Worldwide Web of Human Surveillance: Identity Resolution and Customer Data Platforms

July 24, 2024

Author

Lisa LeVasseur

Contributors

Researchers: Lisa LeVasseur, Saish Mandavkar, Bryce Simpson

Acknowledgements

This research is funded by Wrethinking, The Foundation.

Title image generated by Microsoft Designer with prompt "personal information sharing by hundreds of digital platforms worldwide".

The Worldwide Web of Human Surveillance

Table of Contents

1 Executive Summary

In February 2024, Cracked Labs published "Pervasive identity surveillance for marketing purposes", an in-depth analysis of LiveRamp's RampID identity graph. One of the most superficial yet most powerful functions of this excellent report was to guide attention towards industries responsible for pervasive consumer surveillance. The timing was excellent as I'd already committed to present "The Hidden Identity Infrastructure" at Identiverse (May 2024) and prompted by the report, I dug in to better understand the two industries underpinning hidden identity infrastructure, namely, Identity Resolution (ID Res) and Customer Data Platforms (CDPs).

There are nearly $9T worth of industries worldwide that rely on persistent, *hidden* identification of people. Naturally, demand of this magnitude fueled the now mature industries that perform pervasive, universal identification of people and their personal information. ISL identified over 350 companies providing either identity resolution platforms, customer data platforms, or both.

This paper explores the magnitude and reach of these two industries, how they came to be, and most importantly, why, from a human well-being perspective, it's crucial that these kinds of platforms be held to higher regulatory standards of scrutiny, transparency, and accountability. One identity resolution company alone out of 93 such companies (worldwide) boasts the collection of 5,000 data elements for [each of] 700 million consumers in 2021. To put this in perspective, the number of user accounts breached worldwide in 2023 was about 300 million[1]. Is there an appreciable difference between stolen user data and undisclosed "legitimate" personally identifiable information sharing?

Indeed, after reviewing the research, we must ask ourselves, is this the kind of world we want to live in: a world where everything about us is always known by industry; a world where the ongoing surveillance of people is deemed necessary in the name of capitalism. Is this the kind of world in which humans and societies will flourish or self-destruct? Are humans more than capitalistic consumers? Are we more than our purchasing potential?

We firmly believe so, and we hope that this paper illuminates the entities hiding in plain sight for years so that they may be held accountable for their troves of data on all internet users.

[1] https://www.pcmag.com/articles/2023-was-the-worst-year-yet-for-data-breaches

NOTE: Two companion databases are published with this report: (1) list of identity resolution and customer data platform companies, and (2) list of apps found in the ISL 2022 safety benchmark to have network traffic to identity resolution or customer data platforms.

2 Online Identity

Like so many digital behaviors, identity or identification, has been the subject of much technical advancement, scrutiny, and debate for the past few decades. The "on the internet no one knows you're dog" meme from New Yorker artist Peter Steiner in 1993 is one of the earliest artifacts from the online identification quagmire. Despite the levity of the observation, industry has a strong need to control access to users of their digital tools (e.g. employees) and services (e.g. end users), and thus, to uniquely and accurately identify people and their roles relative to the company's digital resources.

The very term "identity" has been fraught, and the history won't be rehashed here, but Internet Safety Labs (ISL) relies on Joe Andrieu's functional definition of identity:

> *" Identity is how we recognize, remember, and ultimately respond to specific people and things."*[2]

For practical purposes, this paper prefers the term *identification* to *identity.* When technologists speak of *identity* they are referring to the bundle of technologies that enable identification, authorization, and authentication of users of technology. However, in a colloquial sense, identity refers to something much more personal and relational. When I consider my identity, I'm not including notions of the technologies that enable online identification; thus the preference for the term identification. Online *identity* is usually in the context of the singular relationship between the technology platform and the user of the platform[3].

However, what this paper will show is an unstoppable global push for the *universal* identification of people through massively networked identification systems—across the digital and the physical worlds, whether people are logged in or not—largely for

---

[2] "*A Primer on Functional Identity*", Joe Andrieu, Rebooting the Web of Trust #7. We would add in the word: *personally* before "respond" in the definition, to read: "ultimately personally respond to specific people and things."

[3] We refer to this as the digital Me2B relationship. It's technically the relationship between the user of technology and the maker of technology as facilitated through interactive software. The software behaves as a *programmed* or *programmatic* agent of the company who made the technology. See also section 3.2

marketing purposes. The primary industries developing these capabilities are marketing and advertising, though many industries benefit from the fruits of universal identification as will be explored later in this paper.

2.1 Visible and Hidden Identification

Most of the online identity discussion has happened amongst the so-called "Identerati", identity technology experts, people like Kim Cameron and others whose work lies mainly in the Identity and Access Management (IAM) industry and standards arenas. IAM is concerned with the technologies that allow an organization to control user access to digital services through the familiar tools of login credentials. IAM is the collection of technologies that enable *visible identification*, as it is performed in a ceremony in which the user participates, namely online account creation and account log-in. Thus, in this paper, visible and hidden is from the user's perspective. Visible identification is visible from the user's perspective because they are actively involved in it.

In contrast, there is an even larger domain of hidden identification, which is done without involvement—or even awareness—of the individual or data subject (Figure 1).

Figure 1: *The Iceberg of Hidden Identification*

Whereas identity and access management is the enabling industry behind visible identification*, identity resolution* and *customer data platf*orms are the core enabling industries behind hidden identification. This paper focuses on exposing the hidden identification technologies and industries.

> *Identity resolution and customer data platforms facilitate the worldwide web of surveillance of all internet users.*

2.2 Digital Me2B Relationships

Internet Safety Labs started out as the Me2B Alliance, a standards developing organization (SDO). The intention was to alter the traditional industry-heavy composition of SDOs, and to provide everyday people ("Me-s") a seat at the table of software safety standards[4] along with industry or "B-s" (businesses). The core ethical foundation for what ultimately became "safety" was the quality of the relationship between the user of technology ("Me") and the company responsible for the technology ("B"), as mediated by the technology itself. This is called the digital Me2B relationship. We emulated a model of human relationships[5] to characterize states of a relationship (and thus stages of identification) throughout the lifecycle of the digital Me2B relationship[6] (Figure 2).

The pinnacle of the Me2B relationship is the creation and use of a personal online account. Ideally, account creation is voluntarily initiated by the user to signal a desire to be remembered, recognized, and personally responded to. We haven't conducted research, but it would be interesting to poll consumers to determine if they view the creation of an online account to be a point of opting into being remembered, recognized, and personally responded to.

[4] Though we didn't call them safety standards at the time.

[5] Psychologist George Levinger's model of human relationships. https://socialsci.libretexts.org/Courses/Achieving_the_Dream/Child_Development_(Cummings-Clay)/15%3A_Module_13%3A_Frameworks_for_Maturation/15.2%3A_Relationships_and_Families_in_Adulthood

[6] We used the term "Me2B" to underscore the deliberate attempt to alter the power dynamic between the vendor of the technology and the user of the technology, putting the "Me" first in the Me2B relationship. This was a deliberate synonym to the term Vendor Relationship Management (VRM). See also Flash Guide #7: The Me2B Lifecycle: Overlaying Social Norms on the Digital World - Internet Safety Labs

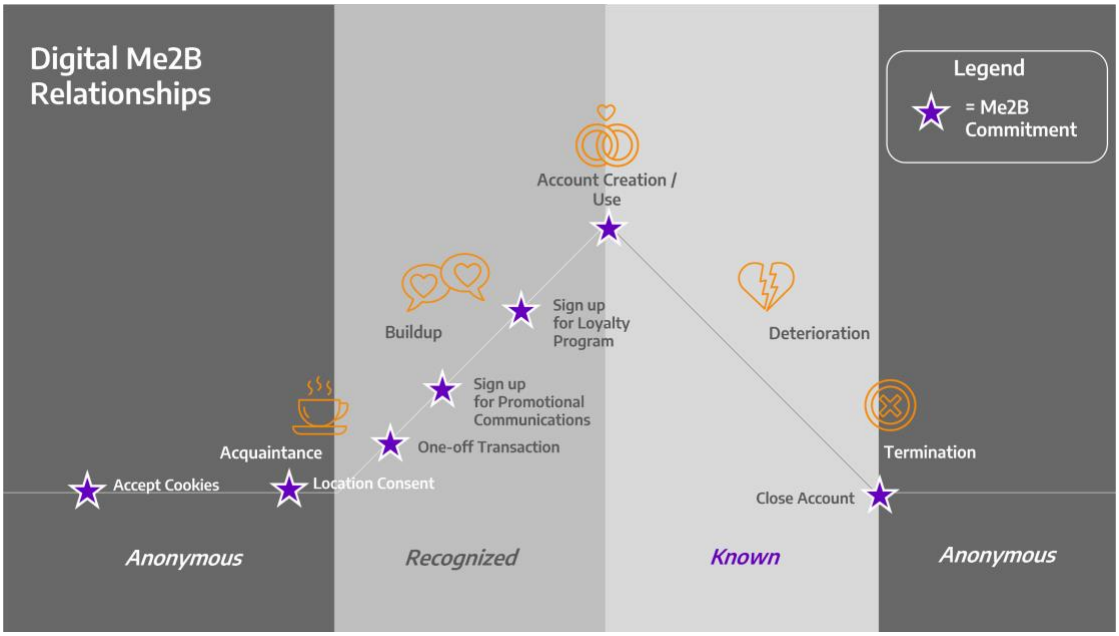


Figure 2: Digital Me2B Relationship Lifecycle

The ISL Principles of Safe Software include as principle #3, Identification Minimization. This means that at the points of "Me2B Commitments"—i.e. points in time during the relationship where we make an agreement to share information or other consideration in exchange for something of value to us—the identification performed is minimized and proportional to the commitment.

ISL Safe Software Principle #3: Identification Minimization

> *Any kind of identification performed **must** be proportional to the particular Me2B Commitment. Thus, the software **must** collect only the minimum set of identity attributes necessary to uniquely identify an individual [or device, or browser session] as needed for the particular Me2B Commitment.*[7]

The point is that the relationship develops organically over time, starting from a point of anonymity. ISL believes that people have a reasonable expectation and human right to only be *known* while in the logged in state. As will be shown in the remainder of this paper, this is far from the actual situation, which looks more like Figure 3.

[7] https://internetsafetylabs.org/resources/specifications/principles-of-safe-software/

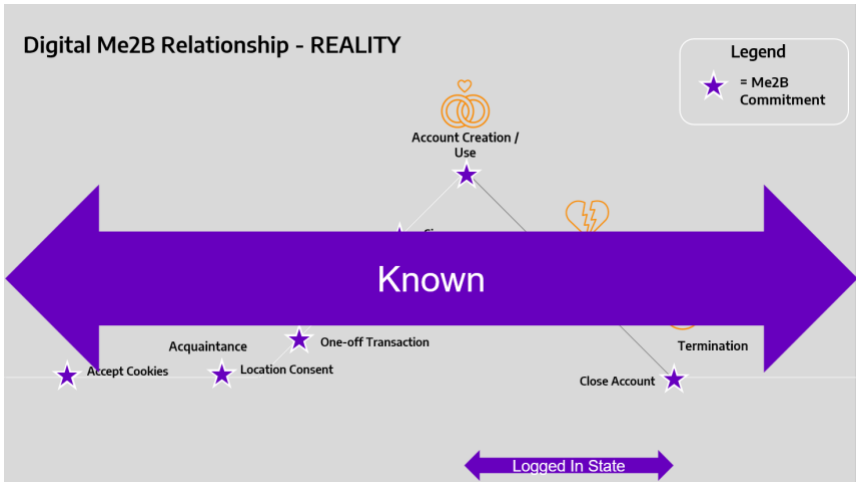


Figure 3: Digital Me2B Relationship Lifecycle in Reality

3 The Mechanics of Identification

There are many methods of digital identification, and new ones are born every day. Section 5 will examine the monetary and industry drivers for identification in more detail, but it's important to understand that, due to massive incentive there is endless creativity for digital identification mechanisms, whether they are visible or hidden. It's not an exaggeration to say that industry and government are addicted to certainty, which includes certainty of identifying people on the internet.

Figure 4 is a loose interpretation of invisible identification that can and does occur at each level of the OSI communication stack[8].

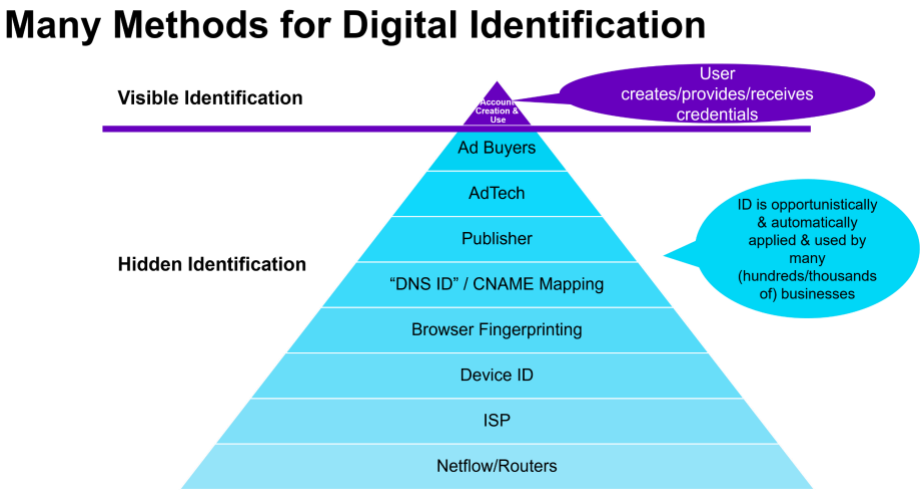


Figure 4: Identification by Communication Layer

[8] https://en.wikipedia.org/wiki/OSI_model

Each of these techniques can be studied in detail and are out of the scope of this paper. This paper's focus is the identification happening as a *coordinated synthesis of multiple applications or services*—i.e. the pervasive, hidden universal identification machinery that *runs above the application layer*.

3.1 Identification Numbering Schemas

For the purposes of this analysis, we'll describe two different scopes of identification schemas: local and universal.

3.1.1 Local Identification Schema

Local identification schemas are used in visible identification. Examples include Facebook's[9] or LinkedIn's internal number schemas. Local identification schemas are generally deployed by first parties, entities who have a data controller relationship with the data subject (the user of the technology). These first parties are typically responsible for the user data processing, including legal bases for such processing. Local identification schemas are enabled by and rely on identity and access management technologies. You'll have seen these local identification schemas at work, you need look no further than the URL for your Facebook or LinkedIn account, which generally ends with a unique identifier (name or number or combination of the two) that is hashed to a unique numeric identifier.

3.1.2 Universal Identification Schema

Universal identification schemas, on the other hand, are found in hidden identification infrastructures. Entities that have universal identification schemas typically have a 3rd party relationship with the data subject[10]. Universal identification relies on several technologies, including identity graphs and statistical methods for identity resolution. Identity resolution is the ability to determine, for example, that user 1234567 from Facebook is the same person as user 7654321 from Google; it's the application of multiple data points and statistical methods to develop quantitative confidence that two separate users from different services are in fact the same person. Identity resolution *resolves* instances of users [in the form of user identifiers] to specific actual identities.

---

[9] Note that a local identification schema can apply to all the products owned by the company. See how Facebook encourages users to correlate their Facebook and Instagram accounts: https://www.facebook.com/help/176235449218188/

[10] Greater scrutiny and legal testing must be performed to clarify if these entities are joint data controllers. Some precedent seems to exist for this.

Universal identification schemas are universal in the sense that they synthesize data from multiple (often hundreds) of service providers' user/customer data, which has local identification numbering on it. Thus, the universal identification schema synthesizes potentially hundreds of service providers' user/customer data into their own proprietary universal identification schema (Figure 5). Two important notes about universal identification schemas:

- They need user data to work. The way that identity resolution happens requires personal information. In fact, personal information is crucial. This is why we see many identity resolution platforms also offering customer data platforms.
- They rely on the first parties for "legitimate" user data collection. The privacy policy language that typically covers this data sharing from the first party will be language like, "we use this information to personalize your experience", or "we share data with our marketing partners", or "we collect this information to provide measurement, analytics and business services[11]. We will come back to this in greater detail in later sections.

Universal identification schemas have historically been the product of identity resolution and customer data platforms. However, there are now industry standards to further institutionalize pervasive universal identification (see section 5.4).

3.2 Example Identity Resolution and CDP Companies

In February 2024, Cracked Labs published their report on "Pervasive identity surveillance for marketing purposes: A technical report on personal data processing for LiveRamp's "RampID" identity graph system based on an analysis of software documentation with a focus on Europe." As the title indicates, the report is a deep dive on LiveRamp's identity resolution capabilities[12], with a focus on their universal identification schema, RampIDs. The report describes in detail the various ways that user data is sent to LiveRamp, through APIs, cookies, javascript tracking pixels, and the RampID itself being sent in the real-time bidding (RTB) protocol used in populating digital ads in online services. In other words, there is LiveRamp user-data-collecting "pixie dust" code scattered across all the various integration and customer partners websites, mobile apps, etc. LiveRamp indicates over 900 such partners[13].

[11] Meta privacy policy, https://www.facebook.com/privacy/policy?section_id=2-HowDoWeUse

[12] LiveRamp is an identity resolution company; provides an identity resolution platform and services.

[13] *https://investors.liveramp.com/static-files/be21b139-8c03-4e8a-b5de-87715b317870, pg. F-4*

Figure 5 below is an image from the Cracked Labs report showing sync partners in France that sync via cookies.

Source: https://crackedlabs.org/dl/CrackedLabs_IdentitySurveillance_LiveRamp.pdf, page 32

Figure 5: LiveRamp Sync Partners (France)

Figure 6 below is a conceptual model of how LiveRamp's universal identification schema integrates and synthesizes local identification schemas into their RampIDs. Again, the channels for sending user data into LiveRamp include APIs, cookies, pixels, and the RTB protocol.

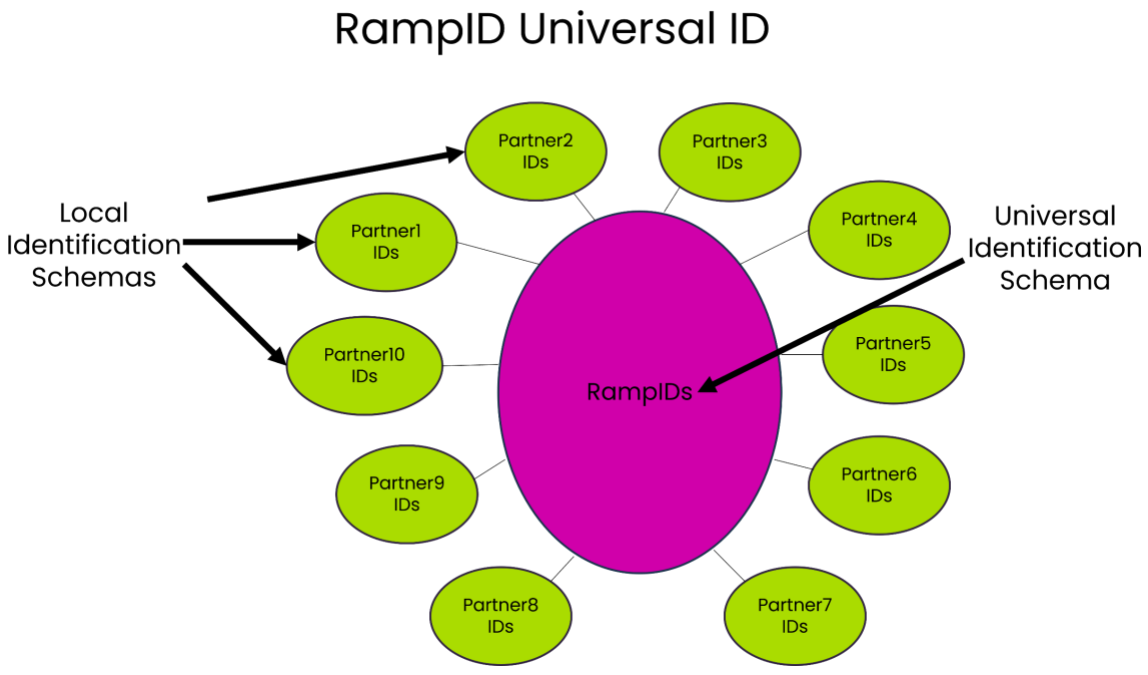


Figure 6: Conceptual Model of RampIDs as Universal Identification

ISL researchers mapped all 367 partners and customers found on the LiveRamp US websites (Figure 7). In the graph, blue lines represent so-called integration partners and yellow lines represent customer relationships.

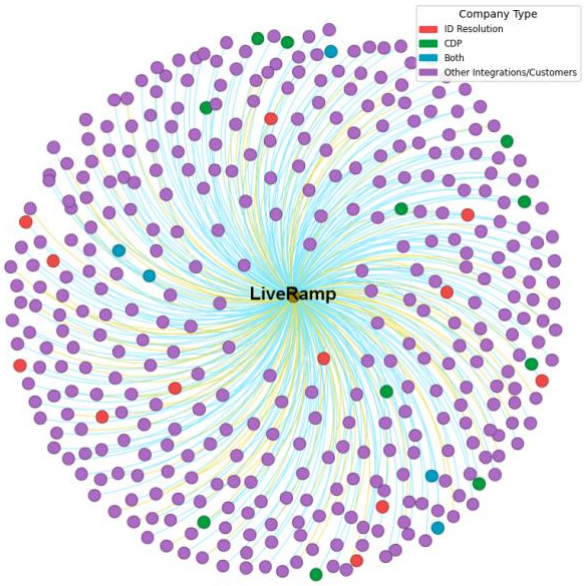


Figure 7: LiveRamp Integration Partners and Customers (n=367)

Similarly, our researchers mapped Twilio's 494 integration partners and customers listed on their US website (Figure 8).

Note that both integrate data from several identity resolution and customer data platform companies. In effect, this means that the networked hidden identification infrastructure (i.e. trans-platform communication) has the potential to be staggeringly large and comprehensive. (See sections 5.2 and 5.3 for discussion on the size of these companies.)

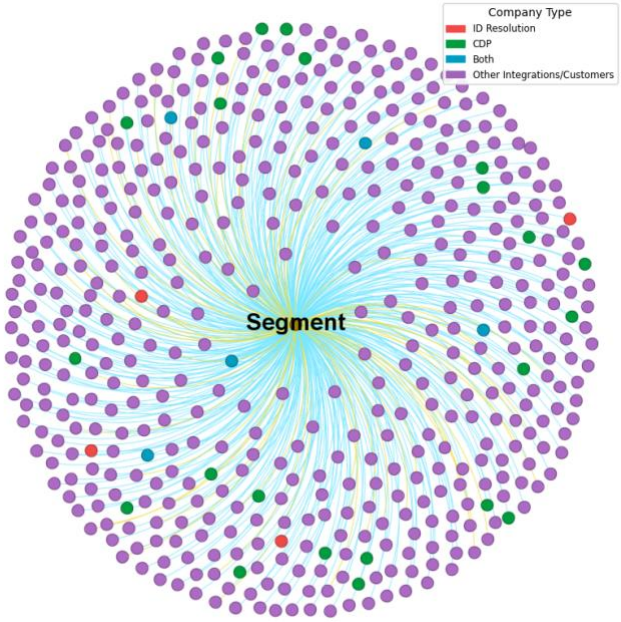


Figure 8: Twilio's Segment Integration Partners and Customers (n=494)

Figure 9 below shows the common integration partners across three identity resolution platforms (Segment, LiveRamp, and mParticle) analyzed by ISL. The cross-pollination of data sources/integrations across these platforms is evident. Note that of the nine common partners for all three platforms, five are either identity resolution or customer data platform companies.

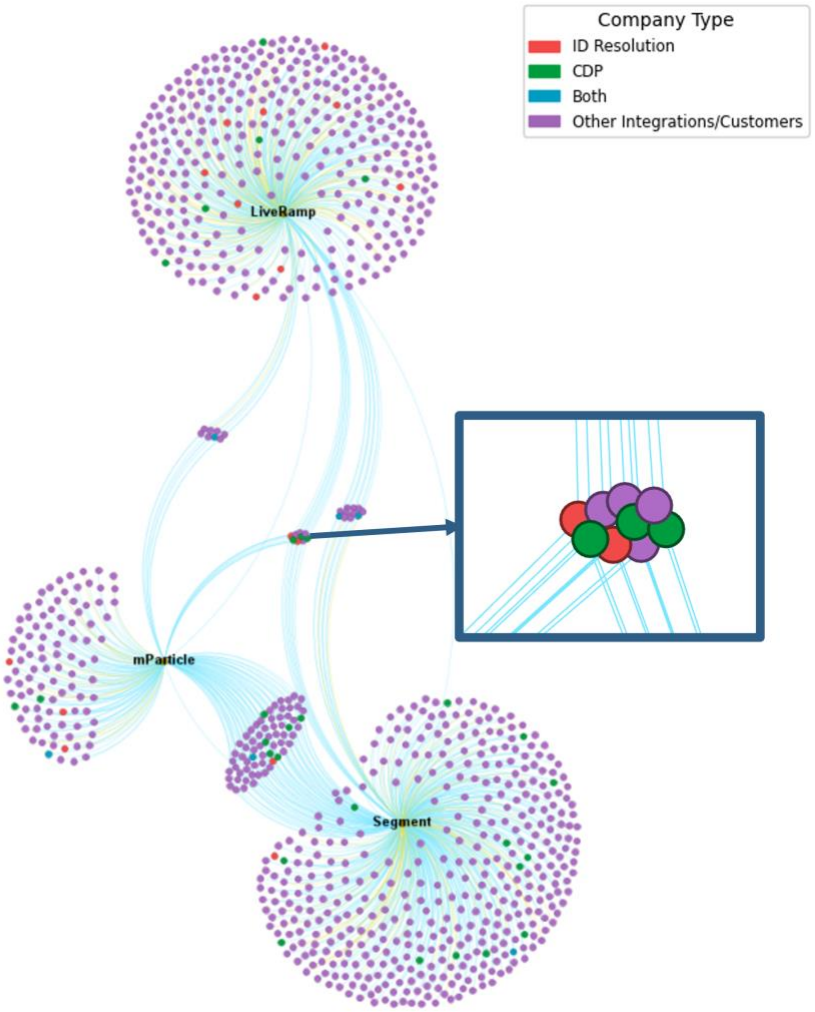


Figure 9: LiveRamp, mParticle, and Twilio's Segment Integration Partners and Customers

There is, however, a big difference between identity resolution or CDPs that sell consumer data and those that don't. LiveRamp, for instance, is a registered data broker[14]. Thus, a person using one of the many services provided by a LiveRamp integration partner will not only have data shared with LiveRamp, but likely is having her data sold to LiveRamp customers in the form of audience profiles aka "segments" (see Figure 10).

[14] https://www.oag.ca.gov/data-brokers?combine=liveramp

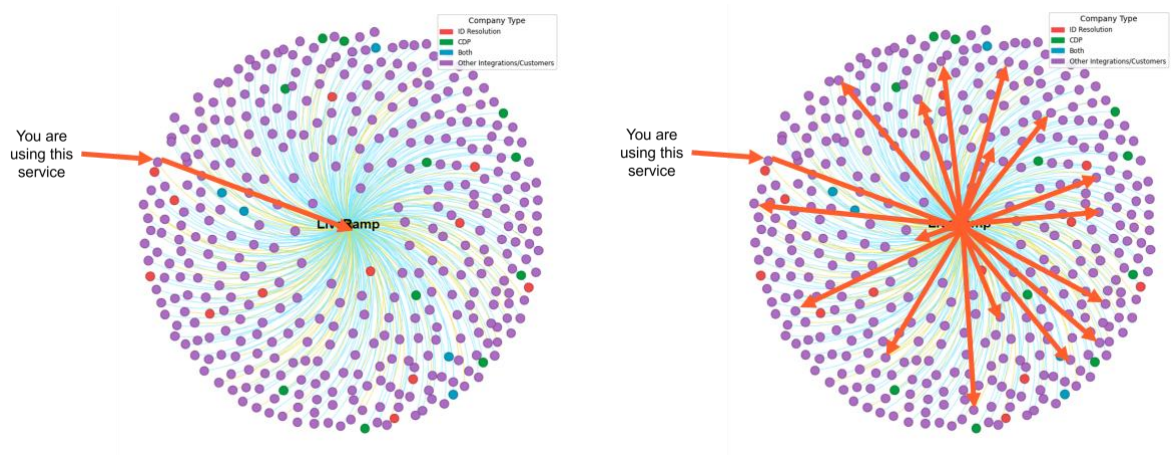


Figure 10: Data Broker Identity Resolution Platform Behavior

Whereas, if the identity resolution platform *isn't* a data broker, the user data is not further shared beyond just the identity resolution platform (Figure 11). Or at least so far as we know; ISL believes these types of platforms warrant greater transparency requirements and a greater burden on data subject permission. This will be discussed further in section 6.

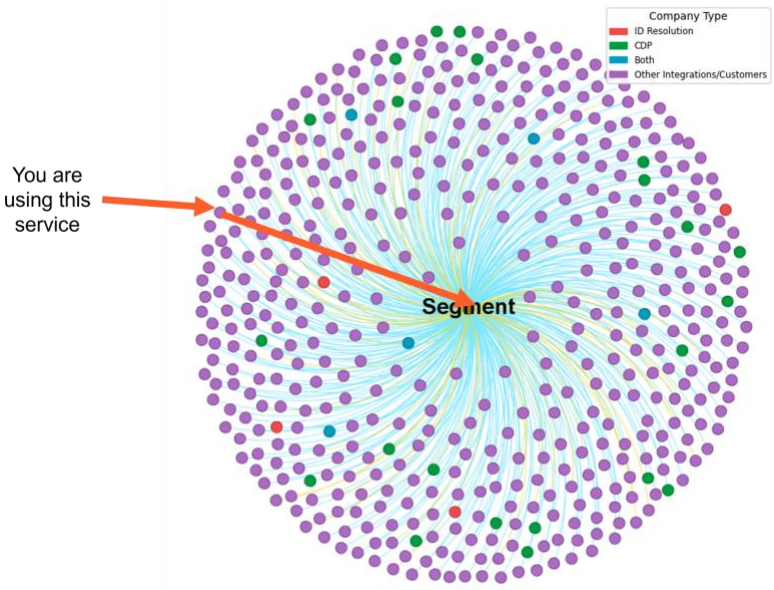


Figure 11: Non-Data Broker Identity Resolution Platform Behavior

4 The Business of Hidden Identification: Identity Resolution Platform and Customer Data Platform Industries

Right about now you may be wondering, how did we get here? The simple fact is that the infrastructure of hidden identification has been building at least since 2013. According to the Customer Data Platform Institute, the CDP category was named in

2013.[15] In their Q4 2023 Identity Resolution Landscape, Forrester calls identity resolution an "established technology market"—neither "emerging" nor "mature". These industries have been in existence for more than a decade.

One of the earliest (and ongoing) needs for CDPs was for intra-company use, to synthesize customer data across multiple fragmented internal systems. Customer relationship management was the original customer data platform. One can see the need for orchestration of customer data across large, complicated businesses with many brands and product lines; this seems somewhat benign and potentially beneficial to end users of technology, especially if they are existing customers. One can readily see how this architecture was easily extended to handle cross-company sharing of customer data.

But there's an even larger incentive driving identity resolution capabilities, and that's the insatiable desire of several large industries to know more about not just their customers, but also about potential customers or citizens—in short, *everyone*.

4.1 Industries that Need Hidden Identification

There are in fact many large industries that rely on hidden identification, including but not limited to industries that need to reduce risk, industries that want to influence existing and potential customers, and industries that amass information as a core business.

1. Industries that need to reduce risk:
    a. Insurance
    b. Government / Law Enforcement
    c. Finance
    d. Fraud
2. Industries that want to influence existing and potential new customers:
    a. Customer Relationship Management
    b. Advertising
    c. Marketing
3. Industries that amass information as a core business:
    a. Social networks
    b. Academic research & publication
    c. Legal research & publication

As can be seen in Table 1, the total worldwide market size of this set of industries that rely on hidden identification is nearly $9T, with an average annual growth rate of 15%.

[15] https://www.cdpinstitute.org/about/cdpi-backstory/

(To put this in perspective, the US GDP for 2023 was about $27.9T.[16]) Note that the customer data platform industry has a massive growth rate projection of 39.9% CAGR from 2024 through 2028.

Table 1: Industries Reliant on Hidden Identification

INDUSTRY	MARKET SIZE (WW)	CAGR
Insurance	$5,938,400,000,000[17]	10.0%
Law Enforcement Software	$1,490,000,000[18]	14.9%
Fintech	$279,740,000,000[19]	16.5%
Credit Reporting	$17,820,000,000[20]	4.8%
Fraud Detection	$25,670,000,000[21]	17.6%
AdTech	$987,520,000,000[22]	16.1%
MarTech	$325,000,000,000[23]	19.8%
CRM	$65,590,000,000[24]	13.9%
Social Media	$219,060,000,000[25]	14.8%
Academic Research Publishing	$27,000,000,000[26]	11.5%
Legal Publishing	$999,900,000,000[27]	5.3%
Identity Resolution	$1,151,000,000[28]	10.3%
Customer Data Platform Market	$5,100,000,000[29]	39.9%
	$8,893,441,000,000	***15.0%***

[16] https://www.bea.gov/news/2024/gross-domestic-product-fourth-quarter-and-year-2023-advance-estimate

[17] https://www.statista.com/statistics/1192960/forecast-global-insurance-market/

[18] https://www.fortunebusinessinsights.com/law-enforcement-software-market-105901

[19] https://www.fortunebusinessinsights.com/fintech-market-108641

[20] https://www.factmr.com/report/credit-reporting-market

[21] https://www.grandviewresearch.com/industry-analysis/fraud-detection-prevention-market

[22] https://www.grandviewresearch.com/industry-analysis/adtech-market-report

[23] https://www.grandviewresearch.com/industry-analysis/marketing-technology-martech-market-report

[24] https://www.grandviewresearch.com/industry-analysis/customer-relationship-management-crm-market

[25] https://www.thebusinessresearchcompany.com/report/social-media-global-market-report

[26] https://www.enago.com/academy/2021-stm-report-global-research-trends/

[27] https://www.contrivedatuminsights.com/product-report/legal-publishing-market-248735/

[28] https://www.businessresearchinsights.com/market-reports/identity-resolution-software-market-105139

[29] https://cdp.com/basics/cdp-industry-statistics/

Another driver for adoption of identity resolution and customer data platforms is the elimination of third-party cookies. Figures 12 and 13 below are notes from LiveRamp's and Segment's websites that explain how identity resolution is the solution for cross-site customer identification previously supported by third party cookies.


If the expectations weren't already high enough, companies are also up against constantly evolving data privacy initiatives, from legislation like the California Consumer Privacy Act(CCPA) to the impending death of the third-party cookie. If companies don't take steps to future-proof their technology in light of these privacy regulations, they'll get left in the dust.

The crux of the issue is this—for a business to get that elusive single view of the customer, you *need* to have good identity resolution.

https://segment.com/blog/identity-resolution/

Figure 12: Segment's 3rd Party Cookie Comment

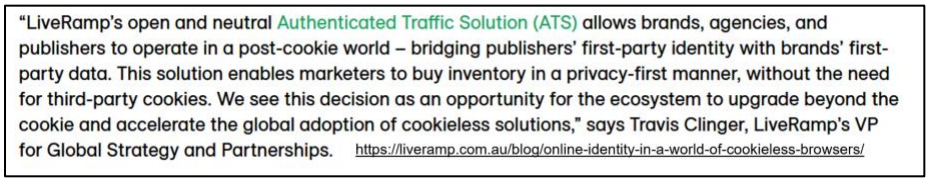
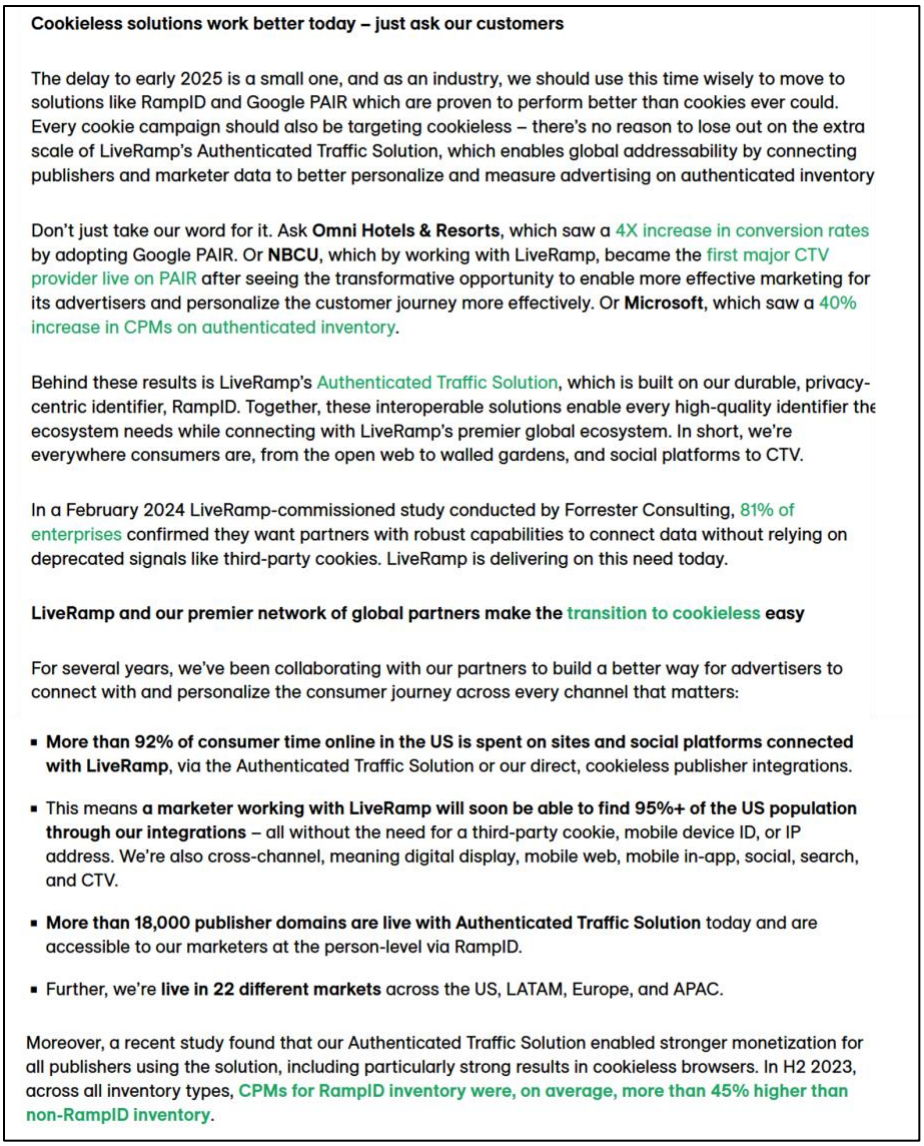

"LiveRamp's open and neutral Authenticated Traffic Solution (ATS) allows brands, agencies, and publishers to operate in a post-cookie world – bridging publishers' first-party identity with brands' first-party data. This solution enables marketers to buy inventory in a privacy-first manner, without the need for third-party cookies. We see this decision as an opportunity for the ecosystem to upgrade beyond the cookie and accelerate the global adoption of cookieless solutions," says Travis Clinger, LiveRamp's VP for Global Strategy and Partnerships. https://liveramp.com.au/blog/online-identity-in-a-world-of-cookieless-browsers/

Figure 13: LiveRamp's 3rd Party Cookie Comment

And one last explanation from LiveRamp about cookieless identification[30]:

[30] *https://liveramp.com/blog/120-days-to-prepare-for-chromes-deadline-why-marketers-should-embrace-cookieless-solutions-today/*



Cookieless solutions work better today – just ask our customers

The delay to early 2025 is a small one, and as an industry, we should use this time wisely to move to solutions like RampID and Google PAIR which are proven to perform better than cookies ever could. Every cookie campaign should also be targeting cookieless – there's no reason to lose out on the extra scale of LiveRamp's Authenticated Traffic Solution, which enables global addressability by connecting publishers and marketer data to better personalize and measure advertising on authenticated inventory

Don't just take our word for it. Ask **Omni Hotels & Resorts**, which saw a 4X increase in conversion rates by adopting Google PAIR. Or **NBCU**, which by working with LiveRamp, became the first major CTV provider live on PAIR after seeing the transformative opportunity to enable more effective marketing for its advertisers and personalize the customer journey more effectively. Or **Microsoft**, which saw a 40% increase in CPMs on authenticated inventory.

Behind these results is LiveRamp's Authenticated Traffic Solution, which is built on our durable, privacy-centric identifier, RampID. Together, these interoperable solutions enable every high-quality identifier the ecosystem needs while connecting with LiveRamp's premier global ecosystem. In short, we're everywhere consumers are, from the open web to walled gardens, and social platforms to CTV.

In a February 2024 LiveRamp-commissioned study conducted by Forrester Consulting, 81% of enterprises confirmed they want partners with robust capabilities to connect data without relying on deprecated signals like third-party cookies. LiveRamp is delivering on this need today.

LiveRamp and our premier network of global partners make the transition to cookieless easy

For several years, we've been collaborating with our partners to build a better way for advertisers to connect with and personalize the consumer journey across every channel that matters:

- **More than 92% of consumer time online in the US is spent on sites and social platforms connected with LiveRamp**, via the Authenticated Traffic Solution or our direct, cookieless publisher integrations.
- This means **a marketer working with LiveRamp will soon be able to find 95%+ of the US population through our integrations** – all without the need for a third-party cookie, mobile device ID, or IP address. We're also cross-channel, meaning digital display, mobile web, mobile in-app, social, search, and CTV.
- **More than 18,000 publisher domains are live with Authenticated Traffic Solution** today and are accessible to our marketers at the person-level via RampID.
- Further, we're **live in 22 different markets** across the US, LATAM, Europe, and APAC.

Moreover, a recent study found that our Authenticated Traffic Solution enabled stronger monetization for all publishers using the solution, including particularly strong results in cookieless browsers. In H2 2023, across all inventory types, **CPMs for RampID inventory were, on average, more than 45% higher than non-RampID inventory**.

Figure 14: LiveRamp's case for cookielessness

The takeaway is that industry need and desire for universal identification [regardless of Me2B relationship status, regardless of consumer awareness and/or permission] is here to stay. Attempts to eliminate it result in the emergence of other solutions—shut down 3rd party cookies, and other methods emerge. The financial incentives and industries are simply too large.

4.2 Scale of Identity Resolution and Customer Data Platforms

When we started this research, we'd never heard of identity resolution or customer data platforms, so we were interested to better understand these industries. Through our research, we identified 93 identity resolution platforms and 267 customer data platforms (ID Res & CDP Companies.xlsx) worldwide. In all, there were 340 unique companies with 20 companies providing both identity resolution and customer data platforms.

As shown in Table 1, the combined current market size of these two industries alone is over $6B, with the CDP industry expected to experience a shocking 39.9% CAGR from 2024 to 2028.

4.2.1 There Be Data Brokers

We were interested to understand the percentage of these companies that are or have been registered data brokers. Using the public data broker registries in California and Vermont, ISL determined that **8.2% (22 of 267) of CDP companies are registered data brokers, and 39.8% (37 of 93) of identity resolution companies are registered data brokers** (see ISL Identity Resolution and Customer Data Platform Companies list).

Should these kinds of companies be allowed to sell personally identifiable information? Should any kind of company be allowed to sell personally identifiable information? Should there be limits on both intended purpose and the types of companies that are allowed to purchase personally identifiable information?

4.3 Volume of Personal Information Trafficked

Another important indicator of the magnitude of the situation is how much personal information is being trafficked by CDPs and identity resolution platforms. Going back to LiveRamp as a representative identity resolution company, it professes to have over 5,000 data elements from hundreds of sources on about 700 million consumers worldwide[31]:

> • ***Extensive Coverage.*** We activate data across an ecosystem of more than 550 partners, representing one of the largest networks of connections in the digital marketing space. We use 100% deterministic matching, resulting in the strongest combination of reach and accuracy. Through our Data Marketplace, we offer multi-sourced insight into approximately 700 million consumers worldwide, and over 5,000 data elements from hundreds of sources with permission rights.

They also mention that they have 700,000 *segments*[32]. In June 2023, The Markup reviewed a Xander database which had 650,000 segments.[33] With this kind of granularity, however, one wonders if these can really be regarded as "segments" so much as personal descriptions. Plus, the data in the segments includes sensitive information such as religion, medical conditions, race, age, and gender. In the hands

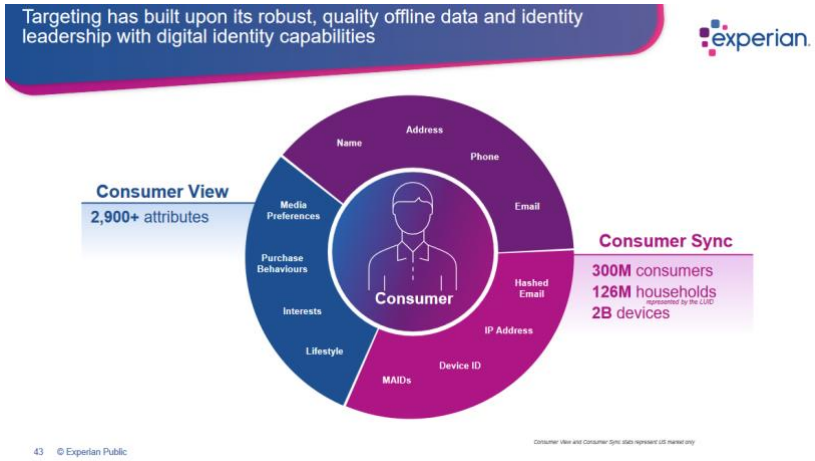
[31] Source: https://www.sec.gov/Archives/edgar/data/733269/000073326921000017/ramp-20210331.htm

[32] *https://investors.liveramp.com/static-files/be21b139-8c03-4e8a-b5de-87715b317870, pg 12-13*

[33] https://themarkup.org/privacy/2023/06/08/from-heavy-purchasers-of-pregnancy-tests-to-the-depression-prone-we-found-650000-ways-advertisers-label-you

of data brokers, this data can be used to target Jewish or Palestinian people, or elderly women with heart conditions, etc.

Another prominent personal data company, Experian, professes to have data on 1.5B customers overall, and calls out 300M consumers, 126M households, 2B devices and 1,900+ attributes per consumer in their marketing data product.[34]



These are just two of 93 identity resolution platform companies. Experian no doubt has unique insight into consumer spending behaviors and data points due to their privileged access to credit card usage information[35]. Other identity resolution companies have special relationships and expertise in other verticals. Thus, the intercommunication between these organizations is vital to paint an increasingly holistic picture of consumers.

Particularly disturbing are the identity resolution platforms that specialize in health care information, such as Redpoint. While recognizing that synthesizing patient data is indeed important for patient care, care must be taken, such as mandatory usage constraints must be in place. Should personal health information ever be monetized?

---

[34] https://www.experianplc.com/content/dam/marketing/global/plc/en/assets/documents/results-and-presentations/2024/experian-roadshow-jan-to-mar-2024.pdf

[35] Since this information comes from retailers and not directly from data subjects, ISL wonders if the ultimate marketing usage of credit card transactions has been consciously consented to by credit card users. This scenario exemplifies key problems with commercial surveillance: (1) tracking data flow from source to sale, and (2) having viable consent for ultimately sold personal data.

4.4 Standardization of Universal Marketing Identification Schemas

The scale and cooperation among identity resolution and CDPs is staggering. To cap it off though, these industries have been developing open universal identification standards since about 2021:

- **Unified ID 2.0** https://unifiedid.com/
- **European Unified ID** https://euid.eu/

These open standards are poised to institutionalize universal identification of every human on the planet for marketing and advertising purposes.

The justification? Throughout companies like LiveRamp and mParticle's regulatory filings one can see repeated mention of consumer demand for more relevant advertising, more personalized services as a justification for this staggering breach of privacy and trust. More will be said of this in section 6 below.

In general, both of these standards codify universal identification schemas for interoperable use throughout all of the marketing and advertising technology in the world—well beyond just the identity resolution and CDP companies.

Unified ID 2.0 (UID2) claims to offer transparency and control to users (data subjects). ISL was curious to see what it entailed and found that clicking on the button opened this page (Figure 15).

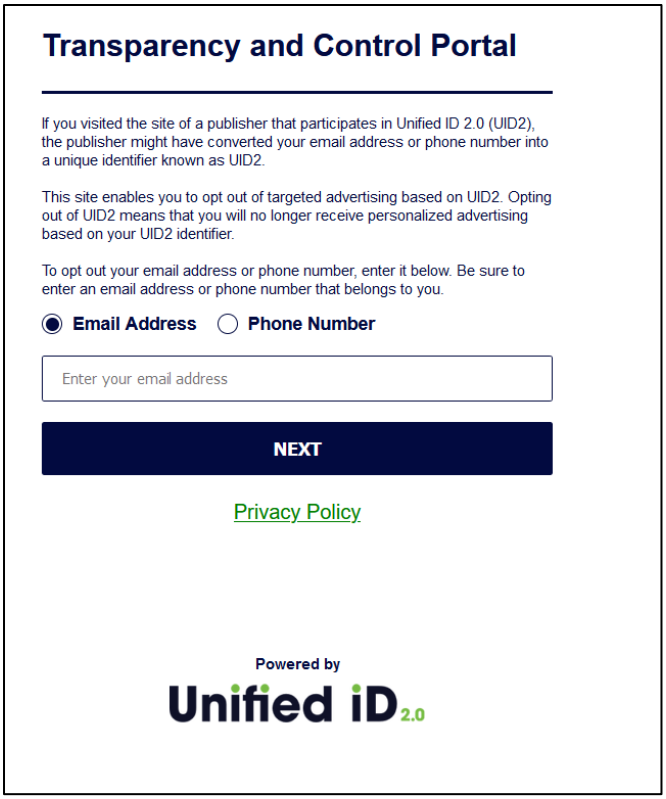


Figure 15: Unified ID 2.0 "Transparency and Control Portal" (https://www.transparentadvertising.com/ or saved page: /web/20240609233528/https://www.transparentadvertising.com/)

ISL has many questions:

- Who are the sites that participate in UID2?
- Why wasn't I notified at the time a UID2 was created? And why wasn't I given an opportunity to opt out from the start?
    - Does the Global Privacy Control signal from my browser count as a signal that I do NOT want a UID2 or EUID created? Can it?
- If I enter my email address or phone number am I then *adding* that information to their identity graph?
    - Weirdly, the privacy policy that applies to the page shown in Figure 15 (https://www.thetradedesk.com/us/website-privacy-policy) explicitly [and in circular fashion] states that:

> *"**UID2 and EUID***
>
> *We may use your email address to create a consistent unique identifier called a UID2 or, in Europe, a EUID. These allow us, and other participants in the UID2/EUID advertising framework, to ensure that the ads you see online are relevant and measure the effectiveness of our advertising.*
>
> *You can read more about how UID2 works on Transparentadvertising.org.*
>
> *You can read more about how EUID works, how to opt-out and your rights on the EU Transparency website. "*
>
> https://www.thetradedesk.com/us/website-privacy-policy (saved /web/20240609233432/https://www.thetradedesk.com/us/website-privacy-policy)

5 Consumer Awareness and "Demand" for Commercial Surveillance

As noted earlier, ISL found repeated claims by identity resolution and customer data platform companies that consumer demand for personalization was the driving force for marketing surveillance. But *do* consumers demand personalization?

Most of the research comes out of industry and thus is designed to support the conclusion that consumers *do* want personalization. In a 2021 survey from

McKinsey[36], consumers seem to welcome the uncanny valley of being "known" on first time purchases.

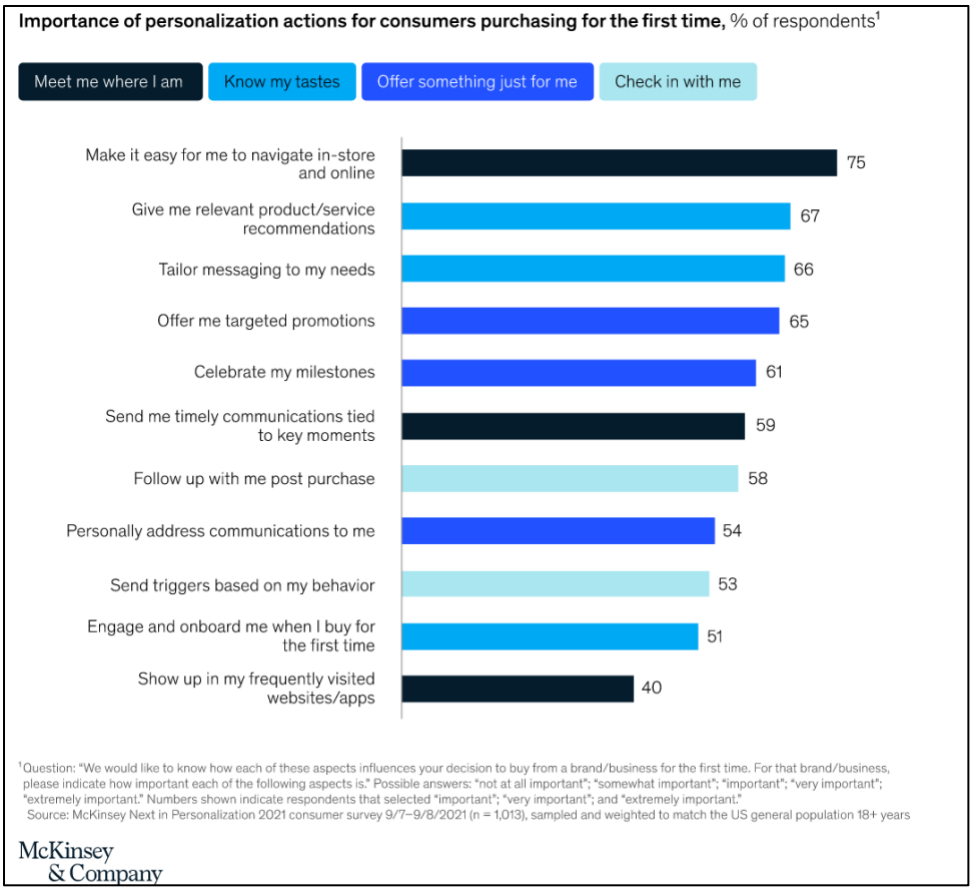


Figure 16: "The value of getting personalization right—or wrong—is multiplying", McKinsey. November 12, 2021

ISL wonders if respondents fully understood what each item means and what each necessarily entails. For example, "Celebrate my milestones" is reported to be important/very important/extremely important *for the first purchase.* This seems a peculiar preference on the *first* purchase. Also, if people knew that the cost of being intimately known in an initial digital transaction means that technology is surveilling them at scale and indiscriminately sharing personal information far and wide, one wonders if consumers would view the tradeoff as worthwhile.

[36] "The value of getting personalization right—or wrong—is multiplying", McKinsey, November 12, 2021 https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/the-value-of-getting-personalization-right-or-wrong-is-multiplying

A 2019 report from the Customer Data Platform Institute[37] shows mixed consumer receptivity to personalization. Figure 17 shows tolerances that seem to support the Me2B relationship where personalization grows with time, i.e. with the state of the relationship, with only 39% of US respondents expecting personalized experiences, and only 26% of European respondents.

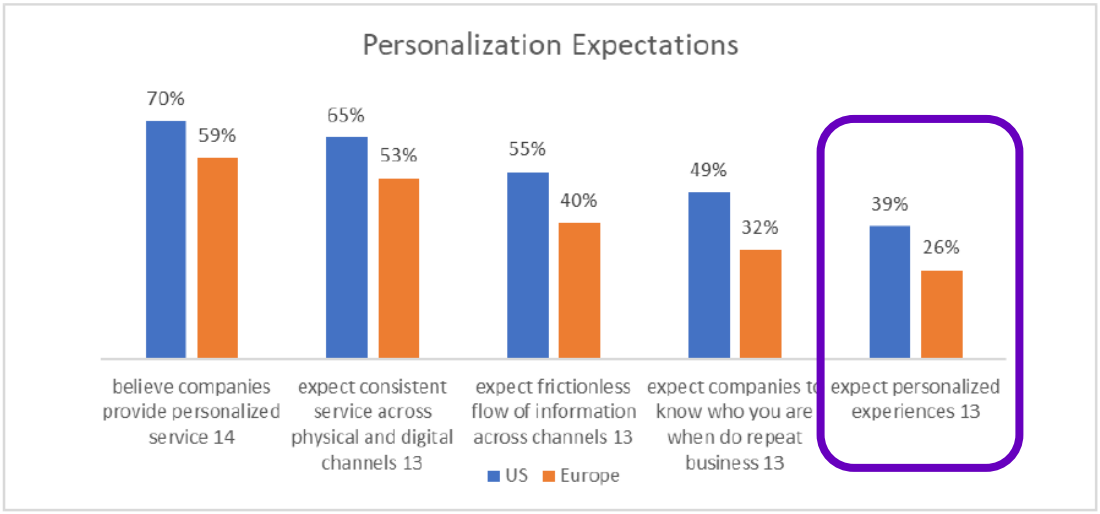


Figure 17: Personalization Expectations, CDP Institute[38]

Another telling finding in the report shows that only 27% (US) and 28% (European) of consumers value anticipation of needs as the important part of personalization. People perhaps don't want to be predicted or predictable (Figure 18).

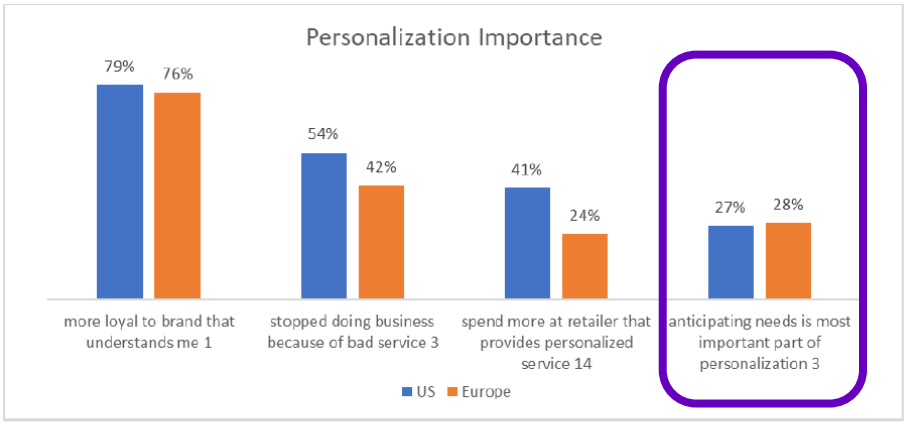


Figure 18: Purpose of Personalization Importance, CDP Institute Report

[37] Report: Consumer Perspectives on AI In Marketing - CDP.com

[38] Report: Consumer Perspectives on AI In Marketing - CDP.com page 2

Overwhelmingly, respondents in both the US and Europe agree that brands should not use personal data to market to them (Figure 18).

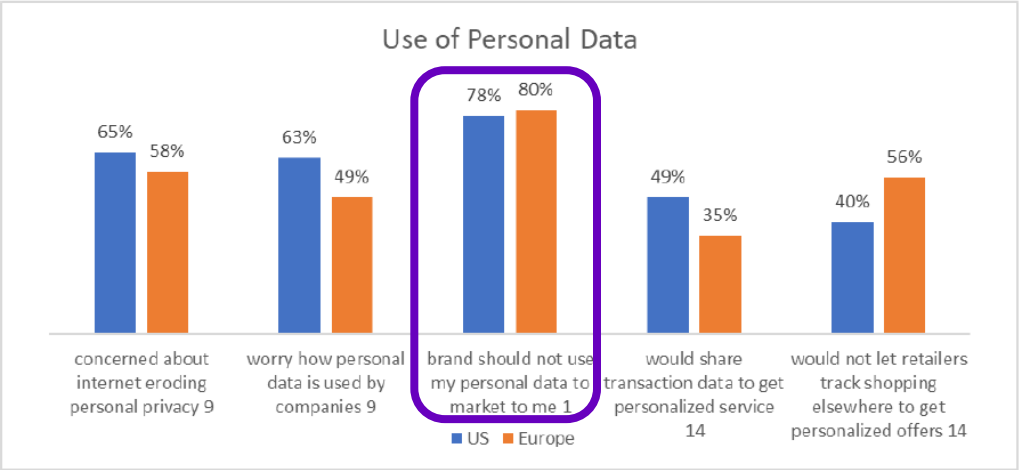


Figure 18 CDP Institute Report

Finally, of note, there is a stark difference between respondents in the US and Europe regarding their belief whether or not organizations sell their personal data with only 45% of US respondents but 68% of European respondents (Figure 19). It seems that Europeans have a more accurate perception of the situation.

Figure 19 CDP Institute Report

Overall, the data from the CDP Institute report suggests that people—especially in the US—are neither aware of nor particularly demanding "always on" personalization.

5.1 What About Consent?

Another consideration of this situation has to do with consent by the data subjects who are being surveilled by this marketing data collection infrastructure. Part of the appeal of the lucrative identity resolution and customer data platform businesses is

that they necessarily have an *indirect* relationship with the data subjects, and they rely on the first-party entities to import personal information into their platforms. Thus, they rely on those first parties for legally obtaining consent for data processing of personal information.

But do consumers really understand the magnitude of what's happening with their personal information—the orchestration and synthesizing? Can they? Do the first parties fully understand where their customer data ends up? Can consent be possible under these conditions?

ISL looks forward to machine generated records of processing activities (ROPAs) that show the data trail from first party to identity resolution platforms three (or more) degrees of separation away.

6 Risks of Identity Resolution and Customer Data Platforms

The sheer magnitude of universal identification without consumer awareness or permission and the volume of personal data being shared, aggregated and processed without appropriate regulation is indeed overwhelming. In this section, we highlight the key risks.

Risk #1: Loss of Consumer Privacy

These platforms amass personal information even when people aren't signed into their accounts.

Moreover, identity resolution platforms synthesize real world personal information (address, property records, etc.) with digital platforms. The result is a complete loss of privacy. People continually ask if their devices are listening to them. The fact is that their devices aren't listening—because they don't have to. Tracking your behavior across hundreds of digital services/websites/devices creates a very robust profile of us. How robust? We're just beginning to understand by exercising rights in states that allow us to obtain information about what companies know about us.

One of the "features" of hidden identification, as compared to visible identification, is that it tolerates inaccuracy. Indeed, data broker data is notoriously inaccurate—or at a minimum, out of date. But ISL wonders with increasing real-time sharing and triangulation of users, coupled with the powers of machine learning and AI, how much longer user profiles will continue to be "noisy" with inaccurate information. Moreover, data being inaccurate isn't the abuse here. The abuse is the massive data harvesting infrastructure and its attendant violations.

Risk #2: Lack of Data Subject Permission / Failure of Consent

As discussed above, it's highly unlikely that a reasonable person would ever agree to the data sharing behavior described in this paper (Figure 10, for example), and no privacy policy in the world adequately articulates the sharing/mixing/matching going on with user data in the hands of identity resolution and customer data platforms.

Software is reaching (if not already surpassing) the limits of "knowability". Executable code like the real-time bidding process entails a different and unpredictable set of third party bid recipients every time an ad is filled and presented to the user. It's arguable if we consumers *ever* really understood technology behavior sufficiently to warrant meaningful consent. In fact, it's arguable that tech *makers* understand their technology behavior enough to accurately describe the behaviors in the privacy policy and terms of service. Technology is only getting more complicated, and consent is an abject failure.

Risk #3: Inadequate Governance Including and Extending Beyond Data Brokers

In the US, data broker governance is fragmented (currently only in four states, and no federal governance) and what exists is inadequate. The penalties are trivial, and under legal definition, 1st parties can't be data brokers, when we know for a fact this happens at massive scale[39]. But as this paper and others have shown, platform behavior goes unscrutinized by external, objective experts. This must change.

Risk #4: Loss of Consumer Autonomy

In the 1950s when television was taking root in US households, there was substantial concern and fear over the use of subliminal messages in advertisements. I find it interesting that we are seemingly so inured to advertising now that we accept pervasive surveillance for the very purpose of manipulating us, as prospective or current customers. Most of us have clicked on personalized ads based on surveilled profiles of ourselves; that is manipulation. Whether we like it or not is beside the point. We no longer seem to recognize where the manipulation begins and ends.

Risk #5: Hidden Identification Indiscriminately Collects Data

The greatest risk of the unfettered infrastructure of hidden identification is that it collects data for any human using connected services, without regard, including the data of children using internet-connected services.

---

[39] https://www.nytimes.com/2020/02/28/technology/fcc-cellphones-location-data-fines.html

6.1 EdTech Apps Communicating with Identity Resolution and Customer Data Platforms in ISL 2022 EdTech Benchmark

We went back through the network traffic that we collected in our 2022 K-12 EdTech safety benchmark and found that **539 apps (35% of 1538 total) sent data to identity resolution platforms or customer data platforms** (see 2022 EdTech Benchmark ID Res & CDP Data.xlsx for the list of apps). The apps were nearly evenly split by operating system with Android apps comprising 44.9% of the apps, and iOS apps comprising 55.1%.

Community Engagement Platforms (aka school utility apps), EdTech "other", and Non-Education Specific News apps had the most apps sending data to either an identity resolution platform or a customer data platform (Figures 20 and 21, below)

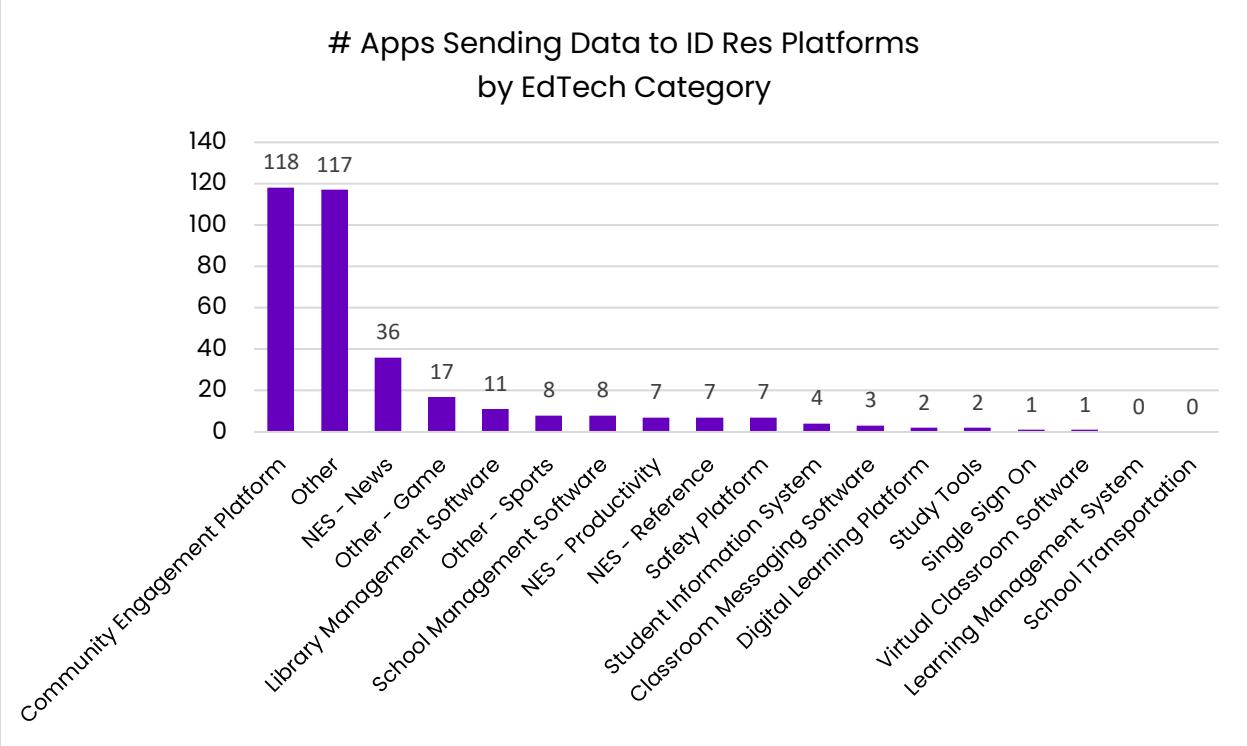


Figure 20: EdTech Apps Sending Data to Identity Resolution Companies, ISL 2022 EdTech Benchmark Data

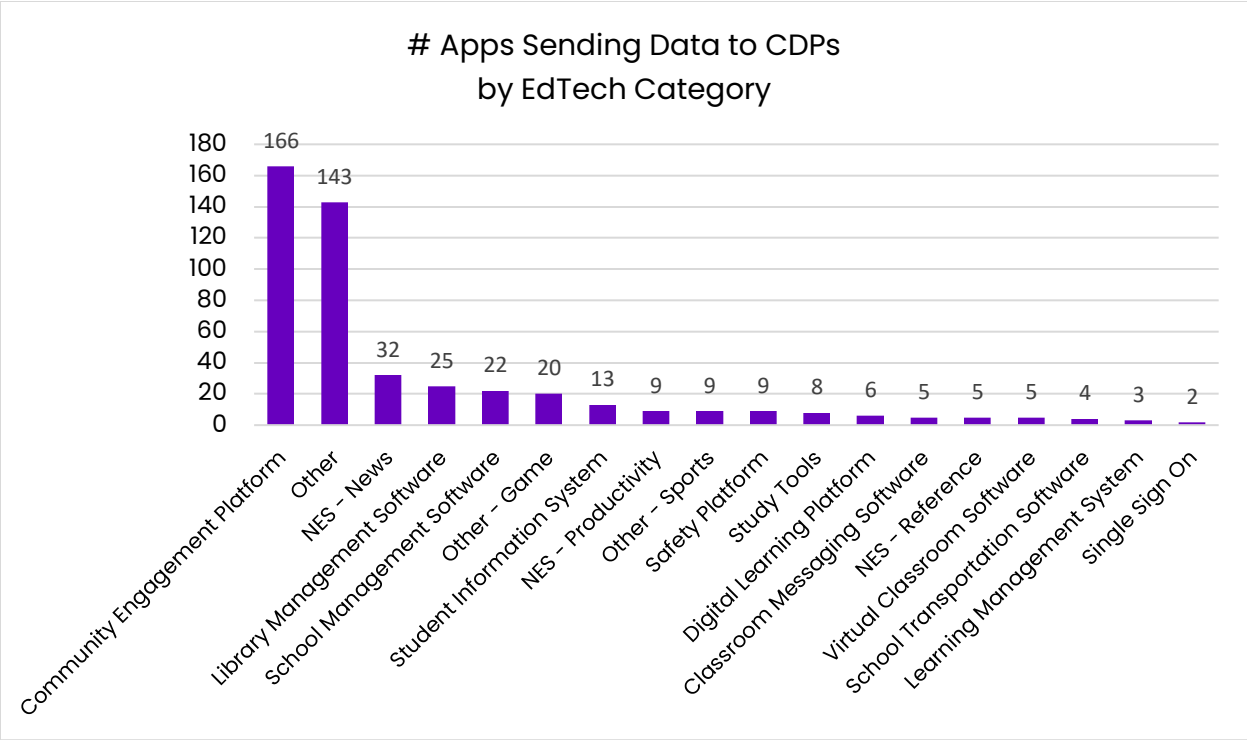


Figure 21: EdTech Apps Sending Data to Customer Data Platforms, ISL 2022 EdTech Benchmark Data

In total, we observed 66 unique companies providing either identity resolution platforms (28), customer data platforms (29), or both (9) (see Appendix A for list of identity resolution and customer data platform companies found in the benchmark data).

The 25 most frequently observed identity resolution or customer data platforms in the 2022 benchmark network traffic are shown in Figure 22. Note that **eleven of the 25 most frequently observed companies are registered data brokers** (the orange-colored bars in the chart denote data brokers).

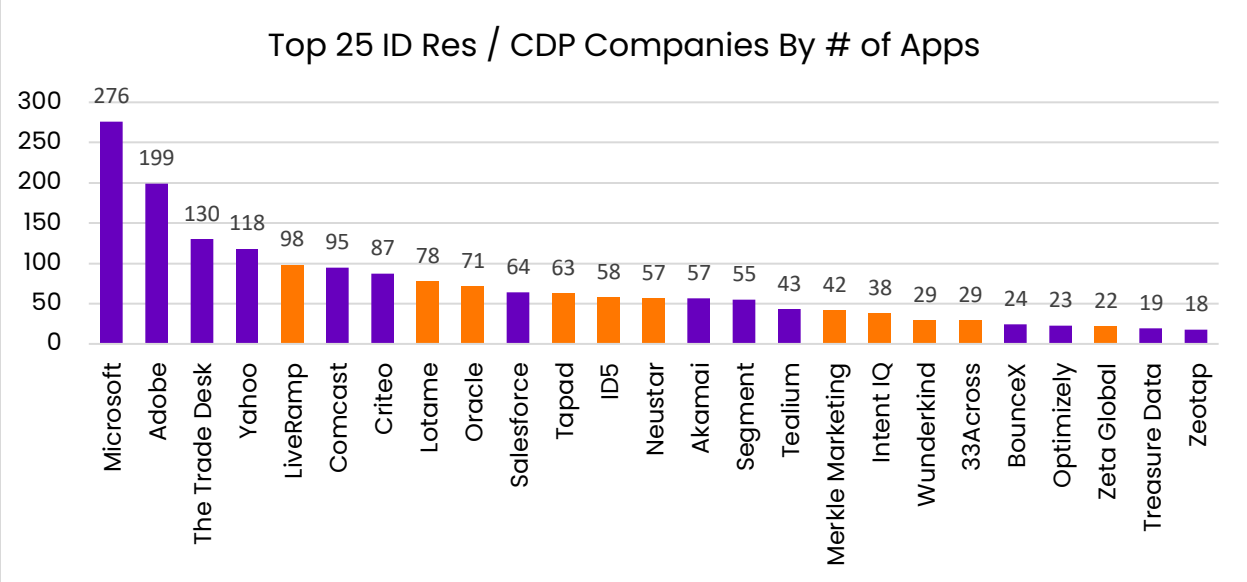


Figure 22: Top 25 Identity Resolution or CDP Companies Receiving Data in 2022 EdTech Benchmark

7 Next Steps

7.1 Call to Action

ISL conducted this research to help illuminate the sizable risk of hidden identification and the worldwide web of user surveillance. ISL believes naming and exposure is crucial to effecting change. Identification resolution and customer data platforms have been hiding in plain sight for more than a decade, and yet even the "identerati" are largely unfamiliar with these industries. How can we expect everyday people to know?

This paper is a rallying call for all privacy advocates to come together to demand greater regulatory scrutiny, transparency and oversight for these industries, in conjunction with more meaningful data broker regulation.

Additionally, this is a rallying call to acknowledge the catastrophic failure of notice and "consent" as a valid permissioning mechanism for digital services.

We must ask ourselves if this is the kind of world we want for ourselves and our children, where our preferences, practices, relationships, behaviors, and beliefs are all up for sale and broadly shared without our awareness. Are we ourselves in fact being sold?

The technologies fueling these capabilities have received billions of dollars; consumers don't have a chance in the face of voracious hunger to identify, know, and manipulate them. We hope that this research shines a much needed light on the forces enabling the worldwide web of human surveillance.

Appendix A: Identity Resolution and Customer Data Platforms Found in ISL's 2022 K-12 EdTech Benchmark

Identity Resolution Companies
33 Across
AddShoppers
Adstra
Akamai
Bluecore
BounceX
Cloud.IQ
Comcast
Crimtan
Criteo
Epsilon
Eyeota
ID5
Identity Resolution
Intent IQ
LiveRamp
Lotame
Lytics
Marketo
Merkle
MetaRouter
mParticle
Narrative
Neustar
OnAudience
Roq.ad
Segment
Signal
Tapad
Tealium
The Trade Desk
Transunion
Treasure Data
Wunderkind
Yahoo!
Zeotap
Zeta Global

CDP Companies
Accoustic Connect
Acquia
Adobe
Apache
Appier
BlueConic
Blueshift
Cloud.IQ
Customer.io
Emarsys
Epsilon
esri
Fullstory
GALE
Klaviyo
Lytics
mediarithmics
MetaRouter
Microsoft
mParticle
ONEcount
OpenText
Optimizely
Oracle
Piano
Precisely
Raptor
Reach
Rudderstack
Salesforce
Segment
Splunk
Tealium
Terminus
Treasure Data
VWO
Zeotap
Zeta

Companies with Both
Cloud.IQ
Epsilon
Lytics
MetaRouter
mParticle
Segment
Tealium
Treasure Data
Zeotap