IDENTITY RESOLUTION FOR ADVANCED TV AND VIDEO ADVERTISING:

A Case for Durability and Transparency

October 2024

CONTENTS

About CIMM

The Coalition for Innovative Media Measurement (CIMM) is a non-partisan, pan-industry coalition of companies from across the media and advertising ecosystem, focused on supporting improvements, best practices and innovations in measurement and currency development, the use and application of new metrics and approaches to understanding the value of media, and data collaboration and enablement.

Every year, CIMM engages with its members to identify the most important and pressing priorities facing the industry. This study, a strategic review of the identity resolution (IDR) ecosystem for advanced TV advertising, was identified as a critical priority for our members in 2023, as identity has assumed a central, vital role in the industry.

Identity resolution is a longstanding, proven capability in the media and advertising industries, but has become far more important in recent users as a foundation for a wide range of critical industry use cases, including measurement and currency, planning, targeting, addressability, and attribution.

Many aspects of TV's fast-developing identity ecosystem(s) are performing well, supporting a growing volume of transactions between publishers, platforms and agencies. Clean room infrastructure has developed rapidly, supporting this growth. There are also important trends and initiatives underway across the market that have the potential to increase the importance of identity, with some of CIMM's members arguing that the future of the TV industry is being built on identity.

However, TV's current identity ecosystem is also complex and fragmented, with very different ecosystems across different platforms and pools of TV and video inventory, a diverse range of identity solutions, and, in at least some cases, limited transparency about interdependencies and data quality. Identity spines and solutions have proliferated, leveraging different audience relationships, identifiers, signals and attributes, resulting in a fast-developing marketplace with a wide range of solutions and applications. Transacting across different identity spines, graphs and solutions and translating between people, devices and households can be complex for buyers. Shifting privacy regulations, changes to the TV distribution landscape and to consumer behaviors, and innovative new approaches to data collaboration and sharing are driving further changes. Many of CIMM's members have expressed concerns about transparency, efficiency, control and governance, interoperability, and, perhaps inevitably, costs.

As such, we're delighted to present this diagnostic study to the industry, in partnership with OpenAP, the first step in a program of work designed to deliver real improvements and to support the effective operation of this new ecosystem. We look forward to further collaboration with companies across the ecosystem.

We'd like to thank the many senior executives and expert advisors who contributed to this study and the team at ThinkMedium, for the expert analysis and diligent work, managing a diverse range of demanding stakeholders.

Jon Watts,
Managing Director,
CIMM (October 2024)

About OpenAP

All Call for Industry-Wide Evolution in Identity Resolution for Streaming

OpenAP was founded with the belief that for data-driven targeting to scale in TV advertising, we – as an industry – needed to make it easier for advertisers to find, reach and measure advanced audiences wherever they were viewing content across multiple publishers. This belief led to coalescing the industry around common standards for advanced audiences, a commitment to interoperability and a level of collaboration that has helped grow the overall market for advanced TV. A once nascent channel is now a multi-billion-dollar marketplace.

Fast forward to the present day and we are, again, at a consequential inflection point that will determine the sustainability of data-driven video with the advent of streaming.

Campaign measurement is becoming more reliant on identifying digital signal versus modeling. As viewership continues to migrate to streaming platforms, the percentage of impressions with signal will continue to increase, reducing the need for complex and expensive modeling as a requirement for reach and frequency. While walled gardens have leveraged signal data to make audience advertising more actionable and measurable for years, television now needs to catch up.

As this transition takes hold, identity will be the most important part of the data-driven video advertising workflow, defining everything from how we build audiences to how we understand which screens ads were delivered to. And as with any aspect of workflow, the process needs to be as standardized, transparent and as trusted as possible so the market can transact at scale – and with confidence.

To help shepherd the industry through the next wave of transformation, it was critical that we at OpenAP, alongside CIMM and the cross-industry group of stakeholders who contributed to this report, illuminate and bring awareness to the next set of issues we together must solve – all centered around how to transparently, consistently and accurately map signal data to a household or person.

As the playbook takes shape for the next three-to-five years for the TV advertising industry, our hope is that this report lays the groundwork for why the industry needs to invest now in clear and transparent guidelines for how to consistently map identifiers to a household for the purposes of both audience creation and measurement. With consistency, we can improve accuracy and – when working together – we can simplify a complex and bifurcated ecosystem, making it easier for advertisers to leverage this medium and help drive growth and scale to streaming video advertising.

Thanks to CIMM, ThinkMedium and our partners who collaborated to provide this important perspective of the current landscape, spanning Experian, LiveRamp, TransUnion, Hearst, Adstra Data, NBCUniversal, BlockGraph, Fox, Truthset, MRI Simmons, DirecTV, Verisk, Cadent and Publicis Groupe's Epsilon for contributing to this critical conversation.

**David Levy,
CEO of OpenAP
(October 2024)**

Authors & Contributors

ThinkMedium

Dennis Buchheim, CEO

Rachel Galvin, Advisor

Alex Roucourt, Advisor

Partners

Kyle Antoian, Founder, *Zeke Strategy*

Claudio Marcus, Founder, *inceptionAS*

CIMM & OpenAP

Ed Davis, President, Product & Operations, *OpenAP*

Jess Hindlian, VP of Ecosystem Solutions and Enablement, *OpenAP*

Brittany Slattery, Chief Marketing Officer, *OpenAP*

David Levy, Chief Executive Officer, *OpenAP*

Jon Watts, Managing Director, *CIMM*

Steering Group Members

Chris Feo, SVP Sales & Partnerships, *Experian*

Tara Franceschini, Head of Industry Strategy, *LiveRamp*

Kimberly Gilberti, Chief Product Officer, *Experian*

Kevin Heindl, Director, Strategic Planning, Media and Entertainment, *TransUnion*

Jessica Hogue, Chief Data Officer, *Hearst*

Emily Kistner, Director, New Business & Identity Solutions, *Adstra Data*

Kevin Kowalick, Vice President Strategy, Media, & Entertainment, *TransUnion*

John Lee, Chief Data Officer, *NBCUniversal*

Jason Manningham, Chief Executive Officer, *Blockgraph*

Mainak Mazumdar, Chief Research and Analytics Officer, *Fox*

Scott McKinley, Chief Executive Officer, *TruthSet*

Joshua Pisano, Senior Vice President, Product Strategy and Innovation, *MRI Simmons*

Peter Randazzo, Senior Vice President, Product, *NBCUniversal*

Dan Rosenfeld, Senior Vice President, Data Analytics and Insights, *DirecTV*

Zora Senat, Chief Commercial Officer, *Verisk*

Darren Sherriff, Vice President, Advertising Technology Solutions, *Fox*

Kate Sirkin, Executive Vice President, Global Data Partnerships, *Publicis Epsilon*

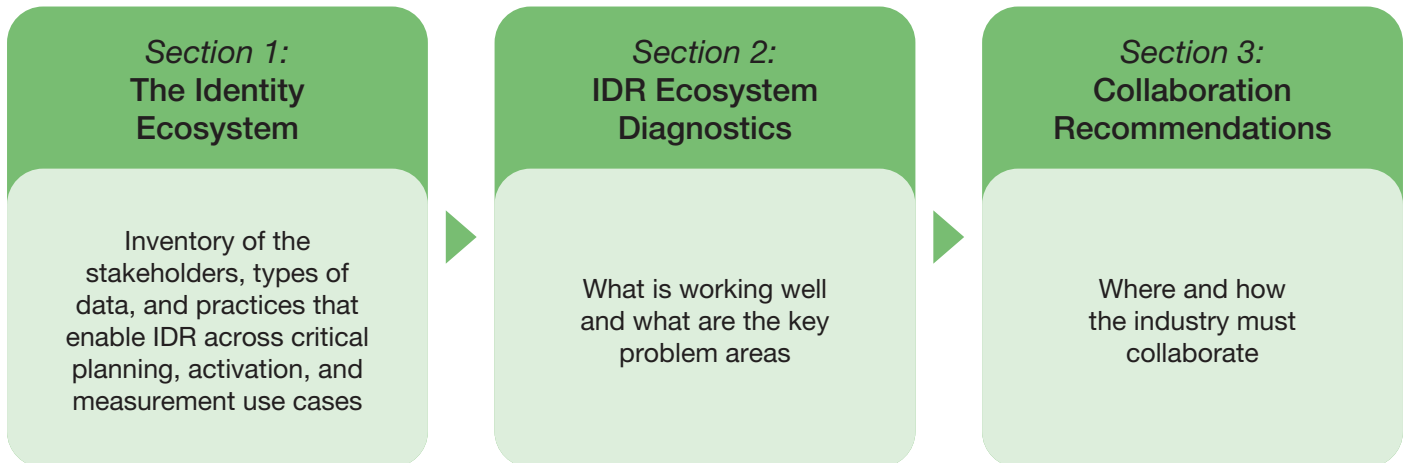Wendy White, Global Lead Product Marketing Manager, Identity, *LiveRamp*

Tony Yi, Executive Vice President, Business Development, *Cadent*

With thanks to additional interviewees from *Nielsen, GoAddressable,* and *Involved Media.*

Introduction

This report is a foundational resource for aligning, guiding, and improving on identity resolution (IDR) in Advanced TV and Video Advertising (ATVA). It provides: (1) an overview of the IDR ecosystem, (2) a diagnostic assessment of the current marketplace, and (3) recommendations for strengthening and improving current ways of working through industry collaboration.

The report was developed collaboratively, with extensive input from an advisory group of CIMM members, OpenAP, and consultation with industry participants across the US media and advertising ecosystem, undertaken primarily during Q2 and Q3 2024.

Section 1: **The Identity Ecosystem**	*Section 2:* **IDR Ecosystem Diagnostics**	*Section 3:* **Collaboration Recommendations**
Inventory of the stakeholders, types of data, and practices that enable IDR across critical planning, activation, and measurement use cases	What is working well and what are the key problem areas	Where and how the industry must collaborate

Executive Summary

Identity and IDR serve as the backbone for planning, activating, and measuring video advertising that is relevant, personalized, and effective. However, the rise of digital streaming, growing privacy concerns, and other trends are challenging traditional approaches in the TV advertising landscape. The increasing complexity and fragmentation of the ecosystem necessitate a reevaluation of strategies and industry collaboration to build a more effective, efficient, and durable future for identity and IDR in the Advanced TV and Video Advertising marketplace.

Although the market is complex, competitive, and developing rapidly, we believe there are important opportunities for improvement, consider:

1) *(Section 1: Identity Ecosystem)* **An advertising use case should determine where and how IDR is applied, as "one size does not fit all".** Identity resolution connects disparate identifiers (and often attributes) to create a unified view of a consumer (via an identity graph) for many planning, activation, and measurement use cases. It is critical to assess IDR in the context of a use case, as that determines the relevant ecosystem stakeholders, data needed, and processes and solutions leveraged.

2) *(Section 1: Identity Ecosystem)* **Data in the ATVA ecosystem and identity graphs can provide useful, but variable, approximations of identity for advertising purposes. However, it is not currently possible to have a fully comprehensive view (of all identifiers and attributes) for any consumer or household.** Data is sourced across diverse media and engagement channels, stakeholders, and devices with varying degrees of data availability, usage, and ID reliability. There may be richness in some elements of an identity graph, but gaps, partial unknowns, inaccuracies, or unmatched data in others. Data sufficiency is determined in the context of a particular use case.

3) *(Section 2: Diagnostics)* **The most significant problem in the current IDR ecosystem, especially among buyers, is a lack of transparency and consistency across the definitions and methodologies related to identity data.** For example, there is a lack of transparency about how a "match" is defined, when vendors report on a match rate, and what principles guide data resolution for audience augmentation.

4) *(Section 2: Diagnostics)* **Resolving these issues is important to addressing market-wide challenges in data quality, durability, interoperability, and usability. However, achieving industry-wide solutions is challenged by rapid growth and development of the IDR ecosystem, fragmentation and complexity, and misaligned incentives across stakeholders.** There is also no industry-wide coordination or central governing/oversight body for this specific space.

5) *(Section 3: Recommendations)* **The most important step that the industry can take to improve the IDR ecosystem is to increase transparency for identity-related data and practices,** especially related to data sourcing, matching, processing, and validation. To enable industry collaboration, the sell-side needs buy-side motivation to overcome competitive concerns, which requires addressing the buy-side's inconsistent awareness, understanding, and urgency. A two-pronged collaborative approach to developing industry resources could help to achieve this:

   a) *Develop buyer education and guides* to facilitate understanding of ecosystem fundamentals and provide curated questions for engaging with identity providers (in part to support the RFP/RFI process).

   b) *Define taxonomies and templates for providers to communicate about definitions and methodologies* for how data is sourced, matched, aggregated (into households, for example), and validated. Simple, flexible templates (updated periodically as practices change) that buy-side stakeholders can include as a part of an RFP/RFI process can motivate vendors and sellers to be more transparent.

Streamlining transparency and standardizing communication among buyers, sellers, and ID providers can establish a solid foundation upon which the industry can increase efficiency and effectiveness to address ongoing and new IDR-related challenges within ATVA (e.g., with tooling, automation, etc).

Identity resolution is critical to the future of TV and video advertising, but, in a privacy-centric market, a new level of collaboration is required to protect ATVA investments, unlock opportunities, and support the industry outside of walled gardens (which are advantaged by a logged-in user base and centralized decisioning systems). Growth, innovation, and shifts from both outside and inside the ATVA ecosystem will continue to shape IDR. To build a robust foundation capable of adapting to future obstacles, collaborative action – most notably related to transparency (as outlined above) – is needed now.

Context

Motivations

Employing *identity resolution (IDR)* to connect data across media channels and sources in order to recognize and understand specific consumers (each having an *identity*) has been fundamental to value creation in advertising. Within *television advertising,*[1] two parallel but historically distinct pathways have emerged – *linear* and *digital* – each relying on different strategies, metrics, and mechanisms for identifying (via *identifiers or IDs*) and engaging with consumers. As digital streaming devices and services proliferate and adoption by consumers rises, a holistic strategy is needed for *advanced TV and video advertising (ATVA)*, which also requires reevaluating existing approaches to identity and IDR
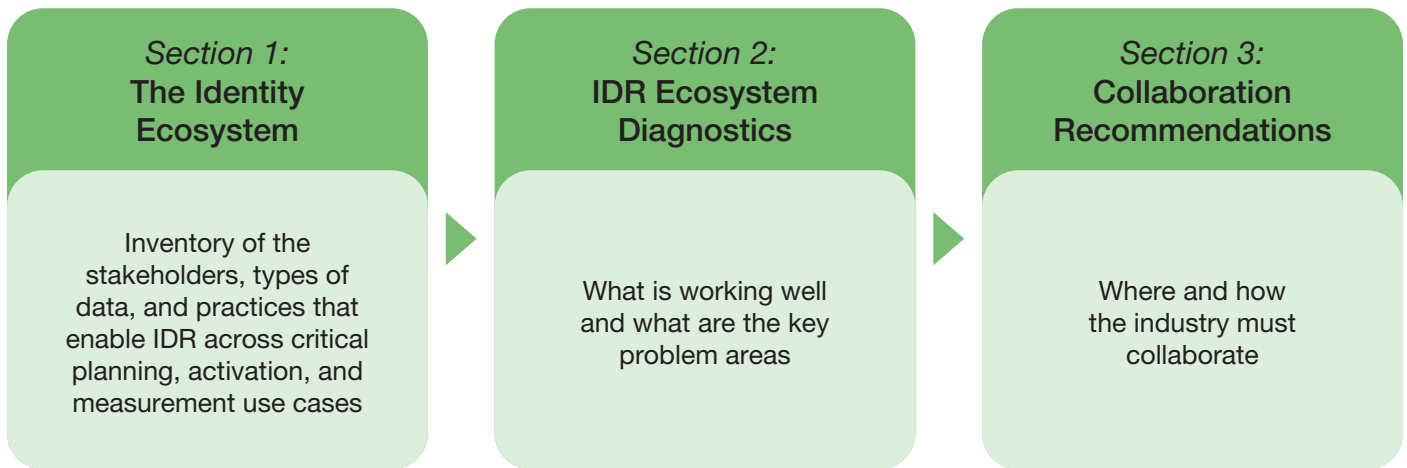
Broader consumer, business, and regulatory trends – especially those related to privacy – further complicate the ability to unify and simplify identity and IDR within the ATVA ecosystem. Consumers' growing awareness of data usage and expectations for privacy in advertising have motivated the creation of an expanding patchwork of piecemeal and difficult-to-understand regulatory and platform requirements. Likewise, consumer usage of tools for blocking ads, obfuscating identity (e.g., one-time-use emails), and limiting data collection, storage, and sharing (e.g., opt-outs, data deletion requests) continues to grow. These trends have collectively led to the phasing out of many identifiers (e.g., cookies, mobile ad IDs), instability among some remaining identifiers, reduced data availability, and limited data utility (i.e., performance, coverage, and accuracy).

Within ATVA specifically, the increasing number of stakeholders and ID solutions adds complexity to data flows, ID usage, and IDR needs. It also adds more players who may have their own interpretations or expectations in regard to privacy policies.

TV advertising and the identity ecosystem that powers it are at a crossroads, facing significant changes and challenges on top of existing complexity and fragmentation. It is crucial to reassess approaches to identity and identity resolution in advanced TV and video advertising to build a more effective, efficient, and durable ecosystem.

---

1   Television is premium, professionally produced long-form video. Linear TV refers to TV accessed on traditional devices through broadcasters, typically on set schedules, such as over the air and MVPDs. Digital TV encompasses all devices and streaming services for accessing TV through the internet – such as connected TVs (CTVs), over-the-top (OTT) devices, and streaming services. (See Privacy and the Future of TV Advertising report.)

Study Objectives

| Section 1: The Identity Ecosystem	Section 2: IDR Ecosystem Diagnostics	Section 3: Collaboration Recommendations
Inventory of the stakeholders, types of data, and practices that enable IDR across critical planning, activation, and measurement use cases	What is working well and what are the key problem areas	Where and how the industry must collaborate

This report is intended to serve as a foundational resource for aligning, guiding, and improving on identity resolution in advanced TV and video advertising (ATVA). The report is divided into three complementary sections providing:

1) *(Section 1)* **An overview of the current IDR ecosystem**, including key use cases, stakeholders, and practices, and processes for working together

2) *(Section 2)* **A diagnostic assessment of what is working well and what are key problems**

3) *(Section 3)* **Opportunities and recommendations** for improving the ecosystem through industry collaboration

The report itself was developed collaboratively, with extensive input from an advisory group of CIMM members, OpenAP, ThinkMedium, and extensive consultation with industry participants from across the US media and advertising ecosystem, undertaken primarily during Q2 and Q3 2024 (*see Appendix for the full project steering group*).

Working sessions, consultation with the industry, input from experts, and secondary research were used to develop and calibrate the frameworks, assessments, and recommendations. Perspectives have been aggregated across stakeholder groups for a holistic ecosystem view; specifics may vary for each stakeholder audience. As inputs were collected in the US, the recommendations provided are focused on the US as well, but the underlying frameworks could be applied to other markets.
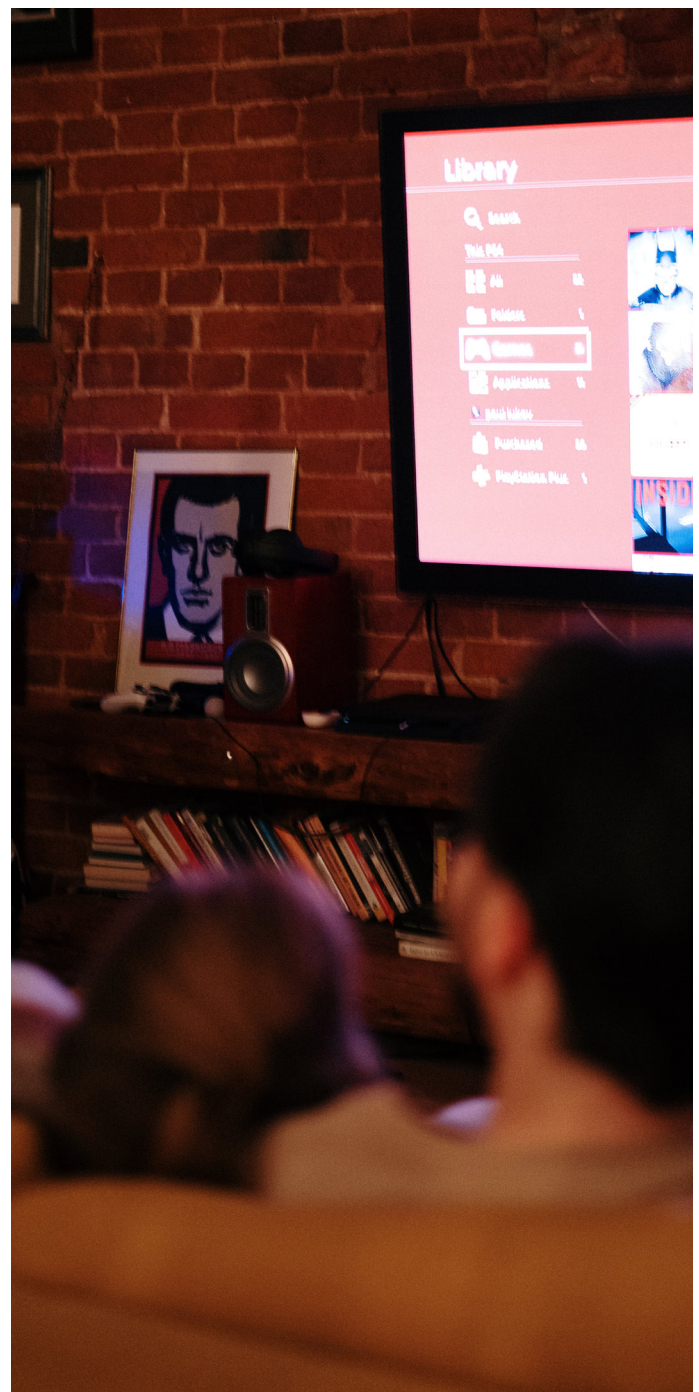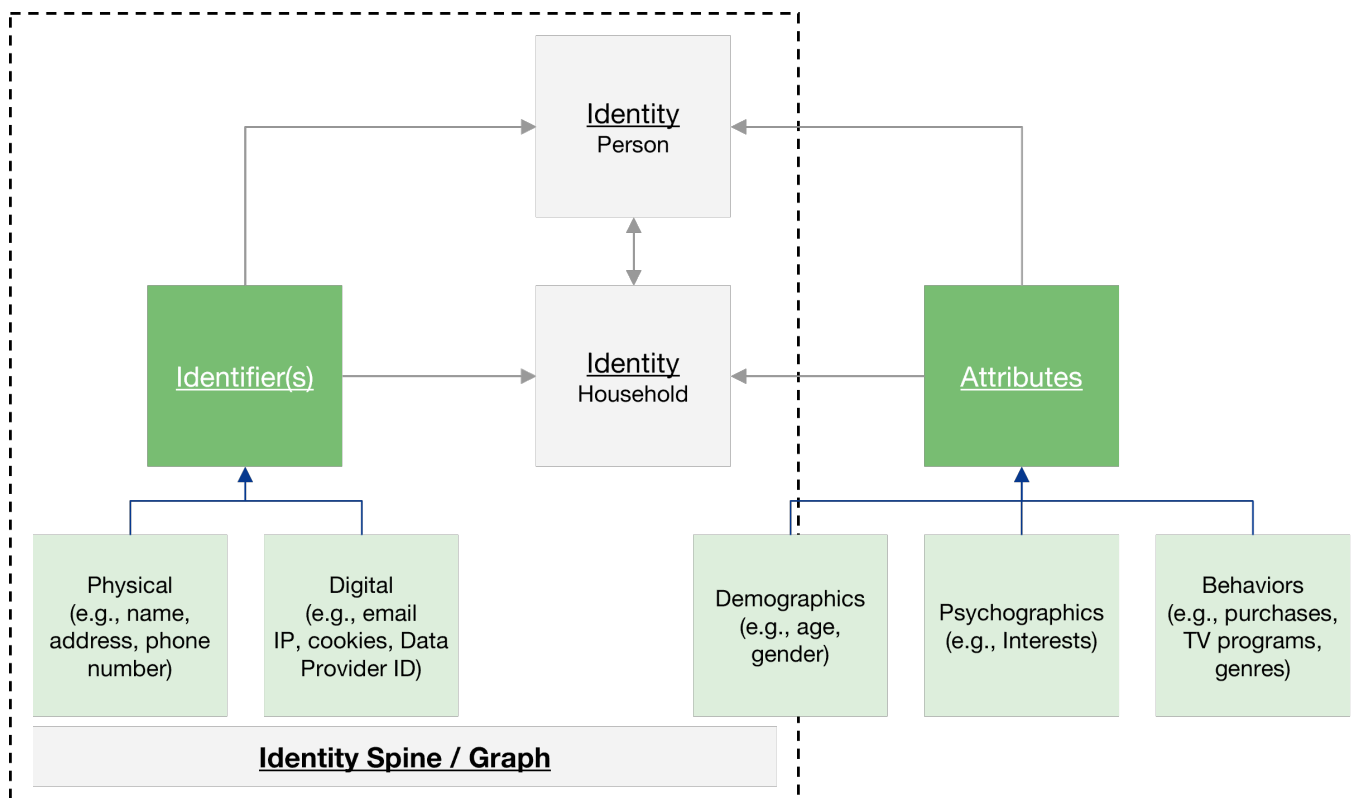
Terminology

Figure: Hierarchy of identity-related terminology

Before diving into the complexities of IDR and how to improve the ecosystem, it is important to level-set on the fundamental elements of identity resolution. Underlying the ecosystem is the concept of an *identity graph* (or spine) which connects a view of a unique individual- and/or household *identity* by linking data across sources and channels. Identity graphs are composed of *identifiers* (keys or partial keys) that can be matched to 1st-, 2nd-, or 3rd-party data that represent *attributes* (e.g., demographic, behavioral, and psychographic data) *associated* with an *identity*.[2] Identity graphs are fundamentally used to make sense of incoming and outgoing consumer data as they apply to specific advertising use cases (e.g., targeting ads or measuring campaign effectiveness).

While the concept is straightforward in theory, in practice it is extremely complex in ATVA for several reasons. First, identifiers and identity graphs provide a useful approximation of identity for advertising purposes, but are far from perfect, one-to-one representations of identity. Second, identity-related data is sourced across diverse media channels and stakeholders, each with varying degrees of data availability and ID reliability. Third, identity in ATVA can take multiple forms, with some having the ability to reconcile consumer/devices, and others at the IP address or household level. **This means it is not possible to have a fully comprehensive view (of all the characteristics, behaviors, and preferences, or even all the associated identifiers) of a consumer or household. Instead, identity-related data exists on a spectrum, with richness in some elements of a graph and gaps, inaccuracies, or unmatched data in others.**

---

2   The term *signals* is used inconsistently across the ecosystem, ranging from components of identifiers (e.g., IP addresses) to data points that assess identity strength to attributes themselves. For clarity, we refrain from using the term in this report.

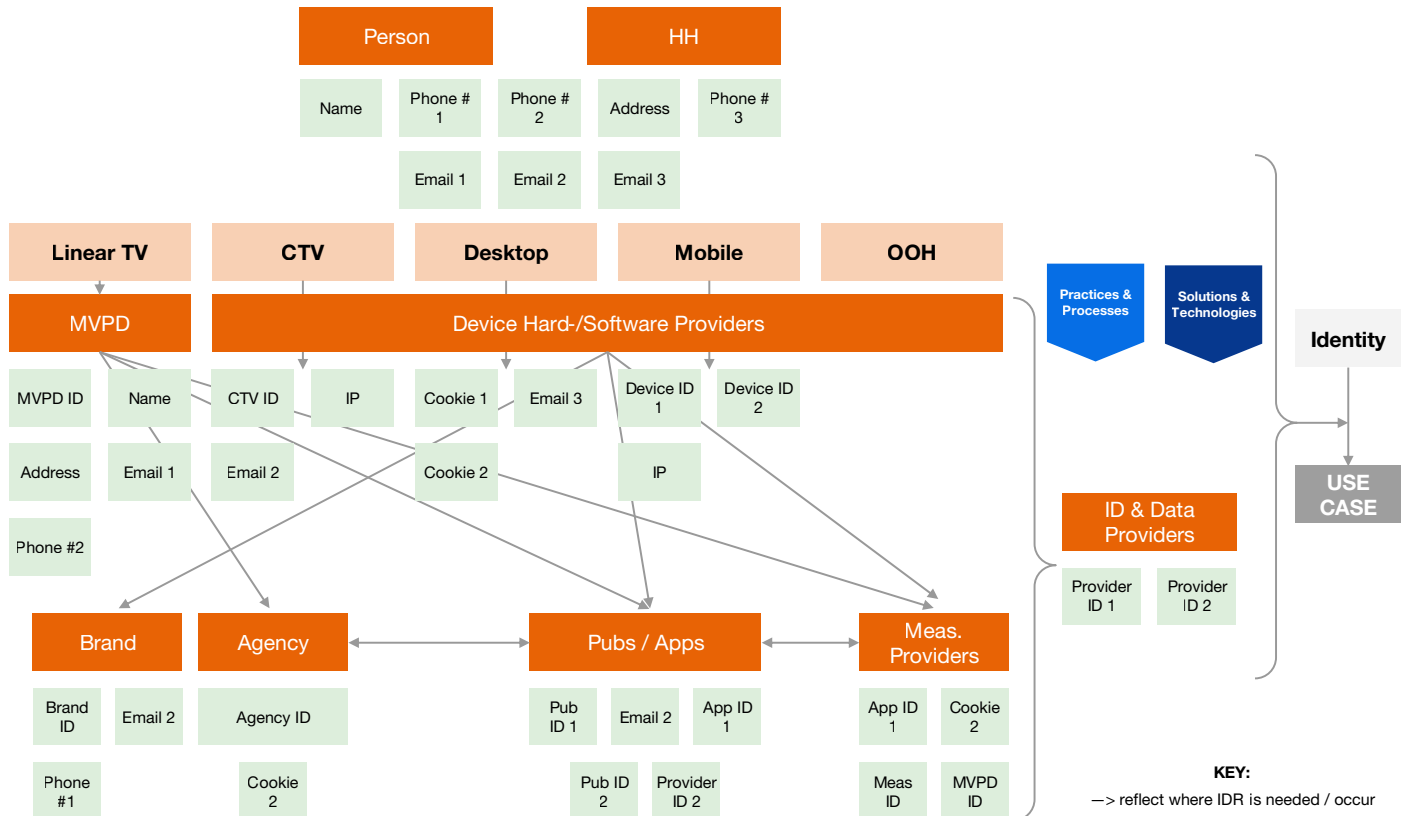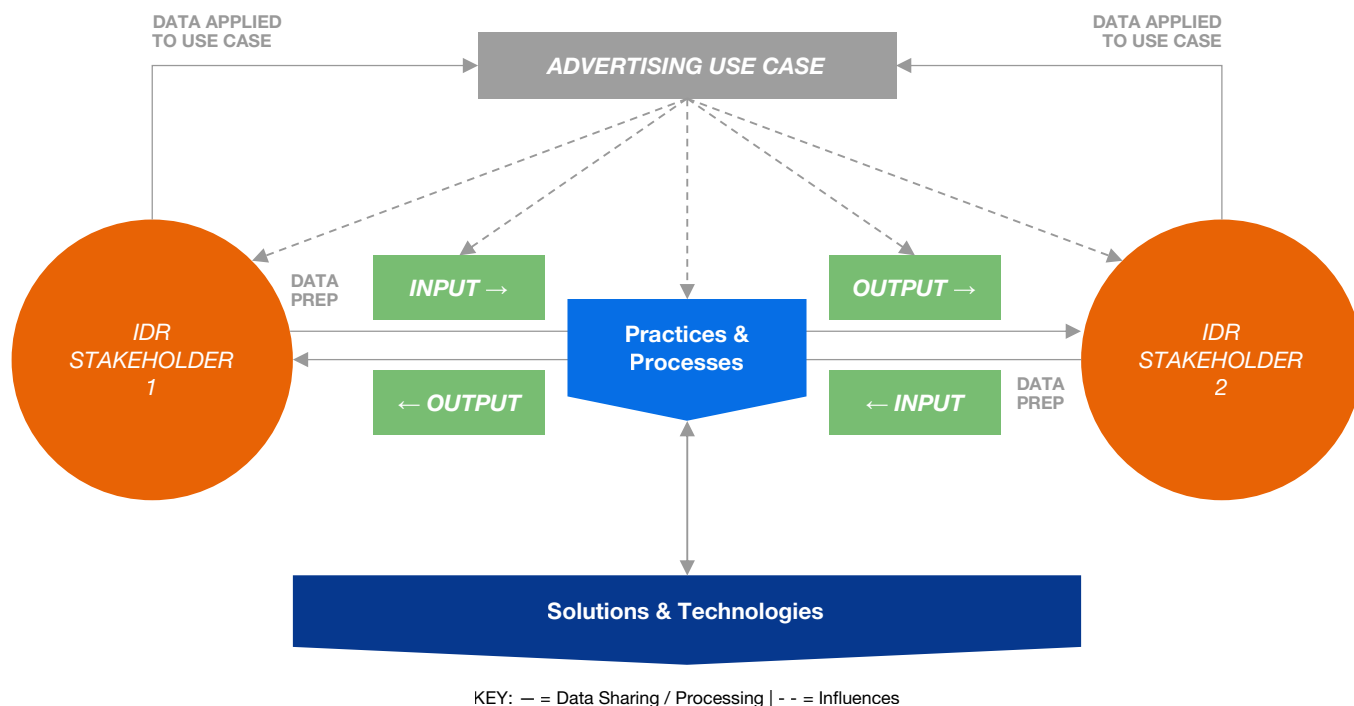# Each Consumer Has a Complex Identity

Figure: Example of the diversity of media channels and stakeholders potentially involved in sourcing identifiers about a single consumer (and their household) for use in the ATVA ecosystem, underscoring challenges in building scaled identity graphs across media channels.

The Identity Ecosystem

Overview

DATA APPLIED TO USE CASE

DATA APPLIED TO USE CASE

ADVERTISING USE CASE

IDR STAKEHOLDER 1

DATA PREP

INPUT →

Practices & Processes

OUTPUT →

IDR STAKEHOLDER 2

← OUTPUT

← INPUT

DATA PREP

Solutions & Technologies

KEY: — = Data Sharing / Processing | - - = Influences

Understanding, navigating, and improving upon the current IDR ecosystem begins with breaking down its parts and recognizing that IDR is undertaken to serve specific advertising use cases:

- A *use case* (Planning, Activation, Measurement at the top level) determines which IDR *stakeholders* engage and what type of identity-related data is needed (*inputs/outputs*).

- IDR *practices and processes* are employed, enabled by *solutions and technologies*, to reconcile one stakeholder's identity-related *input* data (such as IDs or attributes) with another's, and facilitate the sharing of *output* data for application to a *use case*.

A use case influences the inputs, outputs, and applicable IDR practices and processes employed.

IDR Use Cases

Key Takeaways

- IDR is foundational to a variety of the planning, activation, and measurement use cases that create value in ATVA.

- The extent to which a use case relies on IDR depends upon how it is executed and which media channels are included, particularly if collecting or sharing identity-level data is necessary.

- While specifics differ by media channel, stakeholder, and methodology, some priority IDR use cases include:

  ○ Within planning and activation: targeting, ad delivery and serving, optimization

  ○ For measurement: performance/impact assessments, audience validation, and (increasingly) audience counting

Theme	Advertising Use Cases	Implementation(s) Reliant on IDR Today?	
		Yes	No
Planning & Activation (For developing and implementing ATVA)	**Targeting**	✓ For Existing Customers, Prospecting, Reach Extension (using 1st- and/or 3rd-party data)	✓ For Contextual, Audience Modeling
	Suitability	-	✓
	Campaign Activation	-	✓
	Ad Delivery / Serving	✓	-
	Optimization	✓ (Including frequency capping)	-
Measurement (Metrics and currencies used to assess ATVA efficiency and effectiveness)	**(Baseline) Audience Counting**	✓ For Deduplicated Reach, Frequency, GRPs	✓ For Reach
	Protection / Verification	✓ For some Fraud/Security	✓ For Brand Safety, some Fraud/Security
	Audience Validation	✓ For In-target	✓ For Viewability, Attention
	Performance / Impact	✓ For Conversion/ Attribution, Brand Lift	✓ For MMM, onsite engagement metrics

Identity-related data is used to enable a variety of use cases (needs, functions, or capabilities) that create value within ATVA. Identity and IDR are only needed when a use case requires collecting and connecting identity-level (not aggregated) data. This is determined by the methodology for implementing the use case and/or if data must be connected across multiple channels or sources.[3] For example, simple reach calculations on a single channel may not require IDR, but deduplicated reach and frequency within- or across- media channels does.

Ultimately, because of differences in stakeholder needs and methodological preferences – which can even differ by media channel where a use case is implemented, or by campaign goals – there are no definitive rules for which advertising use cases *always* require IDR or which use cases are *most critical* across all stakeholders. However, some use cases are *generally*[4] more dependent on IDR and are a helpful place to start in evaluating current practices and opportunities for industry collaboration:

Within **planning and activation** (for developing and implementing ATVA):

- *Use cases more reliant on IDR:*

   ○ *Targeting:* Delivering personalized ads is often achieved via an understanding of a consumer on a one-to-one basis across channels, devices, and platforms. (While there are alternatives such as contextual targeting and audience modeling, identity-dependent methods have been critical to efficient advertising to date.)

   ○ *Ad Delivery and Serving:* Showing the "right ad to the right person at the right time" relies heavily on the ability to identify consumers.

   ○ *Optimization:* Improving performance based on more holistic consumer behavior and engagement is facilitated by IDR. Likewise, frequency capping at the individual, household, and device-level requires IDR.

- *Use cases less reliant on IDR:*

   ○ *Brand Safety/Suitability:* While understanding one's audience may help determine suitability for an ad/advertiser, granular identity-level data is typically not required to address suitability constraints, and contextual clues are often sufficient.

   ○ *Campaign activation:* This traditionally entails the logistical aspects of campaign planning – scheduling, placement, budget allocation, etc. – where granular identity-level data is not typically required.

Within **measurement** (metrics and currencies for assessing ATVA):

- *Use cases more reliant on IDR:*

   ○ *Performance / Impact:* Measuring the impact of advertising campaigns, for example with attribution and lift, relies on linking ad exposure and outcomes across consumer touchpoints, often enabled by IDR.

   ○ *Audience Validation (In-target):* Ensuring the intended audience was reached requires knowing details about the consumers exposed to a campaign. However, other validation use cases – such as viewability and attention – are not necessarily identity-dependent.[5]

   ○ *Audience Counting:* Historically, basic audience reporting was achieved without identifying individuals, but more advanced metrics – such as deduplicated reach and frequency – required identity resolution. Today, all types of audience counting (and currencies) increasingly rely on identity.

- *Use case(s) less reliant on IDR:*

   ○ *Protection / Verification:* With some exceptions, preventing ad fraud and ensuring ads are displayed in brand-safe environments is not as sensitive to identifying consumers.

---

3   Note: When the lifecycle of a campaign involves combining multiple use cases that leverage different partners for activation and measurement, the complexity, needs, and challenges can increase significantly.

4   Aggregated view across stakeholders, specific stakeholders may prioritize different use cases or have different reasoning for prioritizing a given use case.

5   Generally, viewability and attention measurement for the traditional purpose of post-bid analysis is performed at the impression-level, with no need to layer on unique identities to deliver utility. However, performing pre-bid optimizations based on viewability/attention signals may require identity resolution at the consumer-level.

IDR Stakeholders

The identity resolution ecosystem that enables advanced TV and video advertising is a complex network of stakeholders that collect, match, and/or share identifiers and related attribute data. Stakeholders can be classified based on the primary functions they serve in the ecosystem, the types and level of data available to them, and how data is typically sourced. Even within each stakeholder group, the specific identifiers, practices, and processes used for identity resolution vary. *See the "CIMMScape" for a high-level overview.*

The following sections dive deeper into the specifics of identifier types and how this may map to various stakeholder groups.

Key Takeaways

- IDR ecosystem stakeholders are the entities that may collect, provide, or access identity-related data (e.g., IDs and attributes).

- They can be classified based on their role in the ecosystem, which has implications on where their identity-related data may be sourced, which identifiers are used, and which IDR practices are employed (though this varies across/within stakeholder groups).

- The IDs each stakeholder uses can be a mix of both proprietary internal IDs and IDs that have broader external applicability. Data entering or leaving systems may require resolution across any of these types of identifiers.

CIMMscape

MARKETERS | **CONSUMERS**

Agencies
Type: Digital, Level: Individual, Authentication: Mix

Dentsu
Havas
IPG
Omnicom
Group
Publicis
WPP

MVPD
Type: Physical, Level: Household, Authentication: Yes

| Altice | Comcast Xfinity | DirectTV | DISH Network | Spectrum |

vMVPD
Type: Mix, Level: Household, Authentication: Mix

Fubo
Sling TV
YouTubeTV

CTV Manufacturers & Operating Systems
Type: Digital, Level: HH or Device, Authentication: No

Amazon Fire
Android
AppleTV
Google ChromeCast
Microsoft XBox
Roku
Samsung
LG / WebOS
Vizio

Content Producers / Programmers
Type: Digital, Level: Household, Authentication: Mix

| AppleTV+ NBCU | Disney Netflix Yahoo* | Fox Peacock |

Buy-side Platforms
Type: Digital, Level: Individual or Device, Authentication: No

Amazon DSP
Criteo
Google DV360
The Trade
Desk
Xandr
Yahoo DSP

Sell-Side Platforms
Type: Digital, Level: Individual or Device, Authentication: No

Google Ad Manager
Index Exchange
Magnite
OpenX
PubMatic
TripleLift

Personal Device Manufacturers / OSes
Type: Digital, Level: Individual or Device, Authentication: Mix

Apple: MacOS, iOS, Safari
Google: Android, Chrome
Microsoft: Windows, Edge

Measurement Vendors
Type: Digital, Level: Mix, Authentication: Mix

| Comscore | iSpot | Nielsen | VideoAmp |

ID Vendors and Platforms
Type: Mix, Level: Mix, Authentication: Mix

| Acxiom Adstra | Blockgraph* Criteo* | Epsilon Experian | LiveRamp* Lotame* | The Trade Desk* TransUnion | Truthset Verisk |

Clean Rooms
Type: N/A, Level: N/A, Authentication: N/A

| AWS Clean Rooms Google Ads Data Hub | Epsilon InfoSum | LiveRamp/Habu Snowflake |

Company names per stakeholder group are non-exhaustive.

KEY

Generalized identifier characteristics within a stakeholder group:
- **Type:** Physical, Digital
- **Level:** Person, Device, Household
- **Authentication:** Yes / No

* = Stakeholder offering a proprietary identifier to enable identification of consumers across ATVA

→ = Flow of identity-related data

Figure: "CIMMscape" for IDR Ecosystem

IDR Inputs / Outputs

Key Takeaways

- *Inputs and outputs* cover all consumer data necessary to enable a use case.

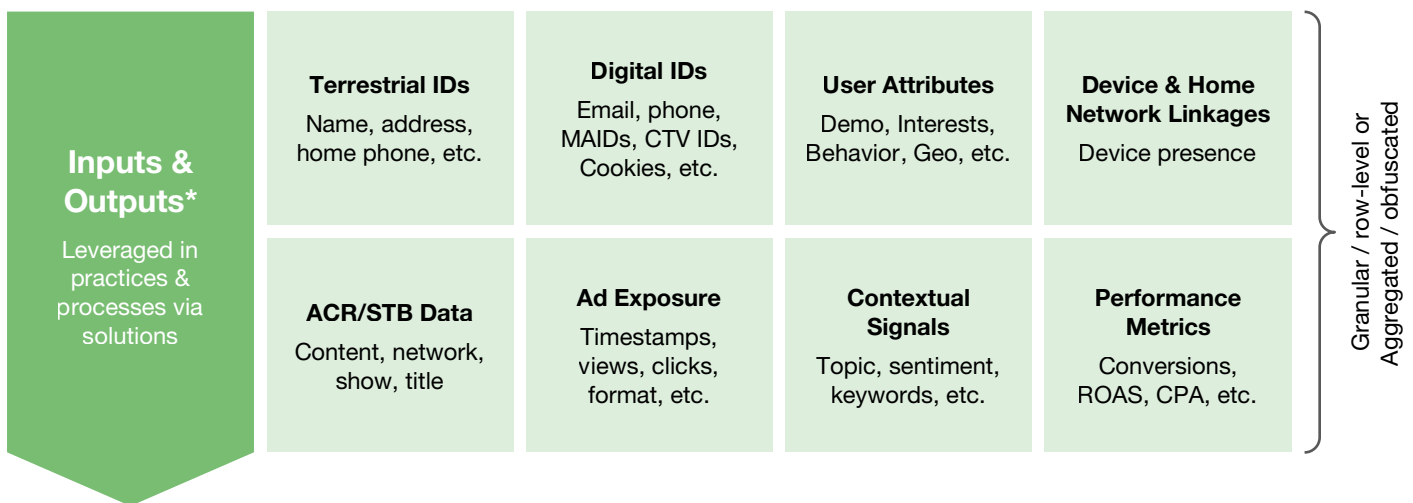- Since identifiers are fundamental to IDR, it is important to understand they vary by class (physical or digital), *level* (individual, household, or even device), and *authentication* (explicitly provided and verified by the consumer or not).

- Inputs and outputs are not equally available or standardized across media channels, devices, or stakeholders.

- This means for IDR purposes it is important to consider the nuances of each of these elements and the intermediaries data must flow through to enable a use case.

Inputs & Outputs* Leveraged in practices & processes via solutions	Terrestrial IDs Name, address, home phone, etc.	Digital IDs Email, phone, MAIDs, CTV IDs, Cookies, etc.	User Attributes Demo, Interests, Behavior, Geo, etc.	Device & Home Network Linkages Device presence
	ACR/STB Data Content, network, show, title	Ad Exposure Timestamps, views, clicks, format, etc.	Contextual Signals Topic, sentiment, keywords, etc.	Performance Metrics Conversions, ROAS, CPA, etc.

Granular / row-level or Aggregated / obfuscated

Inputs and *outputs* encompass all of the identifiers, attributes, and related data that must be resolved to enable a specific use case. While non-identifier data (such as ad exposures or purchase behavior) is needed to support ATVA use cases, identifiers are foundational for leveraging this and other data, so it is important to understand key characteristics in more detail:

1. **Identifier Classes:** Identifiers fall into two primary *classes:*

   - *Physical identifiers*, such as name, physical address, or telephone number.

   - *Digital identifiers*, such as IP addresses, emails, and app/device-specific IDs.

Within ATVA, physical IDs have traditionally been valued as a source of truth for IDR because of their stability (e.g., physical address doesn't change frequently), but accessibility by broad ecosystem stakeholders is limited. On the other hand, digital identifiers as a class tend to be more varied in nature (e.g., the origin and application of IP address versus consented Apple IDFA) which further contributes to the confusion about their provenance and ability to be utilized (or not) by the ecosystem.

2. **Identifier Levels:** Identifiers are typically collected, processed, and shared at the individual, household, and/or device- *level*. Today, how each of these levels is defined can vary by stakeholder.

3. **Identifier Collection/Authentication:** Identifiers can originate from consumers – for example, phone numbers or emails – or be created and applied by ecosystem stakeholders – for example, IPs, app-IDs, or "universal" IDs" (e.g., The Trade Desk's UID 2.0, LiveRamp's RampID). IDs are then *collected* or *assigned* to/from consumers and passed into the ATVA ecosystem:

   ○ Actively *(authenticated):* through login or registration, such as through cable network subscriptions or requiring sign-ins for access to media

   ○ Passively: through technical protocols and requests implemented on devices, browsers, and apps that do not require users to explicitly provide identifiers themselves, such as IP addresses or CTV IDs[6]

For any given use case, the characteristics of ID data available is determined by:

- Medial channel, device, and stakeholder through which consumer data enters the ATVA ecosystem

- Intermediaries through which ID data is shared (each of whom may incorporate their own technical and policy considerations)

As a consequence, even the "same" identifier collected or shared by two stakeholders can look very different – and the more stakeholders data passes through, the greater the risk (intentional or not) of reducing data fidelity.[7] Unfortunately, there is no universal guide for understanding any single identifier across all data sources and flows, nor are there one-size-fits-all rules for the types of data each type of stakeholder may be able to access. However, some general starting points for how this looks today are:

- *Linear TV* has long relied on physical IDs collected directly from consumers (e.g., MVPD subscriptions) and applied at the *household level*.

- *Digital (CTV, Desktop, and Mobile Streaming):* ID class, level, and collection method are more varied and determined by a combination of the device and how a given stakeholder makes content accessible to consumers. For example:

By device:	Is data collected from a TV or another shared device in the home?	Skews household, digital
	Is data from a personal device (e.g., laptop, phone)?	Skews individual, digital
By stakeholder and access requirement:	Login required (e.g., some smart TVs, subscription streaming)?	Authenticated
	Login not required (e.g., FAST without registration)?	Not authenticated

These complexities and inconsistencies can lead to significant gaps and omissions in identity graphs and mapping. For instance, a stakeholder may not be able to connect mobile data on a specific consumer to corresponding CTV data, simply because underlying identity mechanisms may not establish that the devices belong to the same individual. In this example, a mobile or app ID cannot be matched directly to a CTV IP address, but if the mobile or app ID also includes the related IP address then a match is possible – though the quality of the match remains dependent on the relative quality or freshness of the IP addresses being matched.

Characteristics of Common Identifiers

The following table builds on the described frameworks to outline some of the most common identifiers in the ATVA space and highlight key considerations.

---

6   Some of the passively collected identifiers may be associated with an authenticated identifier upstream – such as from credit card information collected during CTV registration – however, this is not always resolved or made available for advertising purposes.

7   More data/party detail in Appendix

Characteristics of Common Identifiers

The following table builds on the described frameworks to outline some of the most common identifiers in the ATVA space and highlight key considerations.

Identifier	Class	Level	Collection	Key Considerations
Full Name	Physical	Individual	Subscriber sign-up Credit bureaus Data brokers Public Records	• Relatively common but not always unique or consistent • Duplicate records, false identities, and homonyms require other data to reconcile • Not directly regulated, but combining with other personal information is scrutinized
Physical Address	Physical	Household	Subscriber sign-up Credit bureaus Data brokers Public Records	• High usability for household targeting but less precise for individual targeting • Subject to privacy regulations but highly stable and can be anonymized
Email Address	Digital	Individual or Household	Subscriber sign-up Credit bureaus	• Widely used, often linked to online activities • Range of reliability based on relative primacy of provided email address (e.g., primary, secondary, or junk email address)[8] • Regulated by governments and impacted by platform/OS-driven privacy obfuscation (e.g. Apple "Hide My Email") • Scrutinized by regulators even with hashing[9]
IP Address[10]	Digital	Household	ISPs Data brokers Public databases Web logs Ad networks Analytics platforms	• Often captured "in the wild" from internet activities and devices but more reliable when sourced from or validated by ISPs that generate the IP addresses • Useful for broad geo-resolution, but less precise for individual identity due to rotating and shared IPs • Increasingly considered personal information under regulations such as GDPR and CPRA, requiring careful handling and/or obfuscation • Increasingly obfuscated by operating systems, browsers, virtual private networks
Proprietary Device-Based IDs	Digital	Individual or Household	Device manufacturers OS providers Ad networks App developers CTV providers IoT platforms	• Limited to specific devices and ecosystems, some with large scale and impact within these proprietary platforms • High availability within specific ecosystems, but low-to-nonexistent interoperability • Subject to privacy regulations with access increasingly curbed by platforms

---

8   CIMM <> TruthSet (Oct 2023) Unlocking the Power of Accurate Data: Truthset-CIMM Study Reveals Opportunities and Challenges

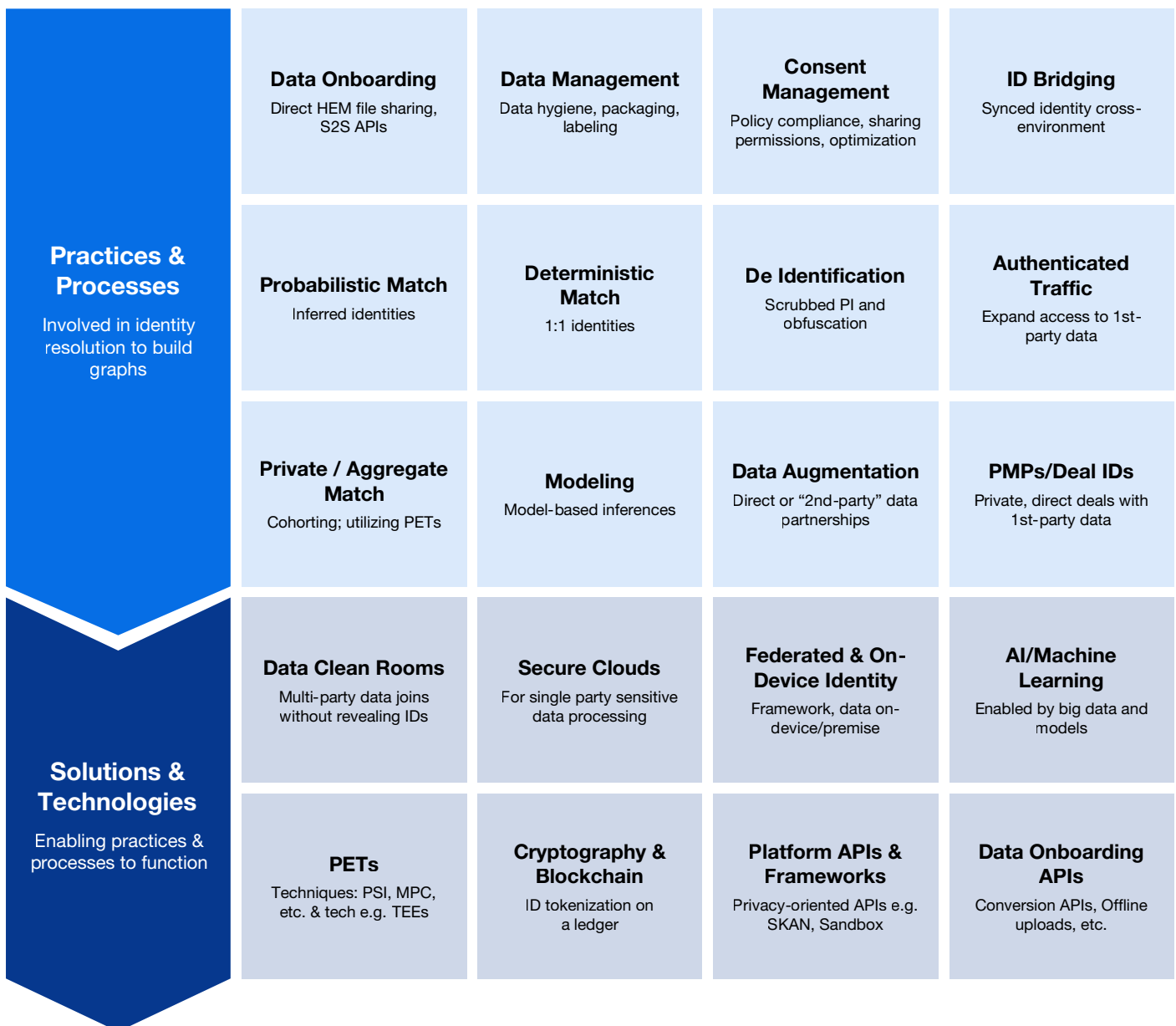9   No, hashing still doesn't make your data anonymous | Federal Trade Commission

10   IP addresses vary in quality & availability: IPv6 is more likely to represent a single device due to its larger address space, which allows for unique public addresses for each device. However, IPv6 has additional privacy features and is harder to parse as fragmentation is handled by originators vs routers for IPv4.

Practices & Processes + Solutions & Technologies

Key Takeaways

- The practices, solutions, and technologies for implementing IDR vary by stakeholder, which is why it is critical to understand any partner's approach to data collection, matching, scaling, and sharing.

- *Data Matching* is a key element of IDR, but underlying processes for achieving it – especially for probabilistic matching – often lack transparency and consistency.

- Privacy-forward *solutions and technologies* are evolving rapidly to offer new ways to ingest and process data more securely but can increase complexity and cost.

Practices & Processes — Involved in identity resolution to build graphs			
Data Onboarding — Direct HEM file sharing, S2S APIs	**Data Management** — Data hygiene, packaging, labeling	**Consent Management** — Policy compliance, sharing permissions, optimization	**ID Bridging** — Synced identity cross-environment
Probabilistic Match — Inferred identities	**Deterministic Match** — 1:1 identities	**De Identification** — Scrubbed PI and obfuscation	**Authenticated Traffic** — Expand access to 1st-party data
Private / Aggregate Match — Cohorting; utilizing PETs	**Modeling** — Model-based inferences	**Data Augmentation** — Direct or "2nd-party" data partnerships	**PMPs/Deal IDs** — Private, direct deals with 1st-party data

Solutions & Technologies — Enabling practices & processes to function			
Data Clean Rooms — Multi-party data joins without revealing IDs	**Secure Clouds** — For single party sensitive data processing	**Federated & On-Device Identity** — Framework, data on-device/premise	**AI/Machine Learning** — Enabled by big data and models
PETs — Techniques: PSI, MPC, etc. & tech e.g. TEEs	**Cryptography & Blockchain** — ID tokenization on a ledger	**Platform APIs & Frameworks** — Privacy-oriented APIs e.g. SKAN, Sandbox	**Data Onboarding APIs** — Conversion APIs, Offline uploads, etc.

Within IDR for ATVA, *practices and processes* are the approaches used to build and leverage identity graphs. *Solutions and technologies* are the underlying infrastructure and tools that facilitate the practices and processes.

Practices & Processes

Practices and processes serve various functions within IDR, such as:

- *Data collection and scaling:* Gathering the foundational layer of information about a consumer's identity and attributes (e.g., consent optimization, authenticated traffic) *(collection)* and increasing the scope and volume of data that can be used for ATVA use cases (e.g., data augmentation) *(scaling)*

- *Data matching:* The processes used to link identifiers and data points to a single individual or household (e.g., deterministic or probabilistic matching)

- *Data sharing:* Sourcing and/or disseminating inputs/outputs used for IDR (e.g., server-to-server sharing and direct file sharing)

Data Matching Processes

Data matching is essential to IDR and identity graph development. Two important elements to consider are the *technique* and *timing* for matching, which require tradeoffs of precision and scale for an intended use case.

Data matching is generally deterministic or probabilistic. *Deterministic matching*, which uses exact matches of "known" identifiers, can offer elevated accuracy, but reduces scale and is limited by identifier availability and reliability/durability. *Probabilistic matching*, which uses statistical models and various attributes to infer matches, can increase scale but accuracy depends upon seed data quality and modeling sophistication. Often, both approaches are combined to balance accuracy and scale, but exact methodologies differ by stakeholder.

With regard to *timing*, matching can be one-time, always-on, or a hybrid. *One-time* matching resolves identities at a specific point in time, typically during data ingestion or processing, or on a schedule such as quarterly or monthly. It is simpler and less resource-intensive, but may not reflect the most up-to-date information (e.g., it presents a challenge in the case of rotating IDs). One-time matching may also be done *as-needed*, such as before the planning or targeting stage of a new campaign and again at the end of the campaign to gauge match "freshness" for measurement and attribution. *Always-on* matching provides continuous, real-time identity resolution, which supports timely insights but requires more computational power, implementation complexity, and cost.

Solutions & Technologies

The technical mechanisms underlying IDR practices and processes enable the sharing of identity-level data across stakeholders. Emerging solutions – such as data clean rooms – are offering ways to ingest and process inputs and outputs more securely and (more) in-line with privacy expectations. Newer solutions and technologies continue to evolve, may not be accessible or efficiently adoptable by all industry constituents, and, without standards for interoperability, can be complex and costly to implement. Understanding which solutions and technologies can and should be leveraged is highly dependent on the stakeholders one is working with and one's own capabilities. For more thinking on a few relevant solutions – including Data Clean Rooms and Federated Identity – please see the Appendix.

IDR Ecosystem Diagnostics

Key Takeaways

- *What is working well in the IDR ecosystem today?* The space is well-funded and competitive, with technological foundations in place and increasing investment to support further innovation.

- *What are the key problems in the IDR ecosystem?* The most pressing problems include:

  - Data quality and durability

  - Interoperability and usability

  - Ecosystem dynamics and expansion

- The overarching trends of ecosystem expansion and misaligned incentives amplify these challenges and create a negative feedback loop that makes finding and implementing solutions more difficult.

Our diagnostic assessment of the ecosystem has focused on four core areas:

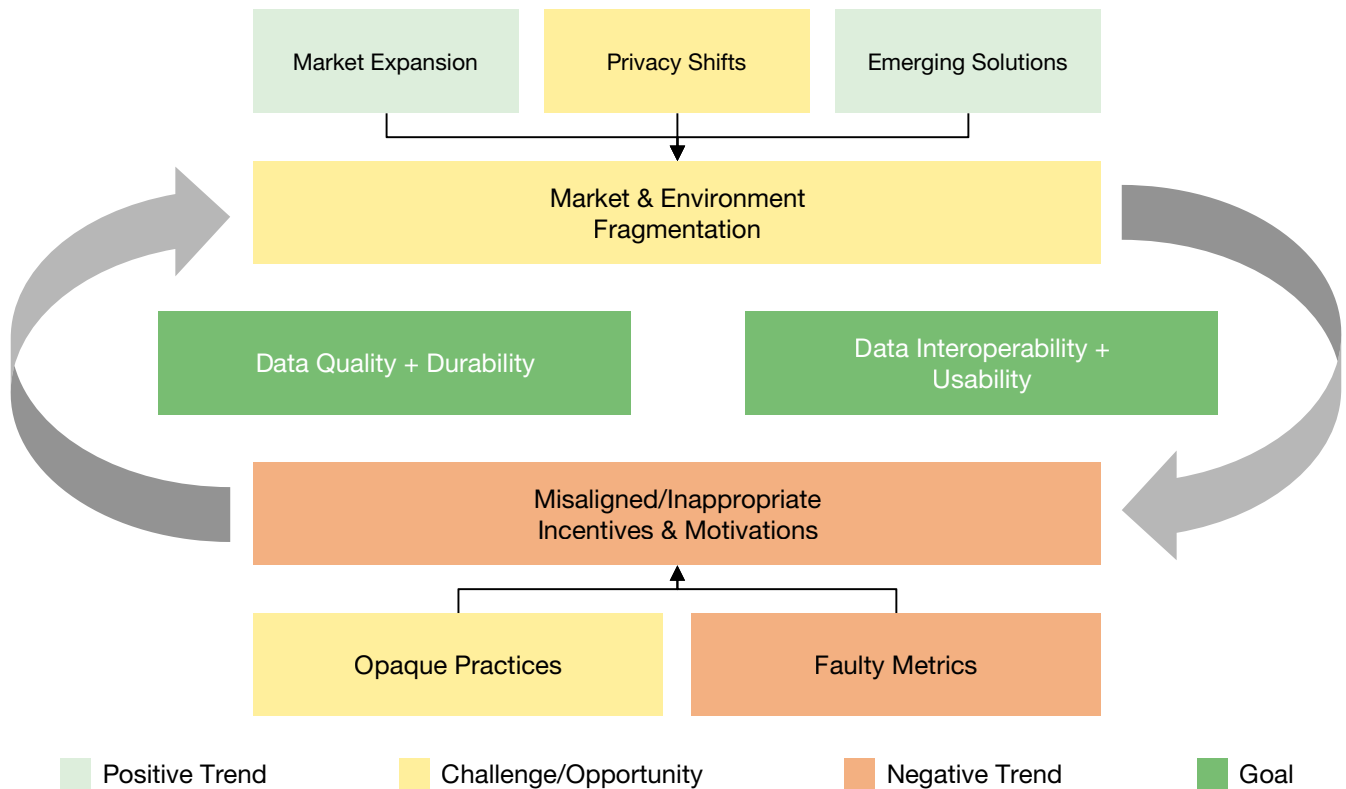| Themes	Descriptions
A. Quality 🟨	Identity data utility and timeliness • Accuracy (point-in-time and over time) • Performance and effectiveness • Reliability of underlying signals
B. Durability 🟦	Data persistence and resilience: • Persistent (vs. transient) nature of ID data • Resilience to evolving privacy regulations and platform policies
C. Scale 🟩	Level of industry adoption and ability to scale • Adoption by key industry stakeholders • Ability to perform IDR for audiences at scale
D. Interoperability & Usability 🟪	Enablement of sourcing and leveraging data across sources • Transparency and compatibility across stakeholders • Industry-standard consistency or lack thereof • Investment required to integrate and leverage on an ongoing basis

What is working well?

The ecosystem is well-funded and competitive, with established technological foundations and increasing investment fostering innovation.

Theme(s)	Working Well
Scale 🟩	**Established Technological Foundations:** *Technologies* are currently in place to enable a variety of advertising use cases. Existing technologies and tech stacks support connecting data across partners, platforms, and channels. *(Inputs/Outputs, Solutions & Technologies)* Newer identity resolution (IDR) *solutions* and *practices* are already on the market or in development to meet ongoing ecosystem shifts. *(Practices & Processes, Solutions & Technologies)* Ongoing integration of AI and cloud solutions is increasingly making data collaboration easier. *(Practices & Processes, Solutions & Technologies)*
Scale 🟩	**Dynamic and Competitive Ecosystem:** A well-funded,[11] competitive ecosystem of suppliers promotes innovation and development of new solutions. Diversity and breadth of available *solutions* offers choice, prevents market concentration, and fosters innovation. *(Stakeholders, Solutions & Technologies)*
Scale 🟩	**Evolving Business Practices:** Business practices are rapidly being developed and iterated. There is growing consensus and prioritization surrounding how to gather and enable identity data across stakeholders. *(Practices & Processes)* Ecosystem stakeholders increasingly invest in gathering first-party data. *(Inputs/Outputs)* Publishers are increasingly finding consensus around identity *practices* (e.g., authentication to capture HEMs). *(Practices & Processes)* Processes and workstreams to deploy identity solutions have been established across use cases (e.g., agencies sourcing providers, publishers leveraging identity solutions) – though not all are mature, efficient, or cost-effective. *(Practices & Processes, Solutions & Technologies)*

---

11   Category spending on identity resolution solutions is growing e.g. expected to reach $8.2 billion by 2024 (source: Statista).

What are the key problems?[12]

Today, the IDR ecosystem in ATVA faces several broad challenges across *data quality, scale, durability, and interoperability,* which are compounded by a fragmented, fast-growing *ecosystem and shifting privacy landscape. Critically, incentives are not always aligned, which can create* a negative feedback loop that complicates the development of pan-industry solutions.

| Market Expansion | Privacy Shifts | Emerging Solutions |

Market & Environment Fragmentation

| Data Quality + Durability | | Data Interoperability + Usability |

Misaligned/Inappropriate Incentives & Motivations

| Opaque Practices | Faulty Metrics |

| Positive Trend | Challenge/Opportunity | Negative Trend | Goal |

The resulting *lack of transparency and consistency* in definitions and methodologies is a priority challenge, although some stakeholders may resist changes because of the benefits they derive from the current system. Implementing transparency is also a challenge in itself, as there is a lack of consensus mechanisms and standards for communicating data quality.

---

12   See Appendix for full diagnostic table of problems and consequences.

Problems identified (with the themes they relate to) include:

Theme(s)	Problems	Consequences

Quality 🟨
 	
Interoperability & Usability 🟪
	Lack of Transparency and Consistency
	
	• Opacity and inconsistent methodologies and processes from vendors and partners for:
	   ◦ Data collection *(Inputs, Outputs)*
	   ◦ Data matching, sharing, and scaling: especially household-individual linkages, probabilistic methodologies, decision hierarchies, and hygiene for deduplication *(Practices & Processes, Solutions & Technologies)*
	   ◦ Data validation, especially screening and assessments of incoming data *(Inputs, Outputs, Practices & Processes)*
	   ◦ Lack of consensus and outdated methods for evaluating IDR data, practices, and processes with too much focus on accuracy and reach at the cost of "performance" *(Stakeholders, Practices & Processes)*
	• Inconsistent "privacy-safe" definitions applied to IDR *practices and solutions* (e.g., ID-based solutions marketed as "ID-less" when using obfuscation technologies, but no technical processes or guarantees) *(Practices & Processes, Solutions & Technologies)*
		Complexity and confusion, wasted investments (time, money) due to a lack of efficiency and effectiveness, enablement of bad actors/fraud, and erosion of (buy-side) trust

Durability 🟦 **Scale** 🟩	**Ad Ecosystem Stakeholder Expansion and Dynamics** • Proliferation of "proprietary" ID solutions *(Inputs/Outputs)* • Increasing number of stakeholders data flows through to enable a use case *(Stakeholders)* • Lack of urgency around alignment, retooling, and adoption of more durable identifiers *(Stakeholders, Inputs/Outputs, Practices & Processes, Solutions & Technologies)* • Lack of or mismatched incentives (e.g., revenue models, viewing IDR as a competitive advantage or concern, rather than as a utility) inhibiting provider collaboration for improvements *(Stakeholders)* • Walled garden data advantages and market-making positions *(Stakeholders, Inputs/Outputs)* • Talent scarcity and competition, especially against walled gardens and AI companies *(Stakeholders)* • Lack of governance for defining standards and coordinating collaboration *(Stakeholders, Inputs/ Outputs, Practices & Processes)*	Reactivity to privacy expectations and policies, erosion of consumer trust, wasted investment (time, money), and potential revenue loss to walled gardens

Theme(s)	Problems	Consequences
Scale 🟩 **Interoperability & Usability** 🟪	**Complexity and Cost (e.g., onboarding, operations, technology stacks)** Immature, manual, and resource-intensive workflows requiring custom implementation for each partner/provider *(Inputs, Practices & Processes)* Inability to reconcile measurement insights and performance across partners *(Outputs, Practices & Processes)*	Wasted investments (time, money), risks of human error and data leakage, and lack of ecosystem accessibility
Durability 🟦 **Quality** 🟨	**Reliance on Brittle or Ineffectual Identifiers and Resolution Practices** • Reliance on at-risk data identifiers OR outdated practices for associating identifiers and attributes to operate identity graphs/spines (e.g., IP-household linkages are not necessarily stable) *(Inputs, Practices & Processes)* • Inclusion of brittle/at-risk identifiers for reporting outputs *(Outputs)* • Inconsistent interpretations of what is "brittle" or not, e.g. IP address *(Stakeholders)*	Variability in identifier quality and durability, erosion of consumer trust, and enablement of bad actors, all of which may lead to even more stringent regulatory and platform policies
Quality 🟨	**Ongoing enablement of fraud**[13] • Outdated methodologies for validating identity *(Practices and Processes)* • Unchecked incentives for scale over reliability *(Inputs/Outputs)*	Wasted investments (time, money), contamination of data sets, and erosion of (buy-side) trust
Durability 🟦	**Consumer shifts, increases in:** • Ongoing viewership shift from linear pay-TV (more reliable physical identifier) to streaming (often less reliable digital identifiers) *(Stakeholders, Inputs)* • ID-sharing (e.g., sharing of email-based logins) *(Inputs, Practices & Processes)* • Privacy awareness and expectations *(Inputs/Outputs)* • Ad blockers and ID obfuscation *(Inputs)* • Generational and economic shifts limiting the reliability of physical identifiers as a "source of truth" (e.g., increase in renters vs. homeowners implies more addresses to reconcile) *(Inputs, Practices & Processes)*	Decreased data availability and utility for advertising use cases, increased fragmentation across channels
Durability 🟦	**Increasing/Misaligned Platform and Regulatory Requirements** • Piecemeal, difficult-to-interpret regulations *(Inputs, Practices & Processes)* • Platforms as decision-makers / gatekeepers for the collection and flow of data, with potentially little input from consumers or other ecosystem stakeholders *(Stakeholders, Inputs/Outputs, Practices & Processes)*	Decreased data availability and utility, compliance challenges and scrutiny, deprecation of legacy identifiers, and enablement of bad actors

13   Ad fraud is the practice of falsifying activity on or related to an advertising impression, which often results in undue financial gains for the fraudster. For example, for a streaming publisher, if a subscriber model is not in place and "viewership" is determined solely by email, it is possible to "fabricate" viewership data by purchasing email lists and including them in bid requests.

Priority Problem Areas

Based on research and consensus from stakeholder interviews, three key drivers of the problems hindering the IDR ecosystem in ATVA are:

1) **Lack of transparency and consistency**

Misaligned industry incentives have led to opaque solutions and practices lacking compatibility, which:

   a) Deter buyers and sellers through diminished trust.

   b) Create wastage and slow innovation.

2) **Brittle and ineffectual identifiers and resolution practices**

   a) Without deliberate plans to align on durable identifiers, many identity resolution practices and solutions lose efficacy over time (risk of obsolescence)

   b) Probabilistic graphs based on brittle or rotating IPs inherently lose accuracy, leading to duplicative reach, inaccurate measurement, and lower campaign performance (risk of degraded utility)

3) **Increasing platform and regulatory requirements**

   a) Legacy identity resolution practices are in many ways antithetical[14] to privacy regulations and require changes to be more durable.

   b) Piecemeal regulatory and platform-driven policy evolution has increased complexity and business risk for IDR providers and their customers.

---

14   This is because identity resolution still in many ways requires and implies a person-level view of identity, which privacy policies are attempting to curb or prevent outright.

Looking Ahead and Industry Collaboration Recommendations

Key Takeaways

- Expectations for the future of IDR in ATVA include 1) continued decline in the scale of physical IDs and the rise of new privacy-compliant physical and digital solutions and processes to fill gaps, 2) increased demand for accountability and interoperability, and 3) walled-garden dominance if industry collaboration is not achieved.

- Opportunities for addressing IDR challenges in the ATVA ecosystem depend upon 1) refocusing market demand and aligning incentives and 2) innovating with technology.

- The most important step the industry can take to improve the IDR ecosystem is to facilitate greater transparency around identity and IDR practices. However, to collaborate, vendors and the sell-side need buy-side motivation to overcome competitive concerns, which requires addressing inconsistent awareness, understanding, and urgency. To resolve this, a collaborative two-pronged approach is recommended:

  - *Develop buyer education and guides to facilitate*/provide 1) an understanding of ecosystem fundamentals and the costs of inefficiencies, 2) decision metrics for assessing IDR practices (including performance outcomes), and 3) curated questions for engaging with identity providers and agencies to gather information on data sourcing, matching, and validation (in part to support the RFP/RFI process).

  - *Define taxonomies and templates for providers to communicate about definitions and methodologies* for householding and mapping to individuals, and for communicating about how data is sourced, matched, and validated overall.

Expectations for the Future of Identity and IDR in ATVA

The identity and identity resolution ecosystem for ATVA will continue to evolve, intensifying existing challenges and introducing new ones. Some trends identified below are inevitable – most notably, the expansion of privacy-related laws and platform policies will continue to push the ecosystem towards novel solutions – but effective collaboration can mitigate the impact of others.

Looking forward, there is a broad consensus that the IDR ecosystem will be impacted by a more complex regulatory landscape and significant changes in many of the technologies underpinning the identity marketplace, such as:

1) **Continued reduction in availability of legacy identifiers** (*Inputs/Outputs*): Privacy-related platform and regulatory policies will further limit the availability, sharing, and utility of identifiers such as IPv4 and hashed emails. This could especially impact CTV and digital video, where IP addresses are often the key to linking data across sources.

2) **Expansion and development of more privacy-durable "identifier" solutions** (*Inputs/Outputs, Stakeholders, Practices & Processes*): To address growing gaps in data visibility, a wider range of identifier solutions will emerge, including anonymous deterministic IDs, alongside existing probabilistic and deterministic IDs. Likewise, investment and exploration of synthetic identifiers that are not tied to a single identity, such as Virtual IDs in the WFA Halo framework, will continue to gain momentum.

3) **Rise of decentralized, local processing** (*Inputs/Outputs, Practices & Processes, Solutions & Technologies*): Building on recent advances in AI / LLMs, decentralized processing – for example, on consumers' devices – with cryptographic matching (not reliant on traditional identifiers such as address or phone number) will gain traction because they provide privacy guarantees and may reduce costs.

4) **Increased demand for accountability and transparency** (*Stakeholders, Practices & Processes, Solutions & Technologies*): As the number of players in the ecosystem expands, understanding who provides what data to whom and the value that data generates will be important in enabling efficiency and establishing value-based pricing/compensation models. This will require greater openness and collaboration among stakeholders but could pave the way for more centralized identity systems and/or linked but decentralized identity systems, while still enabling vendors to differentiate by simplifying processes for the buy side.

5) **Increasing demand to make identity interoperable across planning, activation, and measurement use cases** *(Use Cases, Stakeholders):* Today, there may be limited matching of IDs and audiences across vendors and use cases – which makes the feedback loop of planning, measuring, and optimizing upon identity spines and strategies inefficient and costly. For example, from campaign to campaign, marketers might waste spend targeting ineffective portions of the audiences they purchase because current IDR systems cannot isolate and exclude those segments. Addressing this requires widespread adoption of common best practices and significant collaboration, which will take time and investment to realize.

6) **Dominance of walled gardens barring effective broader industry collaboration** *(Stakeholders):* If a broad base of stakeholders do not work urgently to improve the functionality of the more "open ecosystem", walled gardens – with their logged-in user bases and centralized decisioning systems – will dominate the digital video landscape.

Collaboration Opportunities

Given the diagnostic assessment, there are five areas of opportunity – ranging from aligning the market to improving core technologies – for strengthening and improving the operation of the IDR ecosystem for ATVA.

To align across buy-side and sell-side stakeholders, opportunities include:

1. *Crafting buyer-specific explainers, frameworks, and decision guides* to:

   a. Facilitate understanding of the inner workings of the IDR ecosystem (with this report as a starting point).

   b. Inform engagement with and selection of *ID providers,* including specific ID and ID graph vendors, *plus* any partners that may provide this data, regardless of their "primary" ecosystem role.

   c. Make sense of the patchwork of state and platform policies, such as the value of geo-based identifiers and how to take advantage of modeling and AI.

   d. Educate on how to detect and address fraud in the IDR context.

2. *Developing clear taxonomies, standards, and expectations* for:

   a. Communicating about identity and IDR methodologies (e.g., data collection, matching)

   b. Validating data, practices, and processes

   c. Enabling audits by trusted third parties

3. *Re-evaluating the primary metrics used to assess IDR, prioritizing performance* as a crucial element. While reach and accuracy (e.g., match rates as a contentious measure) have dominated, they may not translate to quality and tangible value for marketers or consumers across use cases. Prioritizing relative IDR quality and related impact on other effectiveness-focused performance metrics would motivate stakeholders to enhance the efficiency and efficacy of their data and IDR practices.

To update and innovate on core technologies, opportunities include:

4. *Utilizing artificial intelligence (AI) and machine learning (ML) technologies* to modernize and optimize IDR processes, while upholding privacy standards. This may include:

   a. Establishing AI/ML standards to provide greater consistency in how companies apply probabilistic identity resolution, shifting focus from reach to quality of IDR

   b. Developing strategies for AI to score, match, and calibrate identity graphs ahead of using the identity or related attribute data

5. *Rearchitecting around priority use cases and data inputs/outputs,* by:

   a. Fostering investment in and development of resilient IDR practices not reliant on cookies, IPs, or hashed email addresses. For example, finding new opportunities to apply distributed processing or modeling.[15]

   b. Defining and aligning on the recommended IDR uses and scope of privacy-enhancing technologies, data clean rooms, and other privacy-forward solutions. For example:

      i. Joining data for audience targeting via Private-Set Intersection (PSI)

      ii. Protecting personal information with Differential Privacy (DP) methods

      iii. Performing data augmentation for audience enrichment in a clean room

      iv. Employing aggregated or modeled outputs for planning and measurement

---

15   A key element of continuing to access any deterministic identity data – which improves modeling – is incentivizing consumers to share their data.

Collaboration Recommendations

Where to get started? We recommend stakeholders prioritize collaboration in areas where their goals are aligned – on maximizing ROI for buyers and revenue for sellers – and where they have the greatest agency to effect change. Today, the sell-side's focus on preserving any perceived advantage over competitors *within* the ATVA ecosystem may be blinding them to the significant, long-term threat of walled gardens. This, coupled with privacy-motivated changes and a legacy emphasis on scale as an indicator of data quality, is hindering collaboration, innovation, and opportunities for meaningful differentiation.

In parallel, limited buy-side awareness and visibility surrounding current IDR practices and the costs of existing inefficiencies is limiting stakeholders' ability to motivate change. While *collaborating and aligning on IDR technologies* (Opportunities 4 and 5) may be a future goal, the ecosystem is not ready to do this at scale because of misalignment and confusion, evolving regulatory and platform policies, and the risk of premature standards hindering innovation.

Therefore, the most important step the industry can take to improve the IDR ecosystem is to facilitate greater transparency and guardrails around identity and IDR practices. Motivating the sell-side to overcome competitive concerns to collaborate requires fostering demand from the buy-side by addressing buyers' lack of awareness, visibility, and urgency. Likewise, it is critical to establish a solid foundation upon which the industry can build more efficient and effective solutions for addressing ongoing and new challenges (e.g., with tooling and automation). To achieve this, we recommend concurrently focusing on a two-pronged approach of complementary collaborative work streams to develop:

- **Buyer Education and Guidance** (Opportunities 1 and 3)

- **IDR-specific taxonomies, templates, and best practices for communicating about inputs/outputs, practices/processes, and solutions/technologies** (Opportunity 2)

At the core of both of these workstreams is the notion that greater transparency, consistency, and specificity are needed to open and interpret today's black box of identity and identity resolution. Within IDR for ATVA, greater openness and information sharing is needed from sell-side players and ID providers[16] on methodologies, practices, and processes behind identity data, to enable effectiveness and find efficiencies.

Key transparency-related questions to answer:

- **Sourcing, collecting, scaling, and licensing data,** for example:

  ○ Does a provider own or license its ID graph?

  ■ If owned, what is the data collection methodology (e.g., sign-up flows)? What consent mechanisms are in place for data collection and sharing?

  ■ If licensed, from whom? How did the partner construct it? How is data from different providers reconciled? What additional processing occurs?

  ○ What *identifiers* are used?

  ■ How are individuals and/or households defined?

  ■ At what level is data collected (person, household, device)?

  ■ How often is data refreshed?

- **Matching and processing data** (for inbound and outbound data), such as:

  ○ How is person, household, and device data mapped, where needed (e.g., IP)?

  ■ IP address handling (what IPs are used, how rotating IPs are handled, how multiple devices with the same IP are resolved/differentiated)? How often are IPs updated? How is the reliability of the IP determined?

  ○ What rules or hierarchies inform how you approach IDR? What is used as the source of truth? How are accuracy, reach, and performance prioritized?

  ○ Which technologies are used for ingesting and sending data to/from systems?

  ○ How are discrepancies between data sources resolved?

- **Validation of data,** for example:

  ○ How is the data validated? By whom? How often?

  ○ What metrics are used to assess "quality"?

  ○ What source systems and client feedback mechanisms are in place?

While the two recommended workstreams ("Buyer Education" and "IDR Taxonomies and Best Practices") overlap in the topics they must address and the goal of facilitating partnerships across the buy and sell sides, they differ in whom it is critical to engage and which specific topics require focus.

---

16   Again, this includes specific ID and ID graph vendors and any other partners who may provide this data regardless of their "primary" ecosystem role.

Recommendations for Buyer Education and Guidance

What: A series of *reports, guides, and live sessions* with targeted messaging and broad outreach to buy-side stakeholders (i.e., marketers, agencies, and some publishers that rely on IDs/IDR from other stakeholders). Education will be especially critical for those who lead RFP/RFIs with agencies/vendors, to motivate adoption of transparency taxonomies (by aiding understanding of why and where transparency is beneficial) and how to interpret provided information.

Who: Must be spearheaded by a trusted, unbiased, multipartite industry trade organization, such as CIMM, that understands the broad perspectives and incentives of different stakeholders.

How: This must address three critical topic areas:

1. *Ecosystem Fundamentals* (with this report as a foundational source) on:

   a. *IDR Ecosystem Basics:* Facilitate understanding of the complexity of the ecosystem and the centrality of use cases for determining strategies, partnerships, and approaches to IDR.

   b. *Assessing IDR Quality:* Provide an updated framework for metrics used to assess IDR, inclusive of accuracy, reach, *and* performance. This should highlight how needs and criteria thresholds may vary by business, use case, channel, etc.

2. *Research reports and case studies:* Highlight and quantify the cost of current inefficiencies and durability risks on business outcomes. It is critical to help buy-side stakeholders understand how they could make better use of their limited resources and how decisions may change as a result of improving the ecosystem.

3. *Curated question guides for engaging with and evaluating partners* (see "key transparency-related questions"): This guide must address the questions outlined across data sourcing, processing, and validation and should complement developed IDR transparency templates (*see below*) that vendors should be asked to adopt. In practice, buy-side stakeholders must be able to easily incorporate these questions into RFPs/RFIs.

Recommendations for taxonomies and best practices for vendor and sell-side transparency

What: While auditable standards may be a future goal, the more immediate need is to align on how stakeholders communicate about IDR. This requires establishing *guidance and tools,* such as the foundational frameworks in this report (*see below*), for vendors and sell-side providers to communicate about their IDR methodologies. To encourage adoption, initial taxonomies and templates must be relatively simple and more open-ended, then grow more complex and specific as their application and the ecosystem evolve.
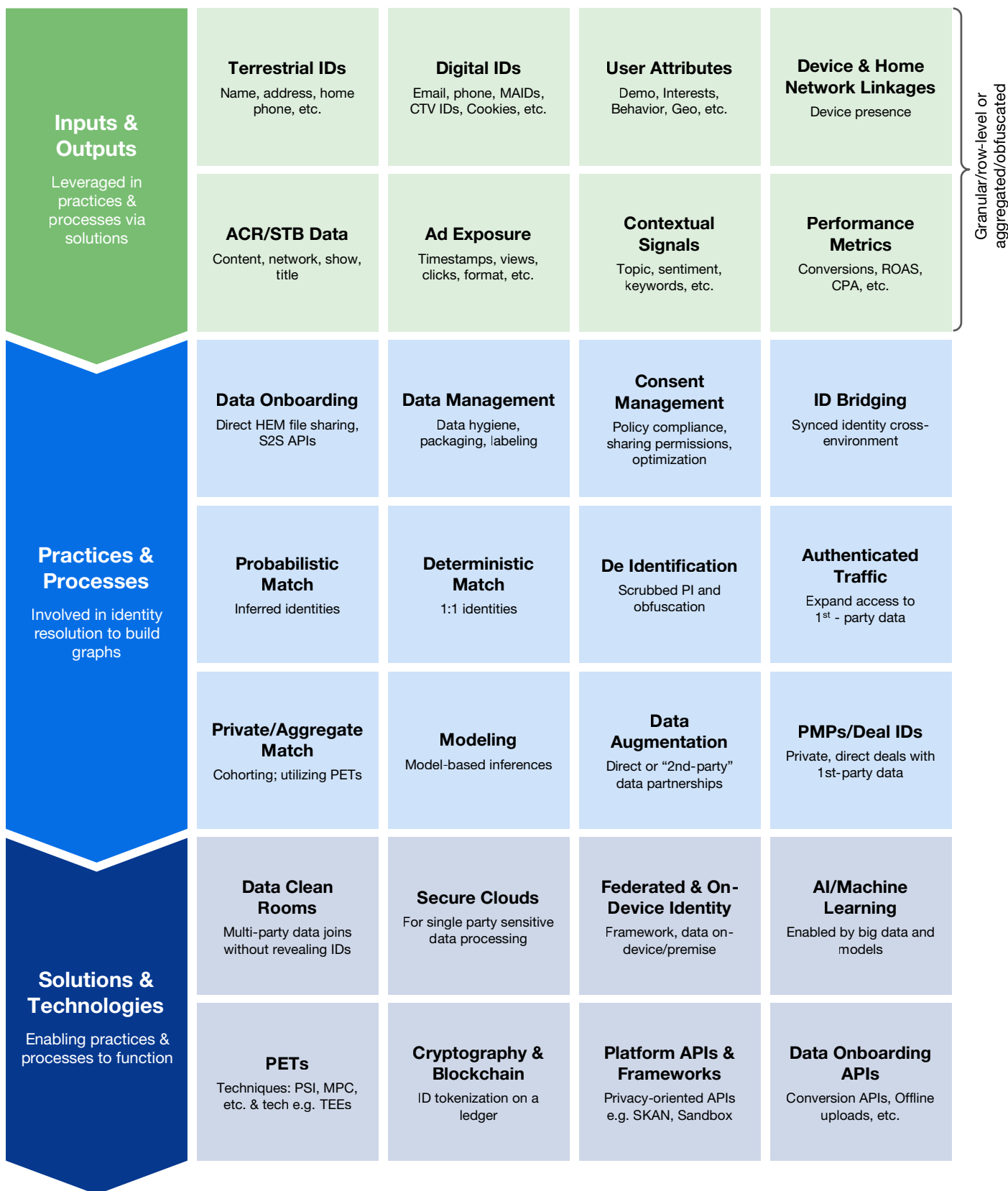
Inputs & Outputs
Leveraged in practices & processes via solutions

Terrestrial IDs Name, address, home phone, etc.	**Digital IDs** Email, phone, MAIDs, CTV IDs, Cookies, etc.	**User Attributes** Demo, Interests, Behavior, Geo, etc.	**Device & Home Network Linkages** Device presence
ACR/STB Data Content, network, show, title	**Ad Exposure** Timestamps, views, clicks, format, etc.	**Contextual Signals** Topic, sentiment, keywords, etc.	**Performance Metrics** Conversions, ROAS, CPA, etc.

Granular/row-level or aggregated/obfuscated

Practices & Processes
Involved in identity resolution to build graphs

Data Onboarding Direct HEM file sharing, S2S APIs	**Data Management** Data hygiene, packaging, labeling	**Consent Management** Policy compliance, sharing permissions, optimization	**ID Bridging** Synced identity cross-environment
Probabilistic Match Inferred identities	**Deterministic Match** 1:1 identities	**De Identification** Scrubbed PI and obfuscation	**Authenticated Traffic** Expand access to 1st - party data
Private/Aggregate Match Cohorting; utilizing PETs	**Modeling** Model-based inferences	**Data Augmentation** Direct or "2nd-party" data partnerships	**PMPs/Deal IDs** Private, direct deals with 1st-party data

Solutions & Technologies
Enabling practices & processes to function

Data Clean Rooms Multi-party data joins without revealing IDs	**Secure Clouds** For single party sensitive data processing	**Federated & On-Device Identity** Framework, data on-device/premise	**AI/Machine Learning** Enabled by big data and models
PETs Techniques: PSI, MPC, etc. & tech e.g. TEEs	**Cryptography & Blockchain** ID tokenization on a ledger	**Platform APIs & Frameworks** Privacy-oriented APIs e.g. SKAN, Sandbox	**Data Onboarding APIs** Conversion APIs, Offline uploads, etc.

Figure: Sample Foundational Taxonomy

Who: A coalition of ecosystem experts (with the power to enact and/or hold partners accountable to outputs) assembled by an unbiased industry organization that can ensure perspectives of the buy- and sell-side will be addressed. An existing industry body, such as CIMM or the IAB Tech Lab (Taxonomy Working Group), may be the most effective options. Regardless of who facilitates the initiative, substantive progress will require bold leaders within the group to take steps and implement the agreed, standardized practices.

*How:* The highest priority areas for industry alignment are:

- *Guidance and taxonomies for providers to articulate how they define, operationalize, and conceptualize identity layers and identifier types, especially regarding "households":* Currently, household(ing) varies by measurement provider and its selected methodologies and data points. Inconsistency in how mapping occurs across households, individuals, and devices reportedly leads to the greatest confusion and issues with interoperability and efficiency. In terms of identifiers themselves, confusion also surrounds IP addresses and their value. The frameworks provided earlier in this report for inputs/outputs, practices and processes, and solutions and technologies (*see example above*) can be used as a starting point for building taxonomies.

- *Tools and templates for communicating about identifiers and data sourcing, processing, and validation:* With buyers leveraging guides for engaging ID providers (as outlined in preceding section), there is an opportunity to streamline and ensure consistent, easily comparable information across providers. Getting started requires developing simple standardized templates for communicating about definitions, data sourcing, matching, and validation – to be shared by providers during the RFP/RFI process and updated periodically as practices change. Adoption of reusable and broadly understood templates to initiate conversations about IDR would alleviate frustration, focus conversations, and reduce the investment currently required by both buy- and sell-stakeholders to gather information about, select, onboard, and maintain providers. Motivating vendors and other sell-side players to adopt these templates will require enlisting major buyers to lead in requesting and holding providers accountable. *See below for an example of how a template might look.*

Ultimately, these recommendations are just initial steps towards establishing a more effective IDR ecosystem within ATVA. While these foundations may lead to better understanding "best practices" and thus greater consistency and standardization across providers, the market will determine how to build on these foundations and what to prioritize next.

Example tool for enabling transparency:

Transparency in IDR should not be overly cumbersome for sellers and ID providers to share or difficult for buyers to interpret. Ideally, what is requested by buyers and provided by sellers and ID providers would be standardized to make this process simpler. Consider the following illustrative data flow, which outlines a targeting use case within CTV, as a starting point for how communicating about IDR inputs/outputs, practices/ processes, and solutions/technologies could work in practice. This simple data flow could be expanded and modified to:

1) Diagnose areas of risks or points of failure for specific use cases.

2) Contextualize the cost of ecosystem inefficiencies within buyer guides (e.g., mapping proof points to numbered callouts in the diagram).

3) Demonstrate/Describe a provider's methodology for data sourcing and processing (leveraging the frameworks from the "Ecosystem Overview", such as practices/processes taxonomy). If each stakeholder that enables a use case provided an advertiser with insights to create a mapping similar to the below, the advertiser would be able to more readily identify and address critical areas of inconsistency across data flows and methodologies.
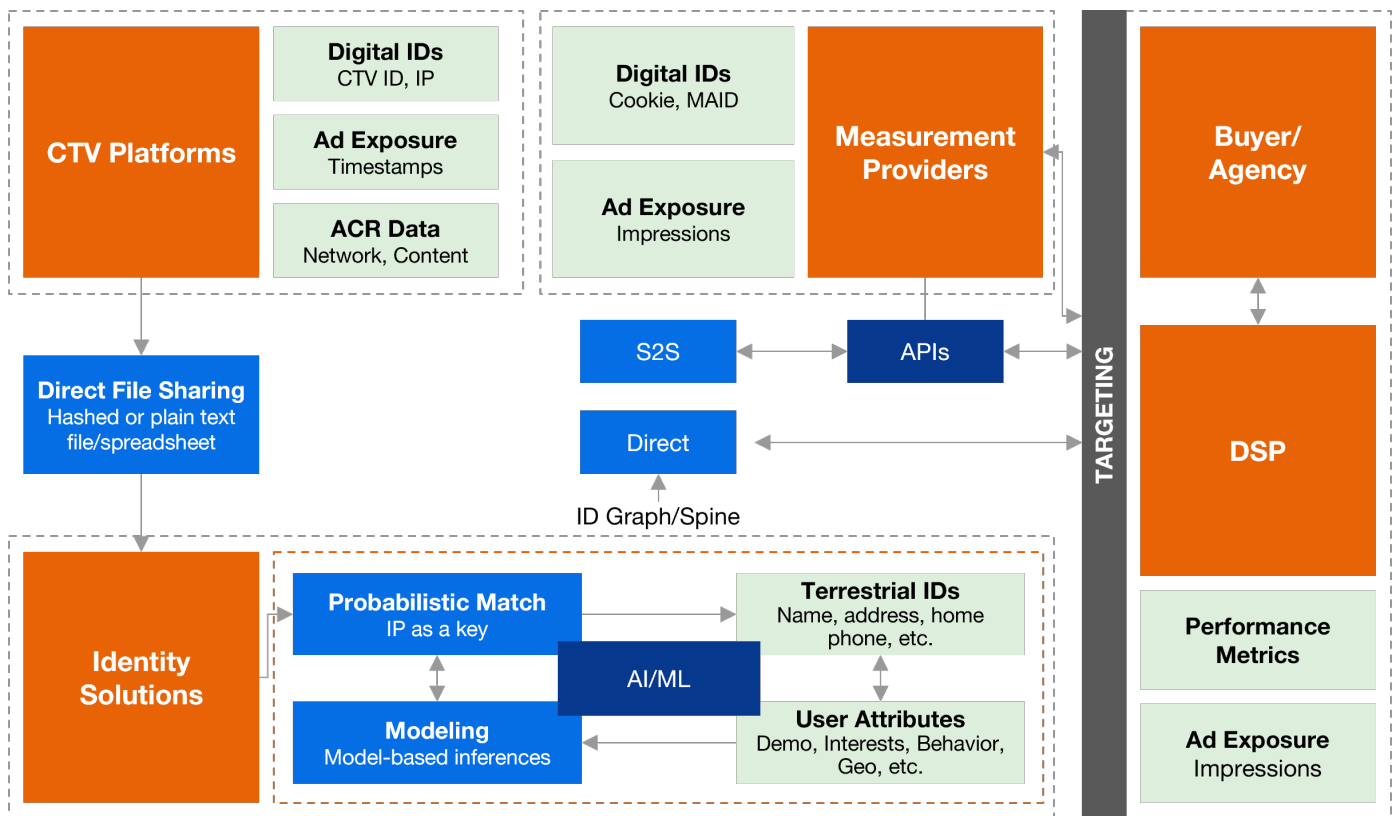
See Appendix for more details

With the right tools in place to facilitate awareness and transparency, sell- and buy-side stakeholders will be well-positioned to focus on follow-on collaboration opportunities, while better investing their own resources to create value for their own customers and business.

Conclusion

The Advanced TV and Video Advertising ecosystem is undergoing a period of remarkable growth and innovation. Identity and IDR serve as the backbone for enabling the planning, activation, and measurement of video advertising that is relevant, personalized, and effective. However, ATVA is at a pivotal moment that requires collaboration to maintain and unlock advertising opportunities and compete against walled gardens. While each stakeholder brings unique strengths and capabilities to the table, as macro-level dynamics and technologies advance, stakeholders that embrace collaboration and actively engage in shaping this landscape will be best positioned to reap the benefits. By working together – especially on increasing awareness and transparency – stakeholders can collectively create a more efficient and effective ecosystem that can continue to adapt successfully to future challenges.

APPENDIX

Inputs/Outputs and Solutions/Technologies (Additional Details)

Inputs/Outputs

Data/Party Relationships

Type	Characteristics	Examples
1st-Party Data	• Sourced directly from users • Owned by the collecting party • Explicit consent obtained • Full control	• Account sign-ups • Email subscriptions
2nd-Party Data	• Another party's 1st-party data shared through an exclusive partnership/agreement • Consent obtained by the original collector • Shared control	• Retailer loyalty program data shared with a brand • Airline sharing frequent flier data with a hotel
3rd-Party Data	• Collected from multiple sources • Owned by aggregators/sellers • May involve broad or inferred consent • Limited control	• Data from data brokers • Segments from data platforms

Solutions and Technologies

Data Clean Rooms
Ex: Snowflake, Habu/LiveRamp, InfoSum

Clean rooms are well-positioned as a solution in the IDR space due to:

- Their cross-environment data joining utility with built-in privacy guarantees[17] that can service multiple use cases,

- Their emerging nature as an established vendor category with robust solutions that can scale,

- Their increasing ecosystem adoption facilitates compliance with platform policies and privacy regulations.

Specifically, in an IDR context clean rooms can be leveraged to:

- Facilitate data joins between buyers and sellers and their identity providers by providing a secure, "neutral" collaboration environment without forcing any party to provide unencrypted datasets to a third party,

- Enable identity resolution without exposing underlying user-level identifiers, as long as data sharing, preparation, and processing best practices are respected by clean room providers, for instance through the use of PETs such as Differential Privacy.

Companies looking to adopt clean rooms to facilitate and/or power IDR should however be aware of several challenges:

1. A persistent lack of standardization and interoperability across clean room providers can cause friction;

2. The need to validate data security claims that aren't always backed by underlying technologies and practices;

3. Generally high technology fees and computing costs, as well as complex service agreement dealmaking and technical onboarding;

4. Only as good as the underlying quality of the data being sourced and validated

---

17 Note that although most clean rooms involve privacy-enhancing technologies, clean rooms are not inherently privacy-compliant and their data practices require careful audit before any selection.

Federated Identity

Federated identity solutions propose to securely and privately pool data from various stakeholders and provide ecosystem utility as a commodity, although in their current evolution, most examples are still for-profit businesses with their own commercial interests.

Key components of federated identity solutions include:

- *Household Identity:* census-level household identifiers derived from first-party subscriber data, leveraging privacy-preserving technologies to avoid leakage.

- A *Platform:* a data collaboration environment and underlying secure cloud/clean room to gather client inputs, perform computing operations, and provide data delivery services.

- *Ecosystem Data:* sourced from a collective of internet operators, publishers, distributors, marketers, platforms, and data partners focused on connected TV advertising;

- Only as good as the underlying quality of the data being sourced and validated.

Similarly to clean rooms, federated identity solutions remain new and complex, and involve potential cost concerns associated with their built-in privacy and security features.

IDR Ecosystem Diagnostics (Full Details)

What are the problems in IDR for ATVA (Full Diagnostic Table)

Theme(s)	Problems	Consequences

Quality 🟨
 Interoperability & Usability 🟪
	Lack of Transparency & Consistency
	
	 • Opacity and inconsistent methodologies and processes from vendors and partners:
	
	 • Data collection *(Inputs, Outputs)*
	
	 • Data matching, sharing, and scaling: especially household-individual linkages, probabilistic methodologies, decision hierarchies, and hygiene for deduplication *(Practices & Processes, Solutions & Technologies)*
	
	   ○ How are person <> household links established? (e.g., tying a CTV device ID to an individual)
	
	   ○ How are inconsistency and opaqueness of probabilistic matching techniques across IDR solutions causing inefficiency, introducing risk, and contributing to poor results? (e.g., probabilistically matching CTV IDs to graphs using IP addresses)
	
	   ○ What are other inherent gaps?
	
	 • Data validation/authentication, especially screening and assessments of incoming data *(Inputs, Outputs, Practices & Processes)*:
	
	   ○ How is data shared by media providers screened/evaluated?
	
	   ○ What data hygiene and deduplication processes occur before sharing data with partners?
	
	 • Lack of consensus and outdated methods for evaluating IDR-related data, practices, and processes *(Stakeholders, Practices & Processes)*
	
	   ○ Primarily industry focus on "scale or reach" (combining deterministic and probabilistic methods) without sufficient focus on quality of reach with related outcome metrics and transparency.
	
	 • Inconsistent "privacy-safe" definitions applied to IDR *practices and solutions* (e.g., ID-based solutions marketed as "ID-less" when using obfuscation technologies, but no standard for technical processes or guarantees behind the term) *(Practices & Processes, Solutions & Technologies)*
		• Confusion and complexity (e.g., of providers, operational processes)
		
		 • Lack of data quality and interoperability, reduced performance
		
		 • Inaccuracies in matching and activating consumer data
		
		 • Focus on potentially irrelevant metrics as a competitive differentiator
		
		 • Poor match rates and downstream performance
		
		 • Enabling and emboldening bad actors in deploying fraudulent schemes
		
		 • Decreased buyer trust, confusion, and frustration
		
		 • Debased product marketing efforts harming high-bar solutions vs less-stringent vendors
		
		 • Challenges validating if solutions meet business, regulatory, and/or platform policies

Theme(s)	Problems	Consequences
Durability 🟦 **Scale** 🟩	**Ad Ecosystem Stakeholder Expansion and Dynamics** • Proliferation of siloed, "proprietary" ID solutions. *(Inputs/Outputs)* • Increase in the number of stakeholders data flows through to enable a use case *(Stakeholders)* • Lack of urgency around alignment, retooling, and adoption of more durable identifiers, with too much focus on what will be lost. *(Stakeholders, Inputs/Outputs, Practices & Processes, Solutions & Technologies)* • Lack of or mismatched incentives (e.g., revenue models, viewing IDR as a competitive advantage or concern, rather than as a utility) inhibiting provider collaboration for more transparent, sustainable IDR practices. *(Stakeholders)* • Walled garden data advantages and market-making position. *(Stakeholders, Inputs/Outputs)* • Talent scarcity and competition, especially against walled gardens and emerging AI startups/companies. *(Stakeholders)* • Lack of governance for defining standards and coordinating collaboration. *(Stakeholders, Inputs/Outputs, Practices & Processes)*	• Reactivity to privacy expectations and policy deadlines (to ensure revenue is maintained) vs. proactively securing data and addressing privacy concerns • Continued erosion of consumer trust in the ad ecosystem to enable IDR in a way that meets consumer expectations, leading to more restrictive policies • Wasted spend and diminishing returns for buyers • Opportunity for walled gardens and others to absorb ATVA revenue
Scale 🟩 **Interoperability & Usability** 🟪	**Complexity & Cost (e.g., onboarding, operations, technology stacks)** • Immature, manual, and resource-intensive workflows requiring custom implementation for each partner/provider (Inputs, Practices & Processes) • Custom implementations required for data onboarding that differ from vendor A to B or channels X and Y implied costs of data preparation where required, etc. *(Inputs, Practices & Processes)* • Data visualization complexity, difficulty of reconciling measurement insights with bottom line performance, actionability challenges, etc. *(Outputs, Practices & Processes)*	• Wasted spending and team bandwidth • Increased risk of human error and data leakage • Lack of accessibility for less sophisticated ecosystem players • Decreased buyer trust

Theme(s)	Problems	Consequences
Durability 🟦 **Quality** 🟨	**Reliance on Brittle or Ineffectual Identifiers and Resolution Practices** • Reliance on at-risk data identifiers OR outdated practices for associating identifiers and attributes to operate identity graphs/spines (e.g., IP-household linkages are not necessarily stable) *(Inputs, Practices & Processes)* • Inclusion of brittle/at-risk identifiers for reporting outputs *(Outputs)* • Inconsistent interpretations of what is "brittle" or not, e.g. IP address regulations *(Stakeholders)*	• Wide variability in quality/ durability within identifier classes (e.g., some forms of IP are more stable than others) • Risk of solution obsolescence, PI leakage, and mismatches • Delaying necessary changes, which may later increase pressure/urgency • Erosion of consumer trust • Enabling bad actors to ignore consumers' privacy expectations • Further damaging the industry's reputation …all of which may result in even stricter regulatory and platform responses
Quality 🟨	**Ongoing enablement of fraud** • Outdated methodologies for validating identity *(Practices & Processes)* • Unchecked incentives for scale over reliability combined with lack of consequences for low-quality/low-value data. *(Inputs/Outputs)* See also "Lack of Transparency & Consistency"	• Wasted ad spend • Wasted time / $ to develop processes to "resolve" fraudulent data • Contamination of higher-quality data sets • Erosion of buy-side trust
Durability 🟦	**Consumer shifts** • Ongoing viewership shift from linear pay-TV (more reliable physical identifier) to streaming (often less reliable digital identifiers) *(Stakeholders, Inputs)* • ID-sharing (e.g., sharing of email-based logins) *(Inputs, Practices & Processes)* • Increases in awareness and expectation around data usage and privacy in advertising *(Inputs/Outputs)* • Usage of ad blockers and ID obfuscation (e.g., one-time use emails) *(Inputs)* • Generational and economic shifts limiting the reliability of physical identifiers as a "source of truth" (e.g., increase in renters vs. homeowners implies more addresses to reconcile) *(Inputs, Practices & Processes)*	• Shrinking pool of addressable consumer profiles leading to decreased data availability and utility • Fragmented views of consumers across CTV, Desktop, and Mobile • Inability to connect ad delivery and measurement

Theme(s)	Problems	Consequences
Durability 🟦	**Increasing (and sometimes misaligned) platform and regulatory requirements** • Piecemeal, difficult-to-understand regulations *(Inputs, Practices & Processes)* • Platforms as decision-makers / gatekeepers for the collection and flow of data, with potentially little input from consumers or other ecosystem stakeholders *(Stakeholders, Inputs/Outputs, Practices & Processes)*	• Decreases in the availability of identity-related data (e.g., data deletion requirements) and decreased data utility (e.g., performance, coverage, accuracy). • Challenges understanding whether IDR practices meet policies and, if not, how they need to evolve. • Increased regulatory scrutiny and fines for violation. • Unpredictability of future requirements, deprecations, and timing. • Deprecation of legacy identifiers (e.g., cookies, mobile IDs) and lack of stability in IDs across desktop/mobile (e.g., IDFA -> IDFV). • Bad actors taking advantage of confusion to continue legacy practices. • Refocus on 1st-party data creating an advantage for larger providers and those with direct consumer relationships over smaller businesses.

Industry Collaboration – Transparency Template Examples

Full Sample Diagnostics

Considering the level of ecosystem complexity, and the opaqueness of IDR practices caused by a combination of misaligned business incentives and genuine data sensitivity concerns, **we propose analyzing IDR challenges through a simplified data flow diagnostics lens** to evidence existing or future failure points. Consider the following examples:
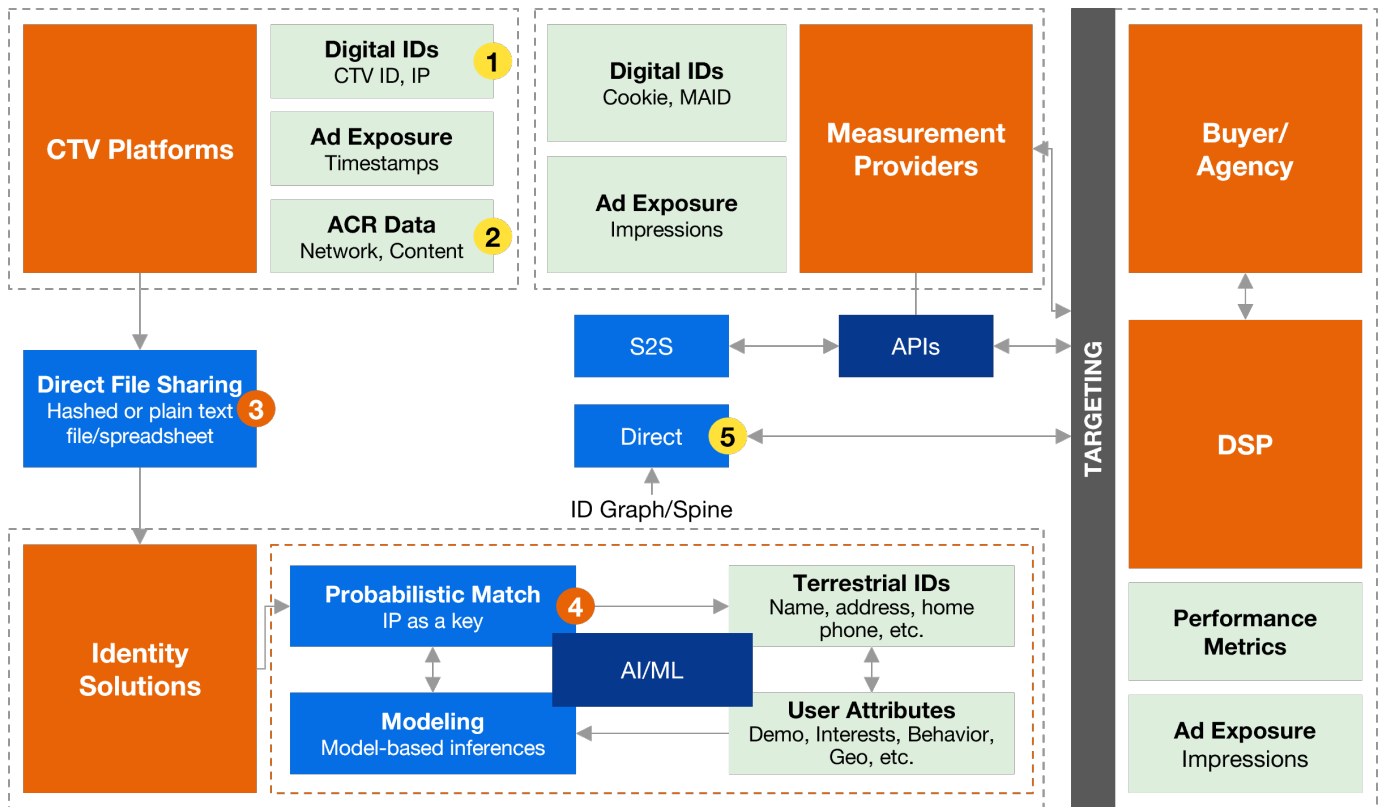
Targeting IDR in CTV environments

Figure: Simplified IDR data flows involved in targeting audiences in a CTV environment:

This illustration highlights five primary concern areas corresponding to the numbers above:

1. **CTV Device IDs**

   a. Help link people with their devices for targeting and measurement purposes, but may be deprecated (similarly to IDFAs by Apple) in the future

2. **Marrying content-level data with identifiers**

   a. Sharing content parameters in the bidstream alongside personal identifiers is often impossible due to VPPA[18] compliance requirements, although there are ways to decouple show-level data from personal identifiers via contextual segments and private deals

3. **Data sharing with IDR vendors**

   a. Legacy IDR practices still involve frequent direct file sharing among involved stakeholders, which creates data compliance and PI leakage risks

4. **Probabilistic matching**

   a. Often relies on IP addresses, which have rotating characteristics leading to duplicative identities throughout a campaign

   b. Frequently lack transparency on how identity resolution employing IP addresses is performed: where is IP address data sourced from, how often is it being updated, how is it validated, how is modeling achieved, with what other parameters, etc.

---

18   https://www.mindingyourbusinesslitigation.com/2023/10/the-return-of-the-vppa/

5. Reporting to DSPs and agencies

    a. May involve direct file sharing, which lacks scalability and creates further sensitive data leakage risk.

    b. Reports may prioritize reach metrics over business outcomes buyers are preoccupied with.
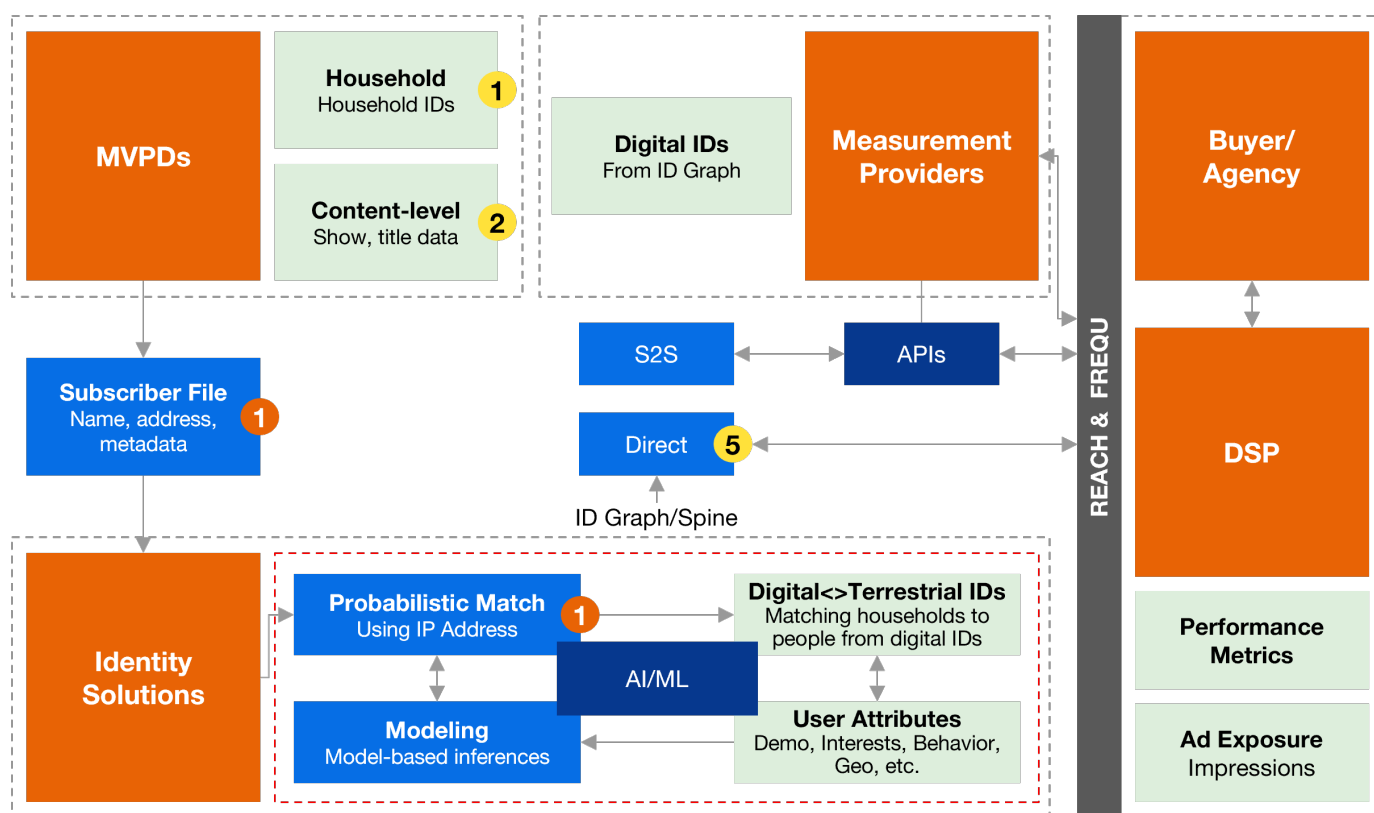
Delivery IDR with MVPD data

Figure: Simplified IDR data flows involved in targeting audiences in a CTV environment:

Three primary concern areas:

1. Household subscriber data sharing with IDR vendors; Household-to-People resolution

    a. Challenge accurately resolving household data with people-based digital IDs

        i. In the context of available IDR vendor's data in identity graphs/spines

        ii. In the context of brittle/rotating digital identifiers used as probabilistic keys

2. Content-level data

    a. MVPD data provenance brings inherent concerns from the risk of marrying identity data with content down the IDR chain, running afoul of VPPA

3. Joins with DSP/Agency data

    a. DSPs and agencies are more likely to transact with digital identifiers obtained via their own identity solutions and data partnerships, requiring durable join keys to match to household-level data. This process often involves probabilistic methods based on IP addresses, which have low durability and rotating characteristics to account for.

    b. DSPs and agencies also generally favor business outcomes KPIs that usually require 1:1 identities (vs household-based reach) for performance measurement.