

Q1 2022

# MediaWallah ID Benchmarks

The Market Is Betting  
on Three IDs, But How  
Reliable Are They?



KEY  
FINDINGS

## No Identifier Is a Home Run, Requiring Expertise and Technology

To achieve scale and accuracy as third party cookie use declines, companies will likely continue to use all three identifiers in the near future. However, each ID comes with unique challenges which should guide future identity strategies.



**6% average mismatched HEMs poses challenges to scale.**

HEMS are highly persistent. But the rate of observed mismatches means millions of unusable HEMs, and challenges for using HEMs as a scalable anchor for linkages.



**6 months is the typical lifespan of MAIDs.**

MAIDs last less time than companies may expect, with persistence decreasing by 3x in the past two years.



**18% of IPs are not tied to a home.**

They require technical expertise to identify one-to-one from one-to-many instances, but can be highly reliable if that issue is solved for — and are valuable for CTV.

“

**Brands, agencies and publishers are ramping up their data collection to support identity-driven digital advertising as third-party cookies become less reliable. Our research provides benchmarks for three of the most common identifiers, and provides insights to help them prioritize to maximize accuracy, persistence and scale.**

**We've found that each identifier has benefits and challenges – challenges including distinguishing the reliable data from the not reliable enough, along with the problems of decreasing identifier persistence and stability. These consideration can be overcome, but require expertise and specialized technology to handle especially given the immense volume of data.**

”

— **BOBBY ATEFI**, *Chief Data Scientist, MediaWallah*



DEEP  
DIVE

# The Market is Betting On Three Leading Identifiers: HEMs, MAIDs and IPs

## *But How Reliable Are They?*

With third party cookies in decline, marketers are turning to three other common identifiers as their replacement: HEMs, MAIDs and IPs:

**HEM**Hashed Email  
Address**MAID**Mobile  
Ad-ID**IP**IP  
Address

These identifiers are common match keys for first party data collection and identity graphs—used to connect other identity data for advertisers to target audience, gain insights and measure campaigns across the digital advertising landscape.

This report looks at the stability and accuracy of these three identifiers, provides the pros and cons of their use and helps advertisers, publishers and tech companies prioritize the identifiers that will be most helpful to their business efforts.

**Methodology:**

MediaWallah's data team analyzed hundreds of millions of US-based HEM, mobile ad ID, and IP data collected between 2020 and 2021.

**DEEP  
DIVE**

# Not all First Party Data is Created Equal

Each identifier is subject to a variety of elements making data easier or harder to collect and use.



## Privacy and Permissioning

HEMs require discreet permission from a user, while IPs and MAIDs are able to be collected passively from a device or internet location.



## Longevity

Different identifiers last longer than others. While many people keep an email for a long time, they may change jobs and get a new one. MAIDs and IPs turn over for a variety of reasons.



## One-to-One / One-to-Many

While most people have one phone, they don't always have one MAID. Conversely, a household or business may have one IP that accounts for several individuals.

Using these factors as guidance for our research, we looked at the relative ease of collection, persistence and ability to reach an individual or household across these three leading identifiers.

DEEP  
DIVE

ID 1: HEMs

## Poor Infrastructure Creates Needless Loss

Median observed mismatch rate is 6% - translating to millions of HEMs that are unused even if they are good IDs.

**HEMs are high-value, but are often handled poorly causing mismatches and missed opportunities:** If the hashes are done incorrectly, HEM data sets will not match together correctly, generating mismatched HEMs, or HEMs that incorrectly align two HEM emails together.

Bad HEMs happen because Publishers and Data Providers providing the data aren't careful about technical procedures:

- Failing to filter out dummy/ "garbage" emails (as when a user signs up with a false email)
- Hashing null or blank values – eg the user leaves a form blank and the data provider hashes that blank spot as if it were an email address
- Casing/trimming before hashing

### GETTING IT RIGHT MAKES A BIG DIFFERENCE:

For one MediaWallah client, a derived, authenticated HEM ID in the bid stream allowed

### **15x match rates**

between brand and publisher audience data sets (in actual bidding) vs. using 3rd party cookies during the same time period.



DEEP  
DIVE

## ID 1: HEMs

# High Value, But Need Structure Across Industry

## Pros:

- Consented / permissioned = privacy-compliant mechanism
- Persistent: Once an email is created, it doesn't change or go away
- Tied to an individual
- Platform agnostic, not tied to a device or browser. Because of this it can be used for targeting and measurement across consumer touchpoints/screens



## Cons:

- Low scale
- Expensive to collect
- Requires privacy regulation compliance
- Data supplier errors can lead to incorrect email hashing, meaning it's difficult to match the data together – rendering it far less usable or even unusable.



DEEP  
DIVE

## ID 2: MAIDs

# Decreasing Longevity, But Easy to Collect

A surprisingly short persistence hurts their value.

When comparing MAIDs to the other IDs, there has been an increased number of MAIDs per HEM over time, which means there are more identifiers to connect in order to define a person. This increases the complexity of the data graphing needed to target somebody correctly.

- 6 MAIDs per HEM in 2021
- 2 MAIDs per HEM in 2020
- 3x more MAIDS per HEM in 2021 vs. 2020

**The median lifespan of a MAID has decreased, from about 1.5 years to only half a year.**

## REASONS FOR DECREASE IN LIFESPAN:

- User resetting their MAID manually
- App / session-based ID being confused with MAIDs leads to inflated numbers



DEEP  
DIVE

## ID 2: MAIDs

# Small Screen ID of Choice, but Low Persistence

## Pros:



- Accessible / integrated into ad tech platforms
- Good for targeting people via mobile apps
- Thought of as more persistent than 3rd party cookies
- Rich data tied to devices (precise location, internet connection) – which in turns translates into more efficient targeting and attribution, including for in-store/brick and mortar sales impact

## Cons:



- Limited to mobile app environment for activation
- Decreasing persistence, caused by certain platforms making it easier for users to block (iOS)
- Like 3rd party cookies, this identifier is “going away” in some platforms (iOS)
- Dependent on large company decisions (Google, Apple, etc.) in terms of viability
- File delivery issues: Often shared in non-standardized formats via batch file without definitive timestamp – not via real-time SDK



DEEP  
DIVE

ID 3: IPs

## Expertise Required to Maximize Value, Tie to Other IDs

**About 25%**

of IP addresses change location in a 2 year span.

Of the IPs that did change their location, over

**one third**

of them spent fewer than 3 months in any given location.

**About 18%**

of IPs are not tied to a home, important to differentiate for ad messaging and targeting.

**HOW MANY IPS ON AVERAGE ARE TIED TO A DEVICE, COOKIE OR EMAIL?:**

**HEM: 6.8 IPs per HEM**  
on average (median = 3)

**Cookie: 2.0 IPs per cookie**  
on average (median = 1)

**MAID: 2.1 IPs per MAID**  
on average (median = 1)

*Lookback of 90 days (Nov 17, 2021 to Feb 15, 2022)*

One to many - It's important to be able to connect IDs to one another in order to create an accurate ID graph that can connect different channels, connect individuals to a home, and increase scale.



DEEP  
DIVE

## ID 3: IPs

# The Household Connection

## Pros:



- Passive / discrete collection similar to MAIDs and third party cookies
- Can be tied to a location that is possibly (but not definitely) a home, making it especially valuable for CTV
- Can link individuals/identifiers together that are in the same location/home – again good for householding/CTV
- More “bang for your buck” - can use data everywhere.
  - Can be used on a cookieless browser (Firefox, Safari)
  - Works across screens / environments (App, Browser, CTV), relevant for cross device targeting

## Cons:



- Less persistent than many realize
- On it's own difficult to use a home IP vs a non-home IP. Other signals are needed to determine whether an IP is a home
- For individual-level targeting, IP would be used more probabilistically rather than deterministically since it's not tied to an individual
- Platforms like Apple looking to obfuscate IP in the future
- Not consented like email



KEY FINDINGS

# Conclusion

## IDs are Dynamic, Require Expertise

### The lifespan of some IDs is shorter than many people may expect,

and in some cases, getting shorter. It's also possible to have many IDs for an individual or household that need to be connected accurately - and little infrastructure exists for brands and publishers to share directly without major loss or mismatching.

It's clear that there needs to be technology in place to aid with ID collection and matching to reduce inaccuracies and waste. It's important to invest in smart ID management, not just data collection.



**Invest more in email**  
due to **consent and persistence**



**Invest less in MAIDs**  
due to **deprecation and less reliability**



**Work with a partner that understands IP addresses**  
They're very valuable if you can differentiate home vs. public

# Methodology

MediaWallah's data team analyzed hundreds of millions of US-based HEM, mobile ad ID, and IP data collected between 2020 and 2021.

- US only
- Filtered out invalid/suspicious data
- Removed partners with insignificant volume



## HEMs

- Filtered out outliers that potentially represented overconnected, fraudulent, or otherwise inaccurate data.
- Also filtered out mismatched HEMs where appropriate. "Mismatched" refers to an MD5 with more than one SHA1 or SHA256 tied to it



## MAIDs

- Took MAIDs that appeared in our network in 2020 and 2021
- MAIDs from United States
- Filtered out outliers that did not reach a critical threshold of connected HEMs or MAIDs, and may thus represent overconnected, fraudulent, or otherwise inaccurate data.
- MAID had to be labeled as either IDFA or AAID
- After identifying the target MAIDs, calculated the shelf life by seeing the difference between the first and last timestamp of their appearance in our network



## IPs

- Took IPs and the ZIP code they were associated within our network from 2020 to 2021
- Filtered to IP addresses in the United States
- Calculated the shelf life for each IP-ZIP combination by looking at regional ZIP code changes
- Used Internet connection type and cookie-MAID-HEM connections to determine household IP

# About MediaWallah

MediaWallah provides innovative identity-by-design solutions for the world's largest brands, publishers, platforms and agencies. Mediawallah helps companies create complex 360-degree views of customers and prospects to continue to market to them across channels, IDs, data types, partners, and their own business silos—all with the most advanced security protections, and regardless of changes to the market landscape.

Companies trust MediaWallah to provide agile privacy-first data capabilities to ensure that their identity data is kept safe while also delivering maximum value even as market dynamics evolve over time.

Learn more at [www.MediaWallah.com](http://www.MediaWallah.com)

