2nd Avenue
Milford, CT 06460

www.cdpinstitute.org

info@cdpinstitute.org