Taxonomy and Data Transparency Standards to Support Seller-defined Audience and Context Signaling

Applying anonymized Audience and Content Taxonomy IDs and Data Transparency Standard metadata within OpenRTB to support privacy-centric addressability and first-party data monetization

March 2021

Draft Open for Public Comment Through May 7, 2021

This document has been developed by the Rearc Addressability Working Group, in cooperation with the Partnership for Responsible Addressable Media (PRAM).

With impending disruption to the identifier landscape, Project Rearc is a global call-to-action for stakeholders across the digital supply chain to re-think and re-architect digital marketing to support core industry use cases, while balancing consumer privacy and personalization. *The Rearc Addressability Working Group* is responsible for the evaluation of alternative technical standards and guidelines to drive "privacy by design" advertising, informed by input from the global business and policy dialogue within the Partnership for Responsible Addressable Media. The Addressability Working Group evaluates responsible technology alternatives to today's short-lived addressability mechanisms, and develops the technology foundations for tomorrow's consumer-centric solutions for ad targeting, measurement and optimization, while enhancing consumer transparency and industry accountability.

Rearc Addressability Working Group Roster

The Rearc Addressability Working Group Roster is made up of 295 individuals representing 146 organizations. Full roster details can be viewed here.

About IAB Tech Lab

Established in 2014, the IAB Technology Laboratory (Tech Lab) is a non-profit consortium that engages a member community globally to develop foundational technology and standards that enable growth and trust in the digital media ecosystem. Comprised of digital publishers, ad technology firms, agencies, marketers, and other member companies, IAB Tech Lab focuses on solutions for brand safety and ad fraud; identity, data, and consumer privacy; ad experiences and measurement; and programmatic effectiveness. Its work includes the OpenRTB real-time bidding protocol, ads.txt anti-fraud specification, Open Measurement SDK for viewability and verification, VAST video specification, and Datalabel.org service. Board members/companies are listed at https://iabtechlab.com/about-the-iab-tech-lab/tech-lab-leadership/. For more information, please visit https://iabtechlab.com.

IAB Tech Lab Contacts

Benjamin Dick
Sr. Director of Product – Identity & Data

Jordan Mitchell
Senior Vice President - Privacy, Identity & Data

Feedback on this RFC can be submitted to addressability@iabtechlab.com

Table of Contents

Goal

This document proposes an approach to addressability that revolves around the use of anonymized taxonomy nodes - sourced from IAB Tech Lab's Content Taxonomy 2.x or Audience Taxonomy 1.x - to signal publisher defined contexts or audience attributes within OpenRTB. The approach aims to support scalable, privacy-centric monetization of open web content and services while also minimizing disruption to responsible business activities and supply chain behavior. In focuses on leveraging existing open standards - including IAB Tech Lab's Content and Audience taxonomies, the OpenRTB specification, and the Data Transparency Standard - in a new way to ensure a dynamic and competitive open web ecosystem while also incentivizing transparent and accountable data access and use that's consistent with regional privacy expectations.

Design Principles

This approach is based on several design principles:

1. **User Transparency and Control** - system design needs to support regional expectations around consumer transparency and control of personal data.
2. **Data Security and Minimization** - approach should not rely on the passing of data that can be used for non-transparent or non-permissioned tracking - such as third-party cookies, mobile / OS IDs, user-provided IDs, or user-agent information. It should also suggest controls to restrict the commingling of data types that could pose privacy security risks without sufficient data protections.
3. **Technical Accountability to Consumer Preferences** - approach needs to be compatible with Tech Lab's Accountability platform, which introduces new tools to demonstrate technical accountability of supply chain participants to consumer preferences and data security expectations.
4. **Backwards Compatibility** - approach needs to minimize disruption to existing business models and competitive dynamics. It should not rely on complex and untested new tools that don't have broad industry consensus and supply chain interoperability, or which require costly / time intensive re-tooling that raise barriers to participation in the open ecosystem.
5. **Complementary to Other Addressability Approaches** - approach should not preclude other viable addressability system designs - including proposals that leverage secure, user-provided identifiers. It should support incremental addressability on devices when these other approaches are not technically feasible.
6. **Supports Industry Growth, Interoperability, and Competition** - approach should be sustainable, support innovation on top of common standards, and provide the necessary incentives for a competitive marketplace. Where possible,

a re-envisioned supply chain, as contemplated here, should provide opportunities for additional orthogonal benefits to consumers, publishers and platforms.

Referenced Documents and Specifications

- IAB Tech Lab - Content Taxonomy 2.1
- IAB Tech Lab - Audience Taxonomy 1.1
- IAB Tech Lab - Data Transparency Standard (DTS) 1.0
- IAB Tech Lab - OpenRTB 2.5
- Magnite “Proprietary Cohort” proposal
- Magnite “Gatekeeper” proposal
- Microsoft PARAKEET proposal
- Chrome “FLEDGE” proposal
- Chrome “Turtledove” proposal
- “Could A Consumer Taxonomy Fill The Identity Void In A Cookie-less World?”, Manny Puentes (CEO, RebelAI), AdExchanger, 7/3/2019.
- “How to Solve For Scalability of Publisher First Party Data”, Rachel Parkin (EVP, CafeMedia), AdExchanger, 9/16/20

Background

Today, online advertising systems rely on algorithms to group information associated with cross-site/app identifiers (e.g., 3rd party cookies and mobile IDs) into multiple audience segments or single cohorts. Sometimes this grouping relies on declared information, such as registration, and sometimes based on observed browsing behavior. Marketers use these audiences to match advertising content to the users who they believe to be most likely to engage and subsequently interact with their brand. This approach also helps publishers more effectively monetize their content while lowering ad loads, and users to see more relevant / less intrusive ads.

The tracking of individual user ids across page domains, made possible by cookies and mobile IDs, has been criticized because it exposes personal data without explicit consumer oversight or control over how their personal data is being collected and processed for advertising use cases. This has led to the deprecation of third party cookies, limitations on the availability of metadata that supports statistical IDs, and restrictions on the availability of mobile IDs by device and OS manufacturers in recent months, and led many industry participants to consider how proprietary grouping of users into anonymized audience cohorts - based on their browsing behavior - can be accomplished and communicated via OpenRTB without a dependency on the browser / OS itself (i.e., an alternative to what’s envisioned by Chrome’s Turtledove proposal,

which retains all decision making). These approaches are intended to address both marketer/publisher needs as well as the privacy and security concerns of sharing data with non-permissioned recipients.

Key Considerations

1. Given the present lack of adoption of the AdCom / OpenRTB 3.0 specification (largely because it's not backwards compatible with previous versions of the spec), this approach requires compatibility with OpenRTB 2.5.

2. IAB Audience Taxonomy was developed *after* the most recent version of OpenRTB, and is not referenced or accounted for in current documentation. Guidance will need to be added should a consensus develop around the in-band use of Audience Taxonomy nodes and DTS metadata as described in the approach below.

3. IAB Content Taxonomy is already supported within the existing OpenRTB 2.5 object model. This document will simply resurface and reframe current guidance based on the goals stated above. This appears at the end of the document.

Approach

Passing Audience Taxonomy IDs and DTS Metadata to Support Audience Targeting

Introduction to Relevant Tools and Resources

There are three existing specifications/resources within IAB Tech Lab's portfolio that can be used in conjunction with OpenRTB to support privacy-protecting audience signaling without exposing personal data beyond directly permissioned parties: Audience Taxonomy 1.x, Data Transparency Standard 1.0, and the DataLabel.org industry repository (www.datalabel.org).

The IAB Tech Lab Audience Taxonomy provides a standardized way to describe segmented audiences across demographic, interest, and purchase intent attributes. It establishes over 1600 standardized attribute nodes that, when used in combination with each other, can triangulate and describe a wide spectrum of niche audience characteristics. It is also intended to help facilitate comparability of "like" audiences across vendors that often have highly discrepant / proprietary naming conventions, and

was developed as a subcomponent of the broader Data Transparency Standard (DTS) program (standardized Audience Taxonomy classifications are one of the twenty required fields within DTS).

The IAB Tech Lab Data Transparency Standard (DTS) 1.0 is a standardized schema of up to 20 fields that establishes for any seller of data - whether independently monetized or bundled with media - a set of minimum disclosure requirements that the industry deems necessary for that sale to be "transparent" to the buyer. As mentioned above, inclusion of standardized naming conventions sourced from the Audience Taxonomy is a required field. These DTS disclosures aim to clarify key determinants of data quality - like provenance, age, extent of modeling, segmentation criteria, etc - but do not themselves constitute a "quality" determination that correlates to market value. This is largely due to the fact that "quality" is subjective and dependent upon the use of the data. As such, the Data Transparency Standard is often described as akin to an FDA "nutrition label". In version 1.0, the 20 fields within the DTS aim to clarify five core determinants of audience segment quality, however these are intended to evolve in future versions based on marketplace needs:

- **Data Provenance**: where was the data attribute sourced?
- **Data Age**: how long ago was the data collected, compiled, and then made available for online activation?
- **Data Modeling**: to what extent was the data manipulated or modeled?
- **Data Segmentation Criteria**: what are the qualifying business rules for a browser or device to be included in a segment?
- **Data Comparability**: when can one data segment be evaluated against another like segment?

Importantly, the Data Transparency Standard requires that the organization monetizing the data segment - regardless of whether that organization is solely responsible for determining the attribute or if it leverages downstream partners to help with that process - self-attests to the fields within the Data Transparency Standard. As such, given the potential for providers to misrepresent their data to buyers, IAB Tech Lab developed in 2019 an associated **compliance program** for the standard that allows providers to demonstrate the quality of their labeling via a Tech Lab "seal of approval" that's issued upon program completion. More information about the DTS standard and compliance program details / requirements can be found at www.datalabel.org. Additionally, the full DTS 1.0 schema and required fields can be found at www.datalabel.org or in the appendix below.

IAB Tech Lab's DataLabel.org repository is an industry resource for data labels produced by data providers that support the Data Transparency Standard. These labels

can either be ingested directly from participating data marketplaces via API, or batch-uploaded and managed directly by data providers within the datalabel.org platform. It provides a centralized location and UI for Tech Lab members to search and discover DTS metadata before making a purchase decision, however does not contain the actual segment data itself (just the descriptive metadata). As such, it can't be used for any form of audience ingestion or activation. Tech Lab members can access the repository via the "Sign In" button on the datalabel.org homepage, which will prompt users for valid Tech Lab tools portal credentials (https://tools.iabtechlab.com) or ask users to register with the tools portal. Associated API documentation for the Datalabel.org repository can be found here.

Application of Existing Tools Within OpenRTB

By leveraging these already-adopted specifications in new ways, publishers or their designated partners can reasonably determine audience attributes based on customer interactions on their properties, map those attributes to standardized taxonomy descriptions and data transparency disclosures, and relay those anonymized taxonomy IDs within OpenRTB to inform downstream signaling by buyers. This can be done at scale, without a reliance on bidstream information that could be used for cross-site tracking, and in a way that provides meaningful differentiation of and competition within seller defined audiences.

Cohort Membership and Size

This approach is based on policy interpretation which suggests that if an audience attribute can be assigned to a sufficiently large number of individuals - so as to not be able to re-identify any one individual, device, or browser that might be associated with an audience cohort - then that "cohort" signal satisfies consumer privacy requirements. This can be done without any personal information leaving the servers of the originating permissioned source. Cohort developers are expected to build and derive these groupings based on regional legislative requirements and expectations around consumer transparency and control features.

To understand what an adequate benchmark might be for the "sufficiently large threshold", we can look at existing policy interpretation from organizations with large privacy ethics and legal teams. For example, Google limits queries against cohorts of 50 or fewer users within Ads Data Hub as described on their developer documentation (see examples here, here, and here).

While this is a helpful directional reference point for the purposes of this primer, this threshold requires additional policy evaluation in three areas. First, guidance will need to account for the concept of "*unique users*" / average device counts per person. It also

needs to consider the concept of a "*lookback window*", which defines the amount of time in the past that devices can be counted within a cohort. Lastly, it needs to consider and account for the possibility that a minimum size within any audience segment could *disproportionately impact smaller publishers and brands* because observed activity might not as easily satisfy the threshold benchmark.

The Role of Transparency To Facilitate Differentiation and Competition Within Cohort Marketplace

Once an attribute is determined, web property owners or their trusted designees can compete with each other for buyer attention based on the quality and/or accuracy of their audience or content signaling. Buyers can learn over time which publisher cohorts generate the best marketing outcomes for individual tactics, then bid ("vote with their dollars") accordingly.

As with any marketplace, standards around transparency are foundational for this approach to be viable and scalable across publisher and format types. Specifically, efficient outcomes and marketplace liquidity requires line of sight into cohort effectiveness, as well as a consistent lexicon and definitional structure to correlate the outcome to the prior exposure. This is because web property owners might have different business rules or segmentation criteria to qualify the inclusion of a device or browser into a cohort, or use different language to describe the same segmentation practices. For example, an "Auto Intender" cohort from Publisher A will likely be unique and differentiated from an "Auto Intender" cohort relayed by Publisher B, despite using the same standardized name. Understanding the different business rules of cohort providers - as well as key differentiating factors like data provenance, age, compilation granularity, etc - and describing them consistently across the ecosystem facilitates pricing efficiency, ease of cohort discovery, and fairness. By improving the availability and consistency of information available among buyers and sellers, transparency standards promote greater accountability, and reduce the possibility for fraud or deceit within the digital advertising marketplace.

Automating Access to Cohort Metadata

IAB Tech Lab's Data Transparency Standard (the "data label") is a well-suited schema to provide the compositional transparency and consistent industry lexicon for efficient decision making within a cohort marketplace. Moreover, the industry repository of data labels at datalabel.org would be an effective, centralized tool to automate delivery of cohort metadata via API integrations. While the full set of 20 data label fields could not realistically be conveyed via OpenRTB real time, given payload implications, they could be retrieved out-of-band from the datalabel.org industry repository should the publisher include a sampling of unique data points from the standard. This combination of data points would allow downstream buyers to identify the unique label within datalabel.org

associated with the cohort and access a rich portrait of metadata associated with cohort provenance, modeling, age, and other characteristics relevant to a buyer's bidding decision. To reduce resource requirements, the full set of label metadata from the repository could be saved and referenced locally by buyer platforms in advance of bidding.

The minimum set of data points from the DTS schema that a cohort provider would need to convey to buyers - in order for those buyers to retrieve unique metadata / avoid collision - are as follows. These were selected based on a) how unique the data points are to the cohort developer, and b) the extent to which they could provide lightweight signals within bid requests:

- **Provider Name** - the unique domain of the provider / business entity making the attribute / cohort determination
- **Provider's [Internal] Segment Name** - the provider's internal ID associated with the audience segment referenced
- **Standardized Segment Name** - the provider's declaration of the standardized ID(s) that best describe its proprietary internal segmentations (as selected from IAB Tech Lab's Audience Taxonomy 1.x)

Because cohort providers often use multiple internal taxonomies to organize mappable audience characteristics, there is an inherent requirement to provide the flexibility to specify which internal taxonomy the provider's segment name / IDs refers to. While this **internal taxonomy name** is NOT currently a field in the DTS schema that would map directly to a data label in datalabel.org, it could be easily included in the API specification used to source datalabel.org metadata. In order for this information to be predictable and actionable by the buy-side, additional alignment is necessary around naming conventions within OpenRTB requests for proprietary taxonomies and possibly a centralized mapping resource of provider names + associated taxonomies in use. This standardization remains an open item for Rearc working groups. Note, additional explanation of suggested system design and data flows can be viewed in the document below.

Data Security and Accountability Within a Decentralized Cohort Marketplace

Within the proposed decentralized cohort marketplace, there are four areas where data security and accountability expectations need to be set to ensure responsible behavior with respect to consumer data access and use:

- Accountability of cohort developers to the accuracy of self-attested labeling
- Accountability of cohort developers to consumer privacy preferences
- Accountability of cohort developers to "sufficiently large" cohort threshold
- Accountability of supply chain participants to minimize commingling of cohort

data with other sensitive data types (including unique device IDs and user-agent information)

Below are descriptions of each, as well as suggested approaches for industry support.

Accountability of cohort developers to the accuracy of self-attested labeling
Cohorts and standardized datalabel.org descriptions are based on self-attested information.Self-attestation opens up immediate financial incentives for property owners to misrepresent the attribute being conveyed, or assign many different simultaneous attribute classifications to a single cohort. For example, a property owner might want to misrepresent a “new mother” cohort as “high net worth individuals” because of the higher value the market places on this attribute. Or a property owner might also falsely tag the “new mother” cohort with additional labels that suggest they’re also “in market for cars”, “hold multiple credit cards”, and are “high net worth individuals” to increase the likelihood of advertiser interest.

The compliance program attached to IAB Tech Lab’s Data Transparency Standard is designed to evaluate and affirm that organizations are completing the labeling accurately and have rigorous processes and technical checks/balances in place. Cohort providers can complete IAB Tech Lab’s Data Transparency Standard compliance program to signal to buyers the accuracy of their labeling. More information about program details can be found on datalabel.org.

Accountability of cohort developers to consumer privacy preferences
Cohort developers are expected to build and derive cohort groupings in accordance with regional legislative requirements and expectations around consumer transparency and choice. Participants are expected to participate in IAB Tech Lab’s Accountability platform, which introduces new tools to improve the auditability of supply chain participants to consumer preferences and data security expectations. More information about the IAB Tech Lab Accountability Platform can be found here.

Accountability of cohort developers to a “sufficiently large” cohort threshold
Cohorts are expected to be made up of a sufficiently large number of individuals so as to mathematically eliminate the possibility of being able to re-identify any one individual, device, or browser that might be associated with that cohort. While the industry has directional guidance on that cohort threshold (see above), more analysis on an industry standard is required. Similar to individual privacy preferences, IAB Tech Lab’s Accountability platform establishes open data sets that allow participants to demonstrate cohort sizes in meaningful ways and participants would be expected to integrate accordingly.

Accountability of supply chain participants to minimize commingling of cohort signals with other identifiers

This approach expects publisher defined cohorts to be conveyed in stream to buyers in isolation from other device-specific data like user-agent information, pseudonymous identifiers, or encrypted user-provided identifiers. This is intended to minimize the possibility of two separate but related scenarios: 1) core consumer privacy concern associated with non-transparent device mapping and behavioral profiling, and 2) commercial sensitivities to publisher business models related to audience data leakage. Below are descriptions of each scenario:

- *Layering Probabilistic Device Maps with Audience Attributes*: consumer transparency into who has access to their data, and choice over how its used, are foundational components of a healthy and sustainable supply-chain. Privacy and security engineers have long established the threats to non-permissioned use of consumer information created when basic machine learning models are applied to openly available publisher bidstream data that contains pseudonymous IDs, user provided IDs, user-agent information, or other data types that could be collected over time and used to re-identify a device across contexts.This ability to maintain non-transparent and non-permissioned probabilistic mappings of devices based on bidstream information becomes especially invasive should distilled publisher-declared attributes - focused on demographic, interests, or purchase intent characteristics - be available in the bidstream to inform these profiles.

- *Publisher Data Leakage*: ad-supported publisher business models revolve around monetizing their web properties by ensuring the opportunities they offer to engage audiences deliver value to marketers. One method some publishers rely upon is to cultivate specific audiences. Publishers expend considerable resources to build and cultivate audiences, and compete with each other for advertiser investment on the basis of the value and size of the audiences their content attracts. Should device specific data be commingled with cohort IDs at scale, it would facilitate publisher data leakage scenarios whereby an audience cohort identified by premiumpublisher.com could be re-identified by an advertiser on bobsblog.org, for perhaps a much cheaper price. This dynamic would inadvertently commoditize publisher audiences, disincentive innovation and investment in online content and service, and erode competition in the marketplace.

There are entities in the supply chain that are well-positioned to systematically constrain

the commingling of device level data - including user agent information, first party identifiers, probabilistic maps of various IDs, and/or encrypted user-provided IDs - should an audience cohort ID be declared by a publisher within a bid request. Assuming consumer transparency and choice has been respected, the choice of whether to convey an anonymized cohort ID versus some other proprietary / commercial identifiers (or vice versa) is a business decision that first parties should control. However, commingling these data points should always be avoided. Descriptions of the entities best positioned to validate and curate bidstream data are below, *however it remains an open question for the industry to align on specific roles and functions.*

- *Header Wrappers* - these are entities that facilitate unified auctions across multiple exchanges. They represent an important technical intermediary between publishers and the SSPs/exchanges that relay bid requests to buyers.

- *SSPs* - sell-side platforms, which often operate header technologies for publishers, are responsible for normalizing bid request signals from publisher clients and optimizing the incoming demand to maximize publisher yield.

- "Trusted Server" - the concept of a trusted server has become a fixture in browser and industry standards conversations. It refers to an external server - usually operated by an organization that does not buy or sell media - that would provide, among other services, tools to anonymize ad requests. Below are relevant proposals on how trusted servers could be implemented to support this use case:
    - Microsoft's PARAKEET proposal
    - Magnite's Gatekeeper proposal

Relevant Fields Within the OpenRTB Object Model

The following sections provide context around what is currently supported within OpenRTB fields, and suggests how these taxonomy IDs and sampling of DTS signals could be mapped to the existing OpenRTB object model to support the passing of audience cohorts without modifications to the specification, and retrieval of the full set of DTS metadata out of band from datalabel.org.

There are three existing objects within OpenRTB 2.5 that - in combination - are well suited to support Audience Taxonomy IDs and select DTS metadata: User Object, Data Object, and Segment Object:

User Object

This object is intended to contain information known or derived about the human user of the device (i.e., the audience for advertising).

Attribute	Type	Description
id	string; recommended	Exchange-specific ID for the user. At least one of id or buyeruid is recommended.
buyeruid	string; recommended	Buyer-specific ID for the user as mapped by the exchange for the buyer. At least one of buyeruid or id is recommended.
yob	integer	Year of birth as a 4-digit integer.
gender	string	Gender, where "M" = male, "F" = female, "O" = known to be other (i.e., omitted is unknown).
keywords	string	Comma separated list of keywords, interests, or intent.
customdata	string	Optional feature to pass bidder data that was set in the exchange's cookie. The string must be in base85 cookie safe characters and be in any format. Proper JSON encoding must be used to include "escaped" quotation marks.
geo	object	Location of the user's home base defined by a Geo object (Section 3.2.19). This is not necessarily their current location.
data	object array	Additional user data. Each Data object (Section 3.2.21) represents a different data source.
ext	object	Placeholder for exchange-specific extensions to OpenRTB.

Data Object

The data and segment objects together can communicate additional data about the related object specified. This data may be referenced from third parties as specified by the id field.

Attribute	Type	Description
id	string	Exchange-specific ID for the data provider.
name	string	Exchange-specific name for the data provider.
segment	object array	Array of Segment (Section 3.2.22) objects that contain the actual data values.
ext	object	Placeholder for exchange-specific extensions to OpenRTB.

Segment Object

Segment objects are essentially key-value pairs that convey specific units of data. The parent Data object is a collection of such values from a given data provider. The specific segment names and value options must be published by the exchange a priori to its bidders.

Attribute	Type	Description
id	string	ID of the data segment specific to the data provider.
name	string	Name of the data segment specific to the data provider.
value	string	String representation of the data segment value.
ext	object	Placeholder for exchange-specific extensions to OpenRTB.

Mapping OpenRTB Object / Attribute Fields to DTS Fields

Given the existing model described above, the following is a suggested mapping of User Object, Data Object, and Segment Object fields to core DTS fields. This can be done without manipulating or modifying current ORTB implementations.

Open RTB 2.5		**Data Transparency Standard 1.0**			
Object	*Description*	*DTS Field Name*	*Field Representation*	*Field Options*	*Description*
User.data.name	Exchange-specific name for the data provider	**Provider Name**	**Provider *Domain***	String	Unique domain of the business entity making the attribute determination (context or audience)
User.data.segment.id	Name of the data segment specific to the data provider.	**Segment Name**	**Provider Segment ID**	String	Provider's internal ID of audience segment referenced (which can be used to retrieve broader sets of metadata associated with DTS).
User.data.segment.id	String representation of the data segment value.	**Standardized Segment Name**	**Standardized Taxonomy ID**	String	List of the most accurate standardized IDs as selected from IAB Tech Lab's Audience Taxonomy 1.x -or- Content Taxonomy 2.x.

JSON Example of Audience Cohort Signaling

User.data{} is an object array that can support the following flexibility per impression opportunity:

- Multiple Cohort Providers - segment name and standard segment ID can be individually listed per cohort provider
- Multiple Cohort Taxonomies - cohort providers have flexibility to define different taxonomies (both proprietary and standardized) that IDs might be associated with. Regardless of the internal taxonomy used, an associated JSON mapping to the standardized Audience taxonomy IDs would be required.

Below is an example of how the JSON would be structured:

```
  "user": {
    "data": [
      {
        "name": "www.dataprovider1.com",                    Provider Name 1 (Domain of entity making attribute determination)
        "ext": { "taxonomyname": "proprietary taxonomy abc" },
        "segment": [
          { "id": "687" },                                  Provider Segment Name (provider's internal Segment ID)
          { "id": "123" }                                   Provider Segment Name (provider's internal Segment ID)
        ]
      },
      {
        "name": "www.dataprovider1.com",
        "ext": { "taxonomyname": "IAB Audience Taxonomy 1.1" },
        "segment": [
          { "id": "687" },                                  Standardized Taxonomy ID(s) (Audience Taxonomy 1.x)
          { "id": "123" }                                   Standardized Taxonomy ID(s) (Audience Taxonomy 1.x)
        ]
      }
    ]
  }
}
```

Prebid.js Support for Cohort Signaling Within OpenRTB

Prebid supports this taxonomy signaling approach and is currently mapping the Prebid.js First Party Data (FPD) object to OpenRTB and providing field validation. Associated documentation can be found with the following issues on the /prebid.js repo:

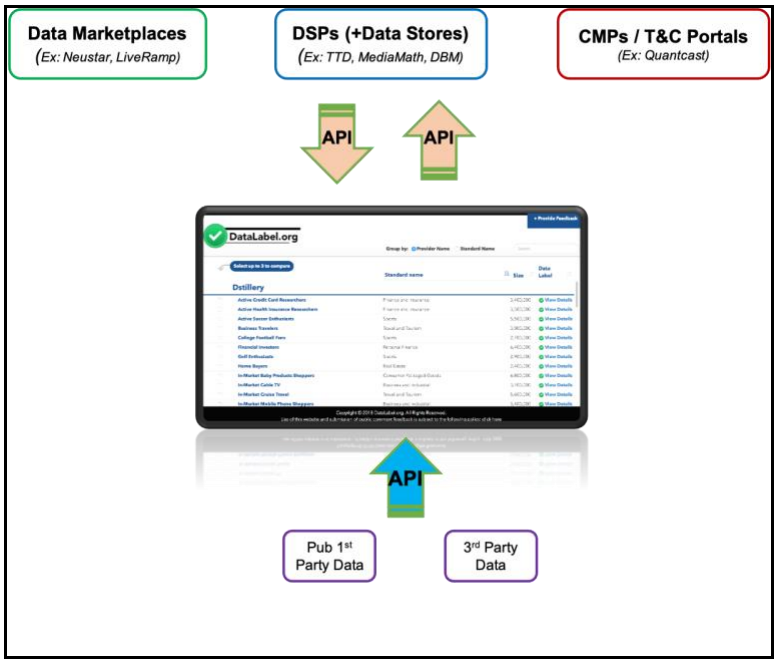
- #6057 Proposal for Taxonomy Segments in FPD
- #5795 First Party Data Revision (for ORTB)
- #6099 First Party Data Module

There are currently ten adapters that read the Prebid FPD object that would need to be migrated.

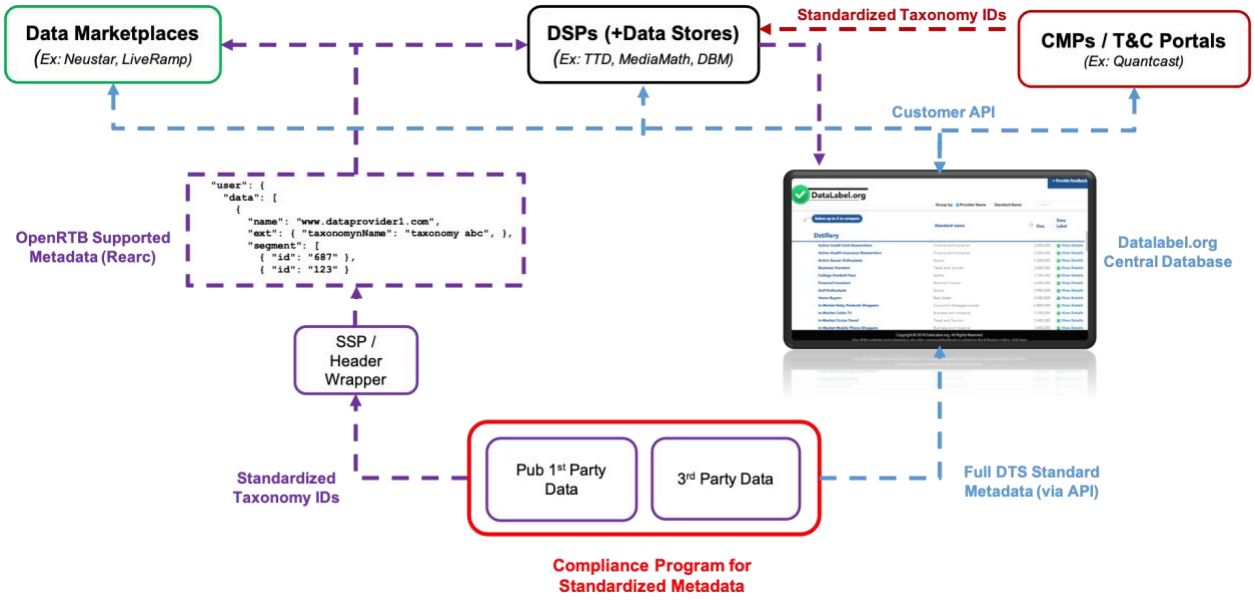
Dataflows Between Cohort Provider, OpenRTB, and Datalabel.org

If a cohort provider were to include the cohort provider's name, provider's internal taxonomy name, provider proprietary ID, and standardized ID mapping within bid requests, a downstream entity could feasibly retrieve the full set of DTS metadata associated with an audience cohort (should the provider support the industry's data transparency standard) via out-of-band API access to the centralized label repository at datalabel.org.

The following assumes there are three primary types of downstream parties interested in the full set of datalabel.org metadata - DMPs, DSPs, and CMP/T&C portals - however this shouldn't preclude utility for other parties that may want to integrate.



Data flows between cohort provider, OpenRTB signals, and Datalabel.org could be envisioned as follows.



1. Publishers or their trusted partners determine audience attributes / cohorts based on local O&O visitation

2. Audience cohort developers populate DTS metadata locally that reflect segmentation criteria and business rules governing inclusion in the cohort (in accordance with the industry standard schema in appendix) and push metadata

to datalabel.org via API

a. Audience cohort developers that have validated the process by which they self-declare metadata via the DTS compliance program will be signaled by the datalabel.org platform.

3. As audiences of sufficient size navigate to publisher pages, relevant provider information (name, internal taxonomy name, internal taxonomy ID) and anonymized audience signaling (Audience Taxonomy 1.1) are relayed to downstream entities within OpenRTB via Prebid header bidding integrations.

4. Parties interested in bidding on anonymized audience signals can call into datalabel.org via API to retrieve relevant metadata associated with that cohort's segmentation criteria, compilation granularity, provenance, age, modeling or other information within the schema's 20 fields. This can be done in real time or based on a local cache that's regularly refreshed from datalabel.org.

5. Upon creative render, publishers receive real time impression level granularity to inform core functions, including ad placement optimization and debugging operations.

Other Industry Use Cases and Utility of Cohort Metadata

Beyond privacy-centric audience signaling, this taxonomy-based approach and centralization of industry cohort metadata facilitates other peripheral industry benefits.

Streamlined integration footprint for DSPs, DMPs, data providers

Currently, marketplaces where audience data is bought/sold need to maintain dozens of API integrations with data providers. Similarly, data providers work with many data marketplaces concurrently. This many-to-many integration footprint introduces significant operational and technical costs for both marketplaces and data providers, and ultimately creates unnecessary duplication of work within the supply chain. A single repository of metadata managed by a neutral industry trade body on behalf of the industry - which serves to broker descriptive segment metadata for all parties that in turn could support many valuable use cases - would reduce that burden to a single integration. The following sections provide more detail on how the extensibility of a common datalabel.org platform could facilitate innovative uses of the metadata and proprietary innovation on top of this industry resource.

More Informed Bidding Decisions

Bidding logic within DSPs and other buy-side platforms - almost always aimed at maximizing over time a wide range of pre-established KPIs like cost-per-metric and

quantity-of metric goals (page visits, clicks, actions, etc), while constrained by pacing (impression, budget), time, budget, exposure frequency and geographic relevance - is informed by a combination of trader parameters and proprietary algorithmic decision-making. If this machine learning were to have greater access to a rich set of DTS metadata - which collectively informs the underlying “effectiveness” and “accuracy” of the attribute determination by accounting for things like data provenance, age/refresh characteristics, modeling, offline data handling details, etc - marketers might uncover new opportunities to increase their effectiveness. They could do this by tactic, inventory source, cohort provider, geography and more. Over time this should produce many desirable outcomes: improve marketing efficiency, re-allocate media investment to the most valuable inventory and data sources, influence data sourcing practices, create healthy monetary incentives around data transparency, and improve consumer experiences online. The timely training of this modeling process will be important to be able to offer marketers value across open web inventory relative to closed, vertically integrated platform publishers.

Connective tissue to Support Consumer Facing Disclosure Information

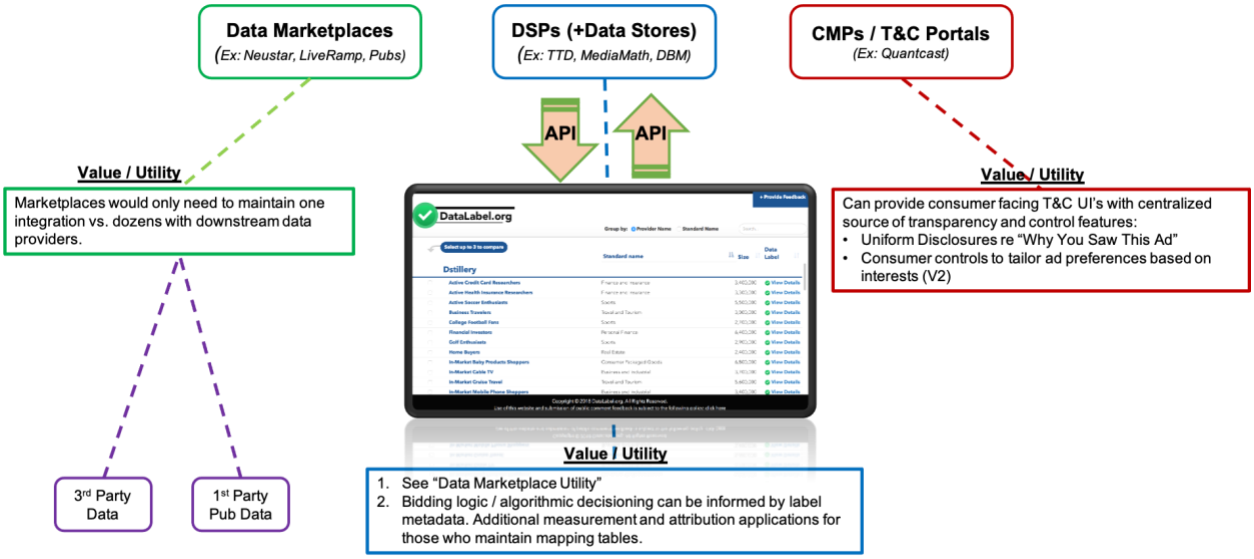
As our industry begins to compete on privacy as a core feature and participants jockey for consumer trust around data collection practices, vertically integrated platforms are inherently better positioned to surface consumer facing transparency information and provide actionable consumer controls over access, use and revocation. Examples of recent data labeling initiatives include Apple’s Privacy Nutrition Label, Facebook’s “Privacy Checkup” and “Off Facebook Activity” features, Google / Chrome’s Ads Transparency Spotlight tool, and Microsoft / Edge’s Transparent Ads Provider program. All of these initiatives have similarities in supported use cases and approach:

- Supported Use Cases: what data is being collected, by whom, why a specific ad appeared, who delivered that ad?

- Approach: all rely on self-reporting, incorporate the concept of data provenance (of “tracking” data), focus on data currencies used for the tracking, and provide information about other data these currencies have been linked to

In an open disintermediated ecosystem, independent ad tech has a much more difficult task in order to execute on consumer transparency features and needs more meaningful B2B interoperability and a reliable supply chain for transparency metadata. In this context, a centralized, extensible industry repository of audience metadata at datalabel.org can become valuable as a source of uniform disclosures about why a consumer saw a given ad (ie, the entity delivering, where the audience data came from, and what business rules apply to that audience data). Additionally, in the short to mid-term, it would be easy to envision datalabel.org supporting consumer control features

using standardized taxonomy signals sourced from IAB Tech Lab's Ad Product Taxonomy and the Audience Taxonomy. The Ad Product Taxonomy provides a standardized way of describing the products or services contained in an ad and could be used as a proxy for the acute/immediate products a consumer might be interested in learning about. The Audience Taxonomy describes the interests/intents of an audience and could be used as a proxy for a consumer to convey a broader array of product interests that revolve around long term hobbies or purchase behavior.



Required Modifications to DataLabel.org Platform and GTM

The approach suggested above would require two modifications to the existing datalabel.org platform and GTM:

1. The data ingestion API for the DTS program is currently designed to source labeling directly from data marketplaces instead of data providers. A slight modification will be required to account for proprietary taxonomy names for the audience cohort provider (as referenced above).
2. A second "Extensibility" API (see "Customer" API above) would need to be developed based largely on the data ingestion API structure, which would allow for industry participants to call into the datalabel.org repository and request data labels associated with the information contained within OpenRTB requests.

Passing Content Taxonomy IDs to Support Contextual Targeting

IAB Content Taxonomy IDs provide a standardized way of describing the "aboutness" of a website or app across browser, mobile, or OTT environments. Importantly, it also delineates "aboutness" from additional attributes of content context that can be signaled

within the spec, such as content language, form factor, origin, and media type. All of these more granular descriptors beyond “aboutness” nodes also have unique, dedicated IDs. If implemented and used correctly, relaying a combination of Content Taxonomy IDs across these vectors can help publishers communicate rich and nuanced content descriptions which can then be used for more informed decisioning by downstream buyers. OpenRTB 2.5 supports the inclusion of multiple Content Taxonomy IDs within the “cat” string array within the **Content Object**:

Attribute	Type	Description
`id`	string	ID uniquely identifying the content.
`episode`	integer	Episode number.
`title`	string	Content title. *Video Examples:* “Search Committee” (television), “A New Hope” (movie), or “Endgame” (made for web). *Non-Video Example:* “Why an Antarctic Glacier Is Melting So Quickly” (Time magazine article).
`series`	string	Content series. *Video Examples:* “The Office” (television), “Star Wars” (movie), or “Arby ‘N’ The Chief” (made for web). *Non-Video Example:* “Ecocentric” (Time Magazine blog).
`season`	string	Content season (e.g., “Season 3”).
`artist`	string	Artist credited with the content.
`genre`	string	Genre that best describes the content (e.g., rock, pop, etc).
`album`	string	Album to which the content belongs; typically for audio.
`isrc`	string	International Standard Recording Code conforming to ISO-3901.
`producer`	object	Details about the content `Producer` (Section 3.2.17).
`url`	string	URL of the content, for buy-side contextualization or review.
`cat`	string array	Array of IAB content categories that describe the content producer. Refer to List 5.1.
`prodq`	integer	Production quality. Refer to List 5.13.
`videoquality`	integer; DEPRECATED	*Note: Deprecated in favor of `prodq`.* Video quality. Refer to List 5.13.
`context`	integer	Type of content (game, video, text, etc.). Refer to List 5.18.
`contentrating`	string	Content rating (e.g., MPAA).
`userrating`	string	User rating of the content (e.g., number of stars, likes, etc.).
`qagmediarating`	integer	Media rating per IQG guidelines. Refer to List 5.19.
`keywords`	string	Comma separated list of keywords describing the content.
`livestream`	integer	0 = not live, 1 = content is live (e.g., stream, live blog).
`sourcerelationship`	integer	0 = indirect, 1 = direct.
`len`	integer	Length of content in seconds; appropriate for video or audio.
`language`	string	Content language using ISO-639-1-alpha-2.
`embeddable`	integer	Indicator of whether or not the content is embeddable (e.g., an embeddable video player), where 0 = no, 1 = yes.
`data`	object array	Additional content data. Each `Data` object (Section 3.2.21) represents a different data source.
`ext`	object	Placeholder for exchange-specific extensions to OpenRTB.

However, in practice, the majority of buy side decisioning relies on signals from third party services that specialize in content categorization via semantic analysis. Examples of these services include MOAT, Integral Ad Science, DoubleVerify, Grapeshot, and more. They are often used in lieu of publisher provided contextual signals because they are considered to be more reliable and objective, given the inconsistency in application of taxonomy IDs across publisher groups, as well as inherent publisher incentives to misrepresent content descriptions to improve perceived value / monetization options.

In this scenario, context signals can use a similar mapping to object values as within the Audience attribute example, however differentiate the context signal from the audience signal by hanging the data object off of the Content Object (vs. the User Object in the case of audience signaling):

Open RTB 2.5		**Data Transparency Standard 1.0**			
Object	*Description*	*DTS Field Name*	*Field Representation*	*Field Options*	*Description*
Content.data	Exchange-specific name for the data provider	**Provider Name**	**Provider *Domain***	String	Unique domain of the business entity making the attribute determination (context or audience)
Content.data .segment.id	Name of the data segment specific to the data provider.	**Segment Name**	**Provider Segment ID**	String (Numeric 5 character limit)	Provider's internal ID of audience segment referenced
Content.data .segment.id	String representation of the data segment value.	**Standardized Segment Name**	**Standardized Taxonomy ID**	String array	Comma separated list of the most accurate standardized IDs as selected from IAB Tech Lab's Audience Taxonomy 1.x -or- Content Taxonomy 2.x.

Summary of Primary Open Questions

There are several open questions pertaining to these approaches to audience and content signaling for the working group to consider, which are summarized below. The Rearc Addressability working group will continue to deliberate these questions during the RFC period.

Appendix

Data Transparency Standard 1.0

Section	Field Name	Field Options	Format Requirements	Description
Data Summary	**Provider Name**	Free text	Alpha-numeric: 100 characters	Name of the business entity selling the data.
	Provider Contact Info	Free text	Alpha-numeric: 100 characters	Email address where provider can field inquiries about segment
	Segment Name	Free Text	Alpha-numeric: 100 characters	Provider's descriptive name of audience attribute contained in segment
	Standardized Segment Name*	Free text *Tier 1, 2, and "final" Tier of Taxonomy naming convention is required to be displayed.*	Alpha-numeric: 100 characters	Declaration of the most accurate standardized name as selected from IAB Audience Taxonomy 1.0 [**LINK**].
	Segmentation Criteria	Free text	Alpha-numeric: 500 characters	Description of the rules applied by the seller that govern inclusion of data points into the online audience segment. Sellers may wish to include provenance, recency, and frequency logic, as well as core differentiating factors that a buyer may want to evaluate during purchase decision
	Audience Precision Level	Individual Household Business Device ID Browser Geography	Multi-select: Dropdown	The level of granularity for audience composition
	ID Count	Free text	Alpha-numeric: 15 characters	The number of IDs contained in the segment.
	ID Type(s)	Cookie ID Mobile ID Platform ID	Multi-Select: Dropdown	The currency of activation IDs
	Geography**	Select from: ISO-3166-1-alpha-3	Multi-Select: Dropdown	Geographies associated with the coverage of the segment.
	Privacy Policy	Free text	Alpha-numeric: 100 characters	Hyperlink to the seller's privacy policy

Audience Details	**Data Source(s)*** **	App Behavior App Usage Web Usage Geo Location Email TV OTT or STB Device Online Ecommerce Credit Data Loyalty Card Transaction Online Survey Offline Survey*** Public Record: Census*** Public Record: Voter File*** Public Record: Other*** Offline Transaction***	Multi-Select: Dropdown	Origin of the raw data used to compile the audience
	Data Inclusion Methodology	Observed/Known Declared Inferred Derived Modeled****	Multi-Select: Dropdown	Description of seller's relationship to the audience attribute / information being sold: **Observed / Known** - The underlying audience attributes are directly observed **Declared** - The underlying audience attributes are self-reported by the audience members **Derived** - The underlying audience attributes are computed based on other known or declared fields on record **Inferred** - The underlying audience attributes are determined from business rules or logic **Modeled** - The underlying audience attributes are calculated using an algorithm, with a seed as the source
	Audience Expansion ****	Yes No	Single-Select: Dropdown	Was look-a-like modeling used to include “similar” IDs?
	Cross-device Expansion	Yes No	Single-Select: Dropdown	Was the segment expanded to include IDs thought to be associated with the devices of the same user, household, or business?
	Audience Refresh Cadence	Intra-day Daily Weekly Monthly Bi-Monthly Quarterly Bi-Annually Annually	Single-select: Dropdown	Cadence of audience refresh
	Source Lookback Window	Intra-day Daily Weekly Monthly Bi-Monthly Quarterly Bi-Annually Annually	Single-select: Dropdown	Period in the past that a qualifying event can occur for inclusion in audience
Onboarder Details* **	**Input ID / Match Key**	Name Address Email Postal / Geographic Code Lat / Long Email Mobile ID Cookie ID IP Address Customer ID Phone Number N/A	Multi-Select: Dropdown	Input ID/ Match Key used by the Onboarder for matching

	Audience Expansion	Yes No N/A	Single-Select: Dropdown	Was look-a-like modeling used to include "similar" IDs before the data was matched to a digital identifier?
	Cross Device Expansion	Yes No N/A	Single-Select: Dropdown	Was the audience expanded to include affiliated devices and IDs before the data was matched to a digital identifier?
	Audience Precision Level	Individual Household Geography N/A	Single-Select: Dropdown	What is the precision level of the data before it was matched to a digital identifier?

* **Standardized Name:** See IAB Tech Lab Audience Taxonomy 1.1 found on IAB Tech Lab's website

** **Geography:** see standardized country codes found within ISO-3166-1-alpha-3

*** **Data Sources:** selection of "offline" sources indicated necessitates completion of "Onboarder Details" section

**** **Data Inclusion Methodology Audience Expansion:** Selection of "Modeling" requires selection of "Yes" within "Audience Expansion" field