// ARTICLE · PUBLISHED 2 NOVEMBER 2026

Identity Data in Marketing and Fraud: The Web of Identity

The traces people leave online become one picture of who they are, which marketers use to decide what to send and fraud teams use to decide whether to believe them

Jason Newell · MAX Research Collective · Article ART-DAT-007 · Published 2 November 2026
Licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/
Original: https://www.maxresearchcollective.com/articles/identity-data-in-marketing-and-fraud.html
Cite as: Newell, J. (2026). Identity Data in Marketing and Fraud: The Web of Identity. MAX Research Collective, ART-DAT-007. https://www.maxresearchcollective.com/articles/identity-data-in-marketing-and-fraud.html

One picture, two uses

Every click, open, form and login leaves a trace, and businesses stitch those traces into a picture of who you are. Marketers use that picture to decide what to send you. Fraud teams use the same picture to decide whether you are who you claim to be. Both depend on one answer: whether two records belong to the same person. AI is about to make that answer more valuable and harder to get.

I've worked on several sides of that picture. At WebMD I ran the portfolio behind consumer and provider portals and the reporting the sales team sold against. At iovation I ran the data warehouse behind device-based fraud and reputation analytics. At Infogroup's Yesmail I ran the data and business intelligence (BI) organization for a national email marketing provider. At Cambia Health Solutions I delivered the omnichannel marketing platform for its Blue Cross Blue Shield brands.

Cambia: one platform, two jobs

At Cambia, a multi-brand health plan, claims, membership, care management and marketing each ran on their own stack. I led the work to deliver a cloud-based omnichannel platform, meaning one system that runs outreach across every channel a member uses, for the plan's Blue Cross Blue Shield brands. It unified campaign execution, customer relationship management (CRM) and member outreach across channels on a shared data and identity foundation.

The platform served two jobs at once:

- Marketing. It reached prospects and members with the right plan message, on the right channel, at enrollment and renewal.
- Business operations. It ran member outreach for eligibility, care reminders and service, tied back to claims and membership data so a message reflected the member's real status.

The key design choice was a single member identity across brands and channels. Without it, one person showed up as four records and got four conflicting messages. With it, marketing and operations worked from the same view of the member. Unifying duplicated platforms and their integrations improved operational efficiency about 50 percent across the affected workflows.

Because this was health data, three things sat underneath everything. Identity federation let one verified identity work across separate systems. Controls for HIPAA, the federal law that governs health information privacy, covered every record. Integration ran over HL7 and FHIR, the standards health systems use to exchange records.

Infogroup: campaigns as a data product

At Yesmail, campaign, customer and response data was the product clients bought. Every send produced a stream of delivery, bounce, open, click, unsubscribe and conversion events. My organization loaded those events, matched them to the right customer and campaign, and turned them into client reporting.

Three things made it work. The first was enrichment: appending demographic and modeled financial attributes to a client's list turned a mailing list into market segments, so offers could fit the household. The second was operational discipline. We automated schema creation for each client's feeds, made loads safe to rerun, and sped biweekly delivery 15 percent through agile practice and test and deployment automation.

The third was A/B market testing. Before a full send, we split a sample of the audience across competing versions of a campaign, with different subject lines, offers, creative or segments. Each version carried its own tracked links, so every click-through was credited to the exact version that earned it. Click-through rate told us early which campaign was generating real interest, well before sales data came in. The winner went to the rest of the list, and conversion to sales then confirmed whether that interest turned into revenue. Opens showed who noticed; click-throughs showed which message worked.

The lesson carried into everything after it. A campaign is only as good as the identity behind it. If you can't tell that two records are one person, you can't target them, measure them or protect them.

WebMD: reading the path people take

At WebMD, the web portals carried an advertising business, and traffic was what the sales team sold. That made usage data a revenue asset. The questions were simple to ask and hard to answer well:

- Where do people come from? The referrer, the page or source that sent the visitor, tells you: a search engine and the query behind it, a partner link, an email campaign tagged with tracking parameters, or a bookmark.
- What do they do? Page sequences, time on page, searches on the site, and the content that keeps them reading all answer it.
- Where do they leave? The exit page is the last page before someone closes the browser or goes elsewhere. A single-page visit is a bounce.

Exit pages are the most honest signal on a site. A high exit rate on an article that answers the question is success. A high exit rate halfway through a registration form is a broken step. Telling those apart meant joining traffic data with page purpose, which is why we built it into a data warehouse and kept it out of a standalone analytics tool. That warehouse also produced the reports advertisers and the sales team relied on.

How the data gets collected

Customer data arrives through four doors. A customer opens some of them knowingly and never sees the others.

Table 1. The four doors customer data comes through, and whether the person notices.

SOURCE	HOW IT WORKS	EXAMPLES	DOES THE PERSON NOTICE?
Declared (first-party)	The person gives it directly	Account sign-up, checkout, loyalty programs, surveys, preference centers	Yes
Behavioral	Code records what the person does	Cookies, tag managers, tracking pixels, mobile app SDKs, server logs, tracked email links and opens	Rarely
Enrichment (third-party)	Records are matched to outside data	Data brokers, demographic and household append, modeled credit attributes, address and phone validation	No
Device signals	The device describes itself	Browser and OS version, screen size, fonts, time zone, IP address, network, hardware traits	No

The tools have shifted more than the categories. A third-party cookie is a small file that a site other than the one you're visiting places in your browser, so it can recognize you across sites. Safari, Firefox and Brave already block them by default. Google spent years planning to remove them from Chrome, then reversed course in April 2025 and announced in October 2025 that it would retire most of its Privacy Sandbox replacements, so Chrome still supports them today.

Even so, the industry has moved toward first-party data, which a business collects directly from its own customers, and toward server-side collection, which records events on the company's own servers rather than through scripts in the browser. Cookies are fragile, and consent laws in the EU, the UK and several US states keep tightening. The data that holds up best is the data a customer chose to give.

The identity web: how fraud teams recognize you

The same signals marketers use to target you are what fraud teams use to verify you. The difference is the question. Marketing asks what this person will want. Fraud asks whether this is the same person, device and history it has seen before, and whether any of it links to known fraud.

Device fingerprinting is the core technique. A script reads dozens to thousands of traits that a browser or phone reveals: operating system, browser version, screen, fonts, time zone, language, IP address and network details, and hardware characteristics. Combined, they identify a device without storing anything on it. A cookie disappears when someone clears their site data, and a fingerprint survives it, which is why the UK Information Commissioner's Office says fingerprinting reduces user choice and control. Google barred advertisers from fingerprinting for years — then began allowing it on 16 February 2025.

At iovation, fingerprinting was the foundation of a device reputation network. Businesses shared reports of confirmed fraud tied to devices, so a device caught abusing one retailer carried that history to the next bank it touched. When TransUnion agreed to buy iovation in May 2018, the network covered nearly 5 billion

devices across more than 35,000 brands. TransUnion now says the product draws on over 185 million confirmed fraud reports.

Modern fraud platforms layer four kinds of signal into one graph:

- Device. Device signals cover the fingerprint, the device's reputation, and signs of emulators, proxies or a hidden location.
- Account. Account signals cover email and phone age, login history, and how many accounts share a device.
- Behavior. Behavior signals cover typing rhythm, mouse and touch patterns, speed through a form, and bot-like timing.
- Identity and demographic data. These checks ask whether name, address, date of birth and Social Security number agree with credit bureau, utility and public records.

The power is in the links. One device tied to 40 new accounts, or one address shared by a dozen unrelated applicants, is worth investigating as a possible fraud ring even when every single application looks clean, though a shared family computer or a fingerprint error can produce the same match.

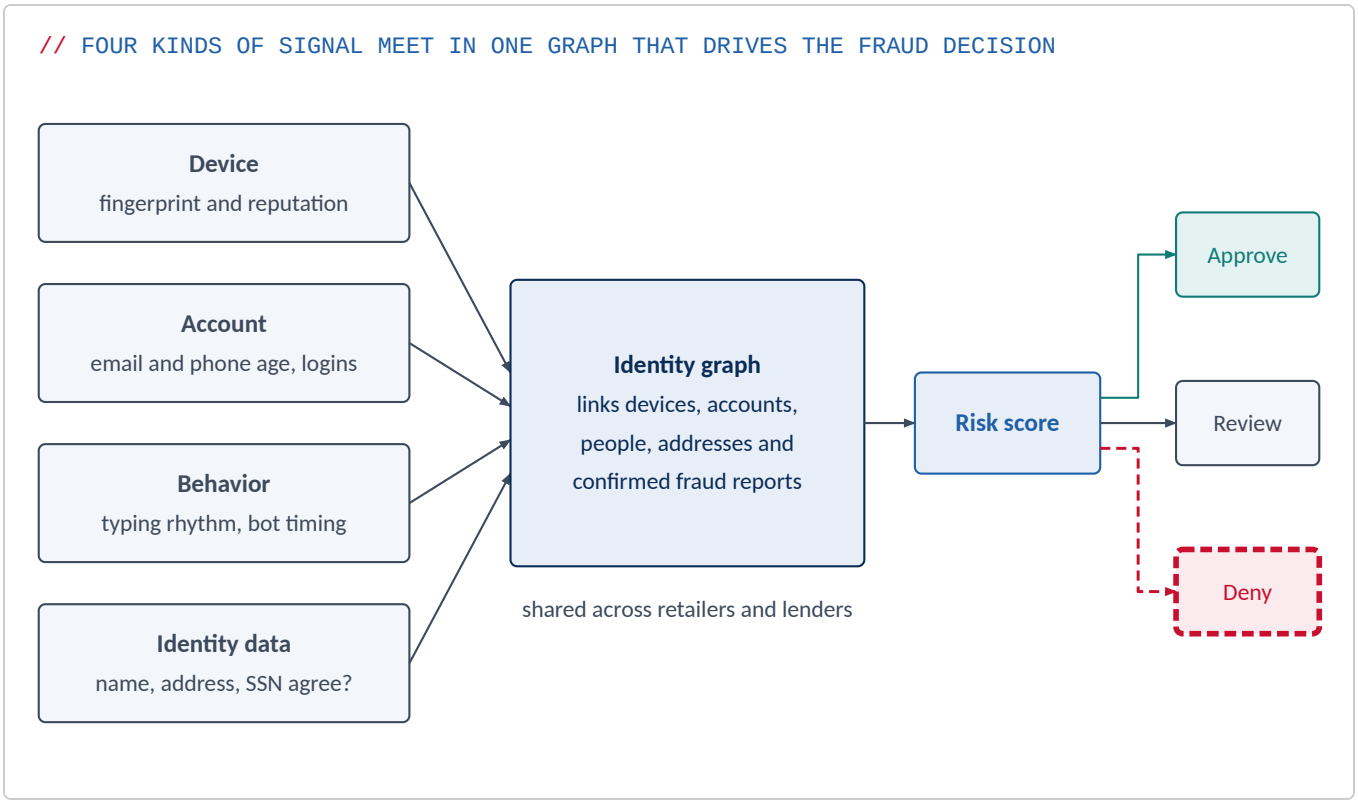


Figure 1. Four kinds of signal feed one identity graph, and the score it produces decides approve, review or deny.

Each signal alone is weak. Linked across businesses, the signals let a device or address with a fraud history raise the score before any money moves.

Retailers use the graph to stop account takeover, card testing, promotion and return abuse, and bot purchasing, while letting trusted returning devices skip the extra checks. Lenders use it at application time against stolen and synthetic identities, and against loan stacking, where one person applies to many lenders at once. A synthetic identity mixes real and invented data into a person who doesn't exist. TransUnion put US lenders' exposure to synthetic identities at $3.3 billion in 2024.

Identity in the hardware

The idea of identifying a machine from the chip itself is older than modern fingerprinting, and it ran straight into public resistance.

Intel's processor serial number came first. In early 1999, Intel announced that every Pentium III would carry a unique processor serial number, and pitched it for e-

commerce and for tracking stolen computers. Privacy groups called for a boycott and filed a Federal Trade Commission complaint arguing it could track users across the web. Intel switched it off by default. Researchers then showed software could read it even when users believed it was disabled. Intel phased the feature out of later chips. A critic at the time made the point that still matters: a chip ID identifies a processor, and people share computers and switch between them.

The Intel Management Engine came next. Modern Intel chipsets include it as a separate small computer with its own firmware and deep access to the system, built for remote IT administration. In 2017, researchers at Positive Technologies found an undocumented switch that disables it, labeled for the High Assurance Platform program of the National Security Agency (NSA). Intel said the setting was added at the request of equipment makers serving US government customers, and it denies building backdoors. The episode shows government interest in switching hidden hardware features off for its own security. It's no evidence that governments store people's identity on processors. I've found no credible evidence that they do.

Hardware identity today is cryptographic rather than a readable serial number. Some Intel processors do keep a unique inventory identifier, but only privileged software can read it, and only where access is enabled; a website can't. Security chips such as Trusted Platform Modules (TPMs) and phone secure enclaves hold keys that prove a device is genuine and unmodified. Passkeys, which replace a password with a key held on the device, rely on them, and so does app attestation, where a device proves to an app that it hasn't been tampered with. Government ID is arriving on devices too, by the holder's choice: state mobile driver's licenses sit in phone wallets that the holder controls and presents. The trend runs toward devices that prove a fact when the owner chooses to, and away from chips that announce themselves to every website.

How AI changes the picture

AI changes every layer described above, and it helps both sides of the fraud fight.

Marketing becomes one-to-one. The segments I built at Infogroup were groups of households. AI can now write and time a message for each person, picking channel,

offer and moment from their history. The constraint shifts from production capacity to data quality and consent. Personalization built on a messy identity graph will be confidently wrong at scale.

The funnel metrics weaken. Opens were already unreliable. As people ask AI assistants for answers instead of browsing, fewer visits reach the site, and the referrer and exit-page analysis I relied on at WebMD sees less of the journey. As AI agents start shopping and booking for people, some of the visitors on a site will be software acting for a customer. Analytics will need to tell a person, that person's agent and a hostile bot apart.

Fraudsters industrialize. Generative AI produces convincing documents, faces and voices cheaply. On 10 June 2026, PYMNTS reported a new category of fraud that combines deepfake video, cloned voices, invented work histories and computer-generated money habits into fake borrowers built to pass automated underwriting. Research from Mitek and Datos Insights, published the same day, puts synthetic identity fraud growth at about 16 percent a year, with US losses on unsecured credit projected to exceed $3.1 billion in 2026.

Defenders lean harder on the graph. A synthetic person can look perfect on paper, so the evidence that still works is what's hard to fake at scale: device history, behavior over time, and links to other accounts. Fraud platforms now apply machine learning across those links to spot rings and bots. Lenders are also separating identity checks from credit scoring. Equifax launched a synthetic identity model and a separate credit abuse model built that way in January 2026.

Privacy pressure rises with capability. The more an AI model can infer from scattered signals, the more a profile can reveal that the person never shared. Regulators already treat fingerprinting as something that needs the person's consent.

Why it matters

The durable path is the one that worked at Cambia: one clean identity per person, built on data they chose to give, used for things they'd expect. A business that gets there can personalize without guessing and verify without watching. A business

that skips it will send the wrong message to the right person, approve the synthetic borrower and turn away the real one, sometimes in the same week.

The tools will keep changing. The core skill stays the same: knowing which records are the same person, which signals can be trusted, and where helpful recognition turns into surveillance.

Sources

- Stape, "Google Chrome: Cookie deprecation explained."
- Google, "Next steps for Privacy Sandbox and tracking protections in Chrome," Privacy Sandbox blog, 2025.
- Information Commissioner's Office, "Our response to Google's policy change on fingerprinting."
- Malwarebytes, "Google now allows digital fingerprinting of its users."
- TransUnion, "TransUnion Announces Agreement to Acquire iovation to Strengthen Fraud and Identity Solutions," 18 May 2018.
- TransUnion, "TruValidate Device Risk," product page.
- CyberInsider, "Synthetic identity fraud exposes US lenders to $3.3B in losses, TransUnion warns," 17 September 2025.
- PYMNTS, "AI Fraudsters Are Building the Perfect Fake Borrower," 10 June 2026.
- CFO Dive, "Synthetic identity fraud surges as criminals weaponize AI: study," 10 June 2026, reporting Mitek and Datos Insights.
- Equifax, "Equifax Introduces Enhanced Synthetic Identity Fraud Detection" and "Equifax Launches Credit Abuse Risk Model," press releases, January 2026.
- Center for Democracy and Technology, FTC complaint on the Pentium III processor serial number, 1999.
- CNN and IDG, "Pentium III squabble continues," 1999.
- Computerworld, "Intel to phase out processor serial numbers."
- IT World Canada, "Identify the user, not the processor."
- The Register, "Intel ME controller chip has secret kill switch," 2017.

Glossary

TERM	DEFINITION
A/B test	an experiment that sends competing versions of a message to random parts of an audience and compares the response. It picks the winning version before the full send.
Account takeover	a criminal gaining control of a real customer's account, usually with stolen credentials. Device history is one of the strongest defenses against it.
App attestation	a device proving to an app that its hardware and software haven't been tampered with. It lets an app trust the device rather than a password alone.
Business intelligence (BI)	the reporting and analysis built on an organization's data warehouse. At Yesmail it was the product clients paid for.
Card testing	running many small charges on stolen card numbers to find the ones that still work.
Click-through rate	the share of recipients who click a tracked link in a message. It measures which message produced real interest.
Customer relationship management (CRM)	the system that holds a business's record of each customer and every contact with them.
Device fingerprinting	identifying a device from the traits its browser or operating system reveals, without storing anything on it. It survives clearing cookies.
Enrichment	appending outside data, such as demographic or modeled financial attributes, to a customer list. It turns a list into segments.
Exit page	the last page a visitor views before leaving a site. Read against the page's purpose, it shows success or a broken step.

TERM	DEFINITION
First-party data	data a business collects directly from its own customers. It holds up best as cookies and consent rules tighten.
HIPAA	the Health Insurance Portability and Accountability Act, the federal law that governs the privacy of health information.
HL7 and FHIR	Health Level Seven and Fast Healthcare Interoperability Resources, the standards health systems use to exchange records.
Identity federation	letting one verified identity work across separate systems. It's what made one member one record at Cambia.
Identity graph	a linked map of devices, accounts, people, addresses and fraud reports. The links reveal fraud rings that single records hide.
Loan stacking	one person applying to many lenders at once, before any one credit report shows the others.
Omnichannel platform	one system that runs outreach across every channel a customer uses, so the messages agree.
Passkey	a sign-in credential that replaces a password with a cryptographic key held on the device.
Privacy Sandbox	Google's proposed replacements for third-party cookies in Chrome, most of which it announced in October 2025 that it would retire.
Referrer	the page or source that sent a visitor to a site, such as a search engine, a partner link or a tagged email.
Server-side collection	recording events on a company's own servers rather than through scripts in the visitor's browser.
Synthetic identity	a person assembled from real and invented data who doesn't exist. Lenders face billions of dollars of exposure to them.
Third-party cookie	a small file placed in the browser by a site other than the one being visited, so it can recognize the person across sites.

Newell, J. (2026). MAX Research Collective, ART-DAT-007. CC BY 4.0

TERM	DEFINITION
Trusted Platform Module (TPM)	a security chip that holds keys proving a device is genuine and unmodified.

DATA ART-DAT-007

// ABOUT THE AUTHOR

Jason Newell leads organizations and people in technology, and is a researcher, deep technologist and builder — from a CRAY to agentic AI. He founded MAX Research Collective to build AI for regulated industries that runs where its owner has control, with healthcare as its strongest focus. About Jason → Connect on LinkedIn →