
How to Close the Data Fragmentation Gap on the Open Web

Inside Realize ID: the identity and signal infrastructure built to recognize users across a fragmented ecosystem

Executive Summary

Advertising has always come down to the same three moves: you recognize a user, you work out what they are about to buy, and you reach them while the decision is still open. Everything else a marketer does is in service of those three, which makes them a fair way to judge any platform you put money behind. Judged that way, the open web has been failing at the first move for years, and the price of guessing at the other two now has a number: \$26.8 billion in wasted spend in 2025 alone.¹

Social and search have solved this in a vacuum. When someone logs in, their identity is settled from the first moment, and the path from first impression to purchase within that platform runs on one infrastructure. If recognition, intent, and conversion all happen there, nothing about the user is lost between the steps.

The catch is that your customers do not live inside one platform. They read the news, check their mail, search, compare prices, move through apps, and increasingly ask an LLM before they buy, and their decision takes shape across all of it. Every environment they pass through knows them only as a new arrival, so what is really a single journey reaches you in disconnected pieces.

The industry has spent years building around those disconnected pieces with third-party cookies, mobile advertising IDs, probabilistic matching, universal IDs built on hashed email, clean rooms, contextual targeting, and advertiser first-party data. What's been missing is the technical ability to map those identities across domains and devices.. That waste figure is up 34% in just two years, over the exact period these tools matured. The tools multiplied and the waste grew anyway, because until now, no amount of spend could buy a continuous view of a user from a company that sees only a portion of their journey.

\$26.8B

WASTED SPEND IN 2025

+34%

GROWTH IN TWO YEARS

Seeing the whole journey takes three things at once:

- The technical ability to map a user across devices and domains
- Presence from where interest first shows to where the purchase finally happens
- An identity layer persistent enough to recognize that the user at both ends is the same user.

The reason so few platforms can hold all three at once is the same reason the journey came apart to begin with. Realize ID is built on all three.

¹ ANA, Q2 2025, [Programmatic Transparency Benchmark \(2025\)](#).

The Open Web's Data Fragmentation Problem

The open web can feel intimidating because until now, no company that touches the consumer journey was built to see the whole of it. Each sees only the stretch that runs through its systems.

Capturing the Whole Purchase Journey is a Challenge

A single purchase decision now runs across mail, news, search, price aggregators, and increasingly LLMs. US adults spend nearly 13 hours a day with media, spread across dozens of devices and platforms,² and a buyer passes through several of them before deciding.

Most Players Only See A Portion of a Purchase Journey

Each environment issues its own identifier, so the same user arrives unrecognized at every boundary, while the record of it is scattered across all of them.

Each sees most of the territory and only a sliver of the user, and the scale puts a whole view out of reach from any one seat. In 2025, a single campaign ran across a median of 26,000 domains bought through 20 different SSPs, and nearly half of marketers could not access their own impression-level data to piece together where it went.³

The advertiser has been left choosing between a complete view of a fraction of the journey and a partial view of most of it.

Every player runs into the same limit from a different seat. The walled gardens see a user completely, but only while the journey stays on ground they own. Everyone working outside these walls has the opposite problem. A DSP holds the bid and the click, an SSP the inventory and the yield, and both lose the user the moment they leave the page. Identity and measurement providers resolve users or rebuild journeys from behaviour someone else observed.

Introducing Realize ID as the Solution

Realize ID closes the gap as a unified identifier for a single user outside the walled gardens, built by clustering the identifiers someone leaves across channels into one profile, and enriching that profile with what Taboola sees hard coded on both the demand and supply sides.

The Realize Identity Graph is the map of every one of those clusters, at network scale, refreshed daily. For an advertiser, that means one user mapped once across devices and domains, their intent visible while it is still forming, and a conversion connected back to the impression that started it.

² eMarketer, [US Time Spent With Media\(2026\)](#).

³ ANA, [Programmatic Transparency Benchmark \(2025\)](#).

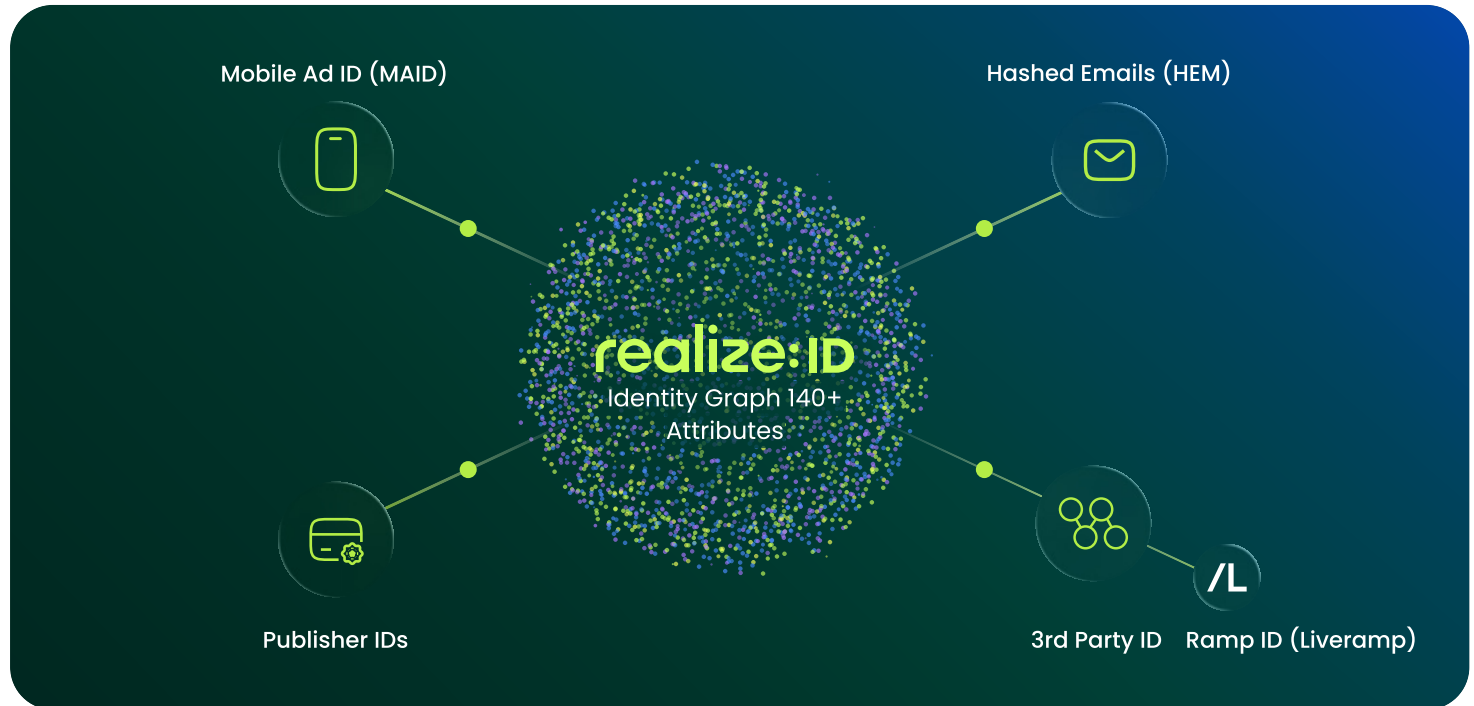
Realize ID's Data Advantage

The industry is converging on identity graphs. Nearly everyone in the market now has one or is building one, but an identity graph, on its own, proves only that the same user came back. When every player can recognize a returning user, recognition is just an entry fee.

The value of an identity graph is what a company can attach to the ID once it exists, which is a matter of position, because a company can only act on signals it can see. Realize ID allows advertisers to see from their first interest to the final purchase.

1 Realize ID Connects a User's Identifiers Into One Profile

Most of the open web isn't logged in. In 2025, 84% of publishers were using or planning identity graphs, and 91% were testing alternative IDs to overcome this challenge. But publishers whose audiences are even 60% authenticated remain uncommon,⁴ so no single identifier can recognize a user across a whole journey.



Realize ID overcomes that constraint. As a user moves across the network, they leave different identifiers behind, including cookie-based IDs, mobile advertising IDs, publisher partner IDs, third-party IDs, and RampID.

Realize ID clusters them into one profile, so one user is recognized as one user, whether they are in mail on a phone or reading news on a desktop. The Realize Identity Graph holds every cluster on the network, processing 240 to 390 million distinct clusters daily to keep each profile current.⁵

While other competitors like UID2 and ConnectID are trying to solve this problem as well, what separates Realize ID is what gets attached to the profile next.

⁴ ID5, [State of Digital Identity \(2025\)](#).

⁵ Taboola internal data (2026).

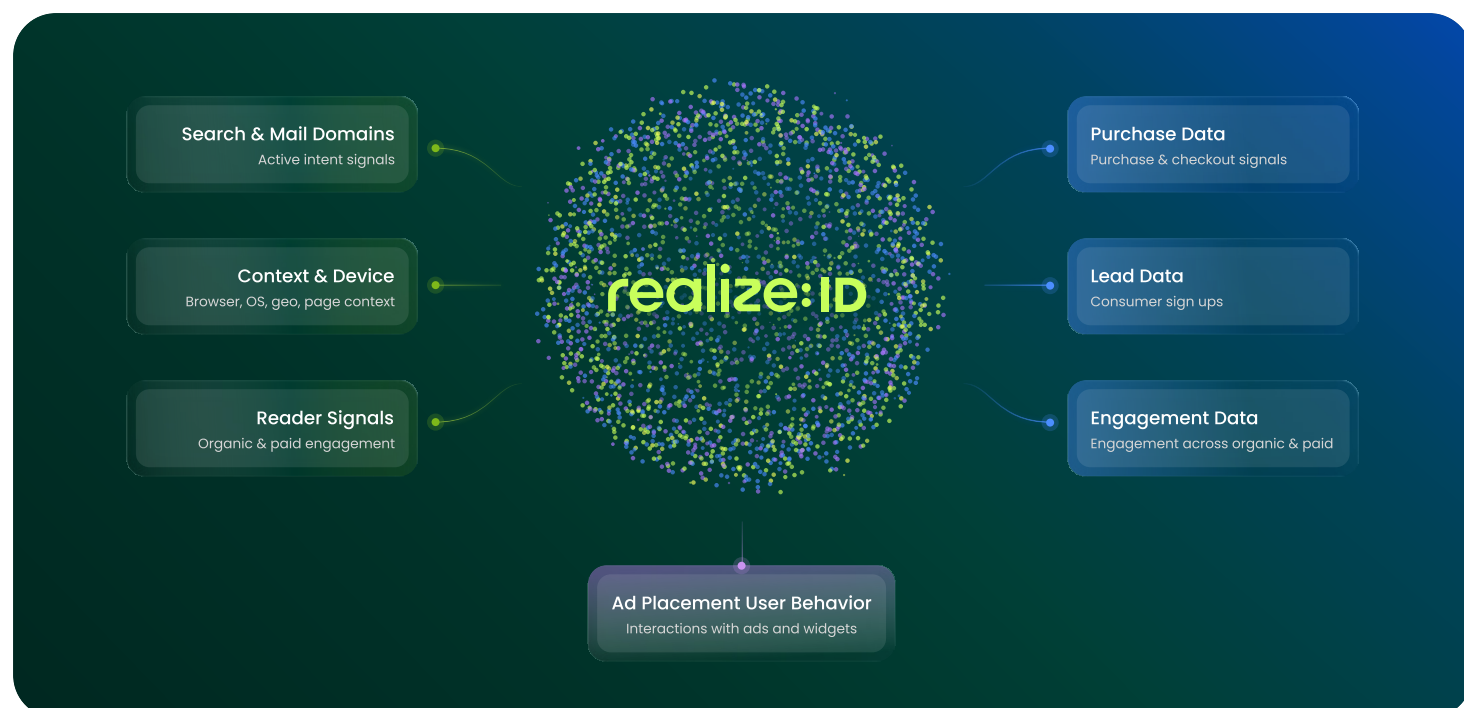
2 Signals Are What Make the ID Effective

Signals are the record of what a user does: what they search for, what they read, how long they stay with it, and what they buy. An identifier (ID) can connect all of that behaviour to one user. Only the behaviour itself shows what the user wants and how close they are to acting on it.

Most identity graph owners buy some signals from data providers, which aren't as effective. In a 2025 study of six major providers, purchased identity linkages were accurate 16% of the time or less, the providers agreed with one another on fewer than 7% of them, and an estimated three cents of every dollar spent on those segments reached its intended audience.⁶

The signals attached to a Realize ID are observed directly, and together they run from first interest to final purchase. Search keywords capture a user actively looking, and mail domains capture the brands they already hear from. Campaigns built on them have driven CPAs up to 40% and 28% lower.⁷

Code on the publisher page then reads how interest deepens, through scroll depth, time on page, and engagement with organic and paid content, framed by the device, browser, and content on screen. The pixel on the advertiser site records how the journey ends, in carts, leads, and purchases.



Every identity graph can say a user came back. Whether it can say what that user wants depends entirely on this layer, and it is the layer where the graphs differ most.

⁶ CIMM, Go Addressable, and Truthset, [Challenges of Using IP Address for Audience Identification and Measurement \(2025\)](#).

⁷ Taboola internal data (2026).

3 Realize Collects Signals From Across the Purchase Journey

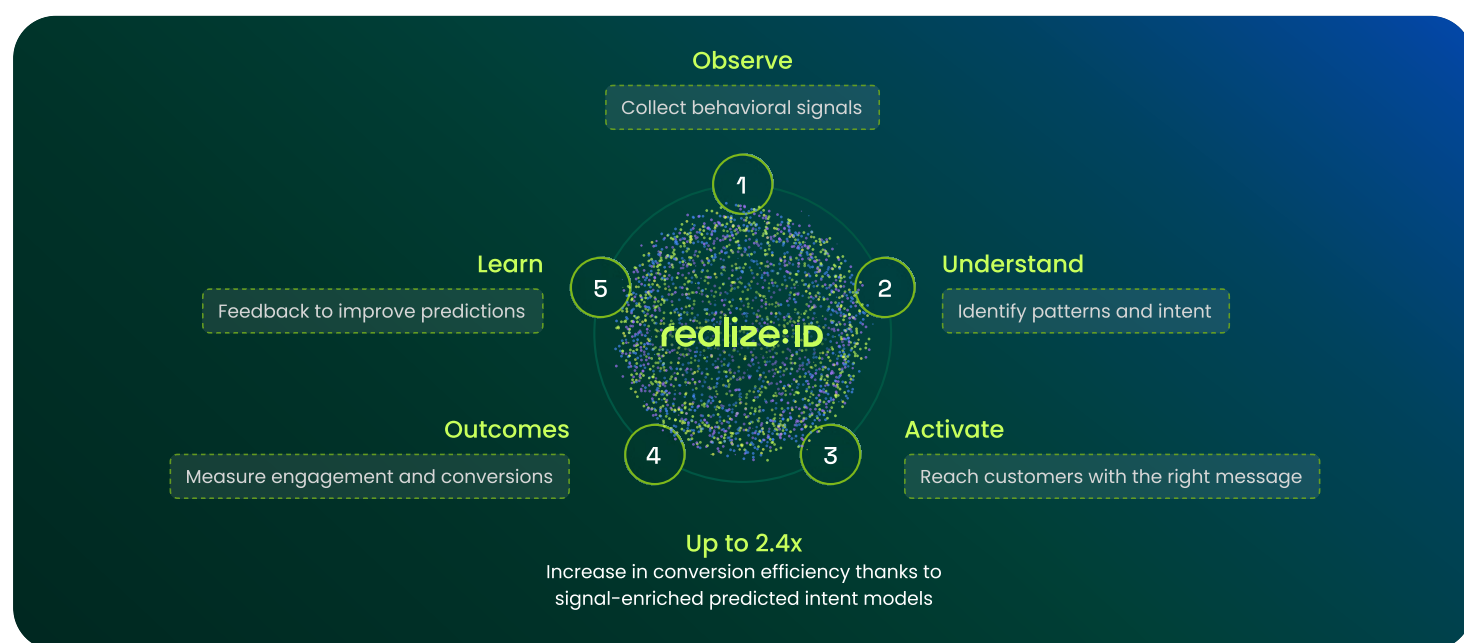
The reason Realize ID signals span the whole journey comes down to where Taboola's code sits. Code on the publisher page observes where interest first shows. The pixel on the advertiser's site observes where the purchase finally happens. Between them, one company watches both ends of the same journey firsthand.

No other seat on the open web holds both.

A DSP, for example, sees the bidstream and the click, and everything after the click reaches it secondhand, reported through someone else's system.

Realize ID allows us to follow the user through their entire journey. They may read a finance article, and our code on the page would know how far they scroll and how long they stay.

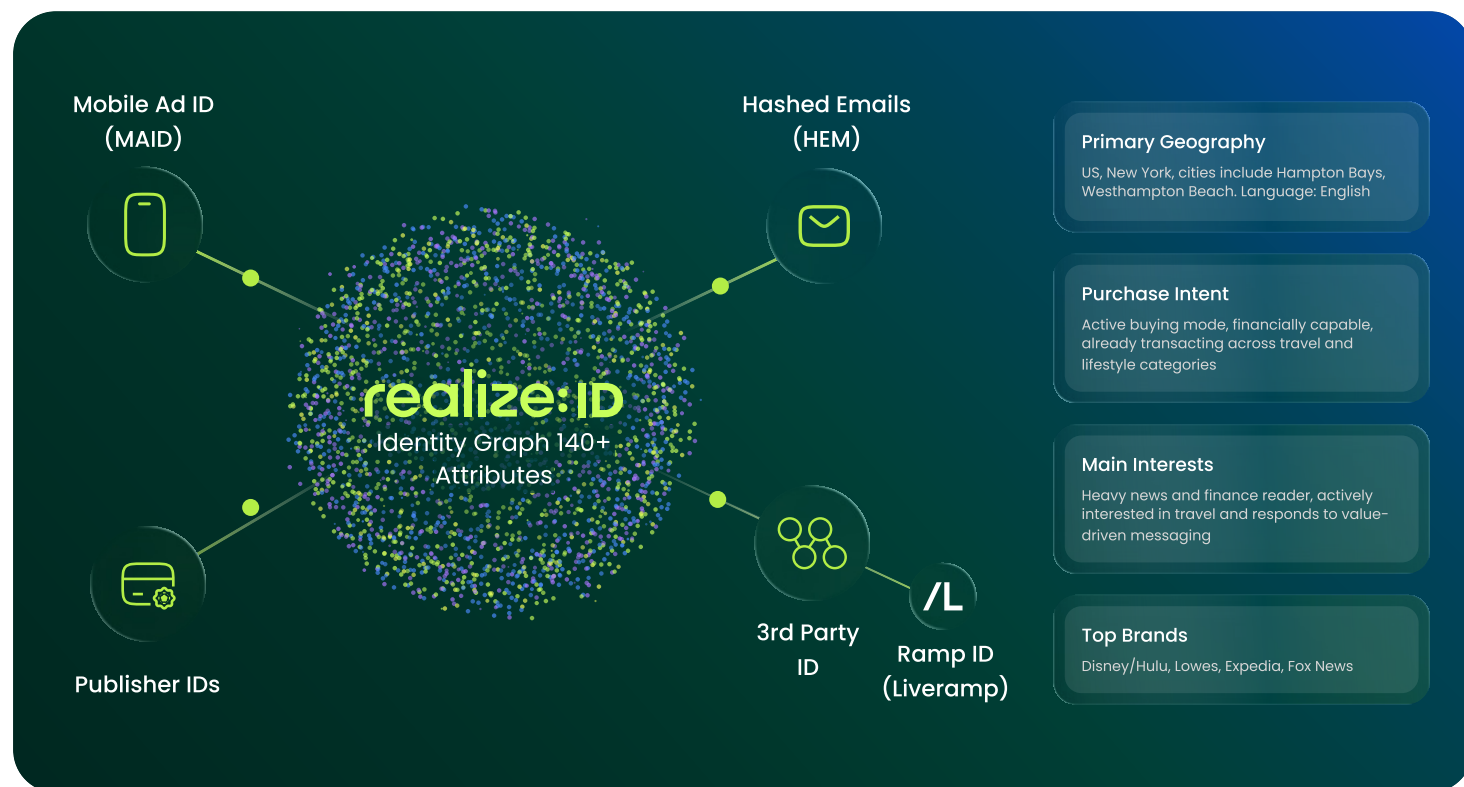
They may then search for life insurance, and the keyword would attach to their ID. They may engage a life insurance ad, start a quote on the advertiser's site, and abandon it, which the pixel would record. They may be retargeted, return, and convert, and the pixel would tie that conversion back to the impression that started it.



Every one of those steps writes back to the same Realize ID, so the profile that informed yesterday's bid is sharper for today's.

4 The Result Is a **Profile** of a Buyer

Each cluster in the Realize Identity Graph resolves into a profile an advertiser can act on. It shows intent to purchase based on things like where a user is, what they read, which brands they engage with, and what products they are in the market for.



That is a different kind of knowledge than a demographic segment carries. A demographic describes who a user resembles. The profile describes what they are about to buy.

Advertisers put it to work in two ways.

- 1 The first is automatic.** The profile feeds the Realize algorithm on every campaign, so bidding leads toward users with intent. On average, we see up to a 144% uplift in CVR, signaling a greater ability to predict intent, which will improve as we continue to add signals to the model.⁸
- 2 The second is deliberate.** The same signals are packaged as a targeting toolkit, with more than 3,000 In Market For segments live across verticals including finance, home and garden, auto, and healthy living, alongside search keyword and mail domain targeting and bundles, Taboola Audiences, and contextual and topics targeting. Users in In Market For segments convert at rates 20% higher.⁸

The profile is built entirely from what Realize observes on its own network. It gets stronger still when advertisers add what they know.

⁸ Taboola internal data (2026).

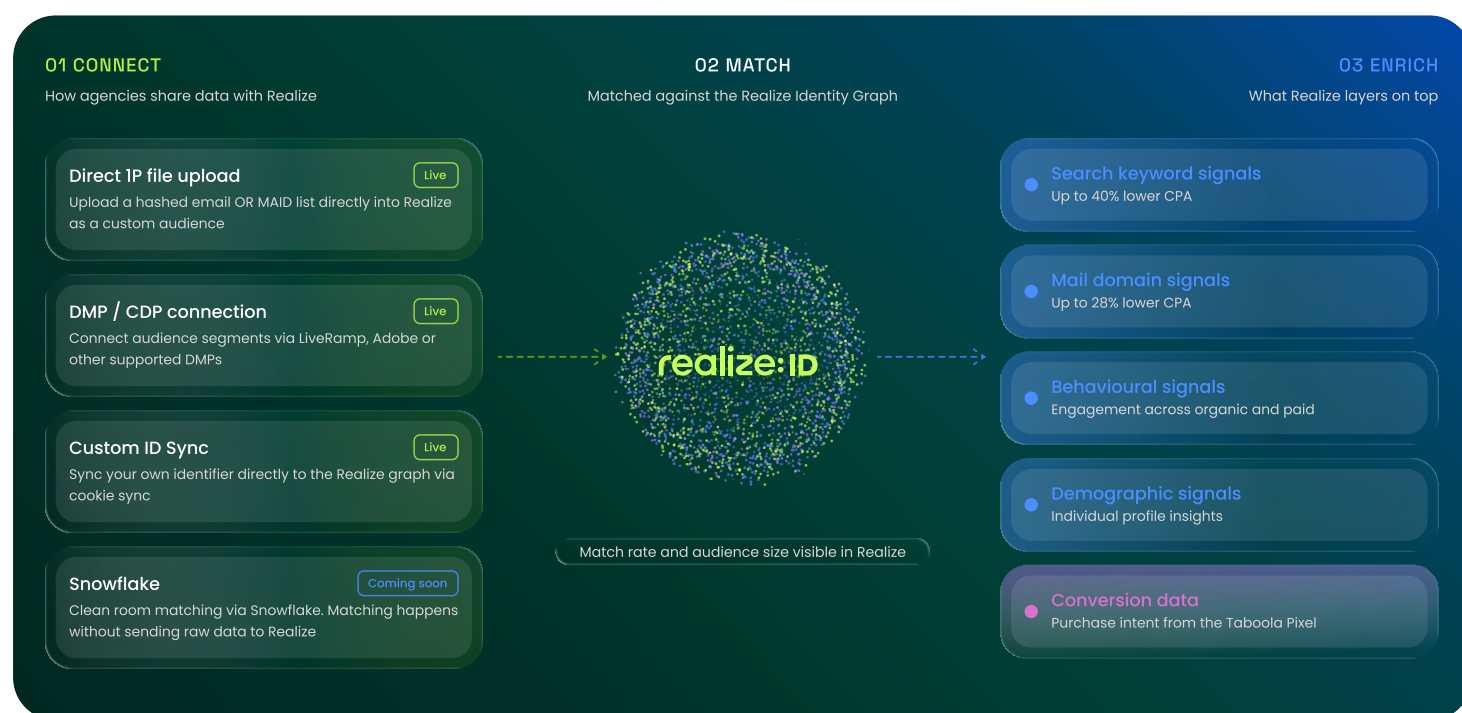
5 Advertisers and Agencies Can Connect Their Own Data

A pixel can record a purchase, but it cannot see the context around it. Only the advertiser knows a customer's lifetime value, their offline purchases, their loyalty status, or that a segment lapsed two years ago and is due to come back.

That knowledge lives in CRM files and CDP segments, and for agencies, often in an identity graph of their own. The industry has spent years trying to put it to work through data collaboration, and adoption is no longer the obstacle. In a 2025 industry study, 66% of organizations were using data clean rooms in some capacity, yet 39% said they struggle to turn what's inside them into anything actionable.⁹

Using Realize ID, an advertiser can make this connection by uploading a hashed CRM file, connecting a DMP or CDP, syncing a custom ID, or matching in a clean room. The match rate and audience size are visible in the platform rather than reported back as a black box.

Once matched, the enrichment starts. The advertiser's and agency's data bring the context. The network brings the live behaviour, what those users are reading, searching, and abandoning. Neither side can see the other's half without the match, and together they turn an audience the advertiser already owned into an audience with live intent attached.



For agencies running their own identity graphs, the custom ID sync makes that graph a first-class input, matched and enriched the same way.

⁹ Skai, [State of Retail Media \(2025\)](#).

How Mitsubishi Is Using Our Data Advantage.

When Mitsubishi launched the Eclipse Cross in Brazil, its performance agency Cadastra needed qualified test drive leads, and the channels that usually deliver them, search and social, were showing diminishing returns. The campaign they built on Realize never targeted an age range. Using Data Marketplace, Cadastra layered third-party segments over Realize's own behavioural data to reach users actively researching SUVs and comparing vehicle features, an audience defined by what it was about to do rather than who it resembled.

The same logic ran through the creative. Testing showed that "Learn More" ads won on click-through while "Test Drive" ads produced better qualified leads, so budget moved to the version with the lower CTR, because the response itself was a signal of intent and clicks were never the outcome being bought.

When those users reached Mitsubishi's site, the pixel picked the journey up where the publisher page left off. Sessions ran 103% longer than social's and bounce rates came in 41% lower, and Maximize Conversions fed that on-site behaviour straight back into bidding, so each round of buying started better informed than the last. Costs finished 9% more efficiently than social across both CPC and cost per visit.

103%

LONGER SESSIONS

41%

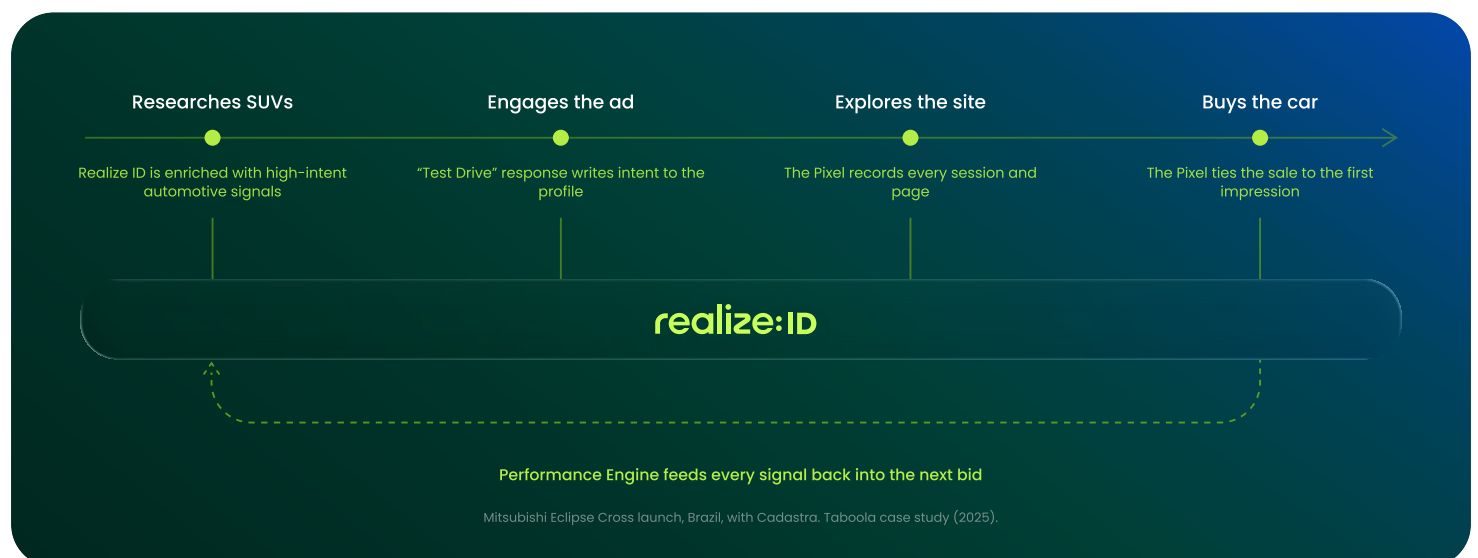
LOWER BOUNCE RATE

9%

MORE COST EFFICIENT

4

CONFIRMED CAR SALES



Cadastra tracked four confirmed car sales to a campaign designed only to drive qualified traffic. The data saw further down the funnel than expected, and Mitsubishi has since brought three more campaigns to Realize.

AI is Only as Good as the Loop that Feeds it.

Search and social have always seen interest, action, and purchase as one record when the entire journey happens within their walls. Everything else that makes them easy to buy, the precision, the measurement, the efficiency, is downstream of that one fact. Outside of those walls, you can see a richer set of data: more of the customer's day, more environments, better context, but could never close the loop.

For an advertiser, a closed loop outside of walled gardens changes the daily mechanics of buying ads:

- A frequency cap set against a Realize ID caps the user rather than each identifier they carry, so a budget stops paying six different prices to reach one buyer six times.
- Prospecting and retargeting collapse into a single program because the new prospect and the returning visitor turn out to be the same user at two moments in the same decision.
- And when that user converts, the conversion is recorded against the impression that started the journey instead of being modelled after the fact.

That kind of targeting is what buyers already expect from search and social, and its absence is the reason its competitions have been budgeted as a diversification line rather than an acquisition program.

Outside of those walls, new environments keep appearing, and LLM surfaces are only the newest boundary a user crosses. To a stack of partial views, each one is another place the journey breaks. In a closed loop, each is one more place the record continues.

**Discover What Realize ID
Can Do for Your Next Campaign.**

Talk to our sales team today

