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

Applying seller-defined audience and content taxonomy IDs and Data Transparency Standard metadata within OpenRTB to support privacy-centric addressability and first-party data monetization

v1.0

Published February 2022

This document has been developed by the Rearc Addressability Working Group.

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 groups. 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 – Privacy, Identity & Data
IAB Tech Lab

Questions 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, Audience Taxonomy 1.x, or proprietary taxonomies - to signal seller defined contexts or audience attributes within OpenRTB. *Referenced hereafter as Seller Defined Audiences or "SDA",* this 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. It 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

The SDA approach is based on several design principles and constraints:

1. **User Transparency and Control** - needs to support regional expectations around consumer transparency and control of personal data.
2. **Data Security and Minimization** - *should not* rely on the passing of user specific data currencies that have historically been used for non-transparent or non-permissioned profile development - including third-party cookies, mobile / OS IDs, user-provided IDs, or user-agent information – and *should* establish expectations to restrict the commingling of data types that could pose privacy risks without sufficient data protections.
3. **Technical Accountability to Consumer Preferences** - 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** - needs to minimize disruption to existing business models and competitive dynamics. It should not rely on complex and untested 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** - *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** - 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.

Additional Reading and Referenced Documents

- IAB Tech Lab - Content Taxonomy 3.0
- IAB Tech Lab - Audience Taxonomy 1.1
- IAB Tech Lab - Data Transparency Standard (DTS) 1.1
- IAB Tech Lab - OpenRTB 2.6
- IAB Tech Lab – Transparency Center Data Set
- 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
- “Wishful Thinking, Meet Pragmatic Planning: A Portfolio Approach To Addressability”, Anthony Katsur (CEO, IAB Tech Lab), AdExchanger, 1/11/22

Background

Today, online advertising systems rely on algorithms to group information associated with cross-site/app identifiers (e.g., third-party cookies and mobile IDs) into audience segments. Sometimes these groupings rely on declared information, such as registration, and sometimes they’re based on observed browsing behavior. Marketers and publishers work together to identify these audiences and match advertising to individuals who they believe to be most likely to engage with their brand. Publishers benefit because they can more effectively monetize their content, while users tend to benefit by seeing more relevant and less intrusive ads.

This dynamic is made possible by the maintenance of audience profiles across apps and page domains, largely via cookies and mobile IDs. Cross-site tracking 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 increased scrutiny over the past several years has resulted in device and OS manufacturers’ removing industry participants’ access to data currencies that could be used for cross-context identification, including third party cookies and mobile

IDs, metadata that supports statistical IDs, and unobscured email addresses used by many to “log in” to personalized content experiences.

This disruption to underlying currencies that support open market addressability has led many to consider how attributes relevant to marketers can be measured by first parties, anonymized, and communicated via open standards like OpenRTB without a dependency on deprecated data currencies or centralized browser mediation. The following document summarizes industry consensus on the most expedient approach to accomplish this, associated design considerations, intended benefits and possible pitfalls, and integration requirements for publishers, SSPs and DSPs. The approach is intended to meaningfully address marketer and publisher concerns around audience addressability within the open ecosystem, as well as underlying privacy and security concerns of sharing data with non-permissioned recipients, but makes no claim to be a universal solution that satisfies all business use cases. As such, it is intended to be a helpful tool within a portfolio approach to addressability.

Updates to Standards Needed to Support SDA

Updates were made to existing Tech Lab standards in order to support the Seller Defined Audiences approach. These updates are based on industry consensus driven within relevant IAB Tech Lab working groups, including the Taxonomy and Mapping working group, Programmatic Supply Chain working group, Rearc Addressability working group, and Data Transparency Standards working group. They include:

OpenRTB Community Extensions
Given limited adoption of the AdCom / OpenRTB 3.0 specification, SDA assumes the use of OpenRTB 2.x.

- IAB Audience Taxonomy was developed *after* the most recent version of OpenRTB 2.x. Guidance was added to the OpenRTB community extensions Github repo regarding how to structure in-band Audience Taxonomy IDs and Data Transparency Standard metadata. Details are described below.
- While IAB Content Taxonomy was already referenced within the existing OpenRTB 2.x object model, additional guidance has been added to the OpenRTB community extensions regarding how to convey content or context signals to support the SDA approach. Additional details use cases for Content Taxonomy appear at the bottom of this document.

Modifications to Data Transparency Standard (DTS) Fields

- The Data Transparency Standard was updated from 1.0 to 1.1 to account for additional privacy compliance fields, and other fields necessary to associate in

stream SDA signals with accompanying metadata housed within IAB Tech Lab's DTS metadata repository (see below). The updated Data Transparency Standard 1.1 specification can be found in the appendix.

Modifications to Repository of DTS Metadata (formerly Datalabel.org)

- The industry database of DTS audience metadata, formerly housed at datalabel.org, was retooled to account for changes to the DTS schema moving from 1.0 to 1.1
- This DTS data set was merged / bundled with IAB Tech Lab's newly launched Transparency Center, which makes available to industry participants a broad suite of metadata aimed at helping the buy and sell side transact with confidence. Specifically, it aggregates and provides API access to structured data sets about ad tech businesses in the areas of industry compliance, business identifiers, and supply chain attributes. DTS metadata is now one of these data sets.
- DTS metadata APIs were updated to reflect new endpoints and DTS schema. Documentation can be found here:
    - Upload API: documentation and endpoints (note, requires Tech Lab Tools Portal login). This allows adopters of the DTS standard to upload their audience metadata labeling for use by the marketplace. See appendix below for more information about mapping DTS and API field names.
    - Retrieval API: documentation (note, requires Tech Lab Tools Portal login). This allows the marketplace to retrieve DTS metadata from the repository within Transparency Center for subsequent analysis.

Approach: Passing Taxonomy IDs and Metadata to Support Cohort 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.x, and the Transparency Center industry metadata repository.

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) 1.1 is a standardized schema of over 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 1.1, the 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 incentive for providers to misrepresent their data to buyers, IAB Tech Lab developed 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 can be found at www.datalabel.org. Additionally, the full DTS 1.1 schema and required fields can be found in the appendix below.

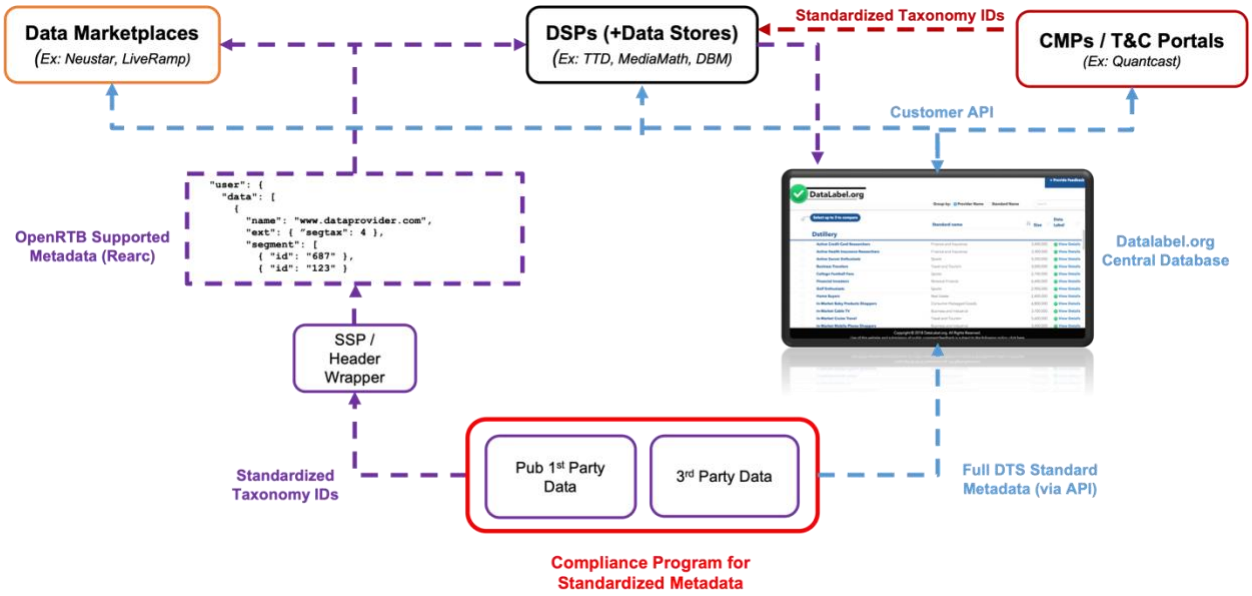
IAB Tech Lab's Transparency Center is an industry resource for, among other things, DTS data labels produced by data providers that support the Data Transparency

Standard. These labels are ingested directly from participating data providers via API. The tool provides a centralized location and UI for Tech Lab members to search and discover DTS metadata before making a purchase decision, and also allows for Transparency Center subscribers to retrieve the metadata via API. Importantly, DTS metadata 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. More information about the Transparency Center can be found here.

Application of Existing Tools Within OpenRTB

By leveraging these already-adopted specifications in new ways, SDA establishes that publishers or their data partners 1) determine audience attributes based on customer interactions on their properties, 2) map those attributes to standardized taxonomy descriptions and data transparency disclosures, and 3) relay those anonymized taxonomy IDs within OpenRTB to inform downstream signaling by buyers.

By using a bespoke integration with prebid.js, this can be done at scale, without a reliance on deprecated data currencies, and in a way that provides meaningful differentiation of and competition within seller defined audiences. Data flows between publisher/attribute provider, OpenRTB, and Transparency Center can be envisioned as follows:



The following sections provide an overview of these mechanics.

Using Segtax Extension to Pass Audience Cohort Signals via OpenRTB

IAB Tech Lab's Data Transparency Standard is a well-suited schema to provide differentiation, compositional transparency and consistent industry lexicon for efficient decision making within a cohort marketplace. Moreover, the industry repository of data labels within Tech Lab's Transparency Center is an effective, centralized tool to automate delivery of cohort metadata via API integrations.

While the full set of data label fields can not realistically be conveyed via OpenRTB in real-time given payload limitations, they could be retrieved out-of-band from the Transparency Center repository should the publisher relay a small, unique set of data points from the label that downstream buyers could use to identify a specific audience label within the broader Transparency Center repository. This metadata portrait can then be used to facilitate improved decision-making based on data provenance, modeling, age, or other characteristics relevant to a buyer's bidding decision. To reduce resource requirements, the full universe of label metadata from the repository can be saved and referenced locally by buying platforms in advance of bidding.

In order for buyers to retrieve unique metadata and avoid label collision, its suggested that the following data points should be relayed within the "Segment Taxonomies" community extension within OpenRTB (aka "segtax"). These values are a) relatively unique data points in combination, and thus well suited to identify the full DTS label within Transparency Center, and b) provide lightweight signals within bid requests:

- **Provider Name** - the unique domain of the entity making the attribute / cohort determination
- **Segment ID (s)** - the provider's declaration of the ID(s) that best describe its internal segmentation
- **Taxonomy Name –** the taxonomy range in which the Segment IDs / values can be found. This taxonomy range can refer to standardized taxonomies like IAB Tech Lab's Content Taxonomy or Audience Taxonomy, or it could refer to custom or internal taxonomies (more below).

"Taxonomy Name" and associated ID values are maintained in an enumerated list within Tech Lab's community extension repo. The following are current reserved values to support SDA signaling:

ID Value	Description
1	*IAB Tech Lab Content Category Taxonomy 1.0. (Deprecated)*
2	*IAB Tech Lab Content Category Taxonomy 2.0 (Deprecated)*

3	IAB Tech Lab Ad Product Taxonomy 1.0.
4	IAB Tech Lab Audience Taxonomy 1.1
5	IAB Tech Lab Content Taxonomy 2.1
6	IAB Tech Lab Content Taxonomy 2.2
7	IAB Tech Lab Content Taxonomy 3.0
500+	Vendor-specific codes.

Leveraging Standard or Proprietary Taxonomies

This taxonomy list supports both standard taxonomies and vendor-specific taxonomies, with ID values of 500+ being reserved for the later. Because cohort providers often use multiple internal taxonomies to organize audience attributes, it is valuable to enable flexibility to specify which internal taxonomy the provider's segment name / IDs refers to. Private taxonomy attribute values are inherently unknown to the broader marketplace, and thus limited in scale, given that private taxonomy values would only be recognizable and actionable by those entities that have a direct integration with the signaling party.

All industry participants can request vendor-specific ID ranges:

- Pull requests (PRs) can be submitted by anyone on an ongoing basis within the Segtax community extensions repo.
- Tech Lab working group leads will regularly review + approve pull requests
- The PR submitter must notify the working group leads of the PR submission. Tech Lab working group leads will approve pending ID conflicts and the number of IDs requested.

JSON Example of Audience Cohort Signaling

Existing objects within OpenRTB 2.6 that are well suited to support Taxonomy IDs signaling: the User, Data, and Segment objects.

User Object: intended to contain information known about the user of a device.

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.

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.

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.

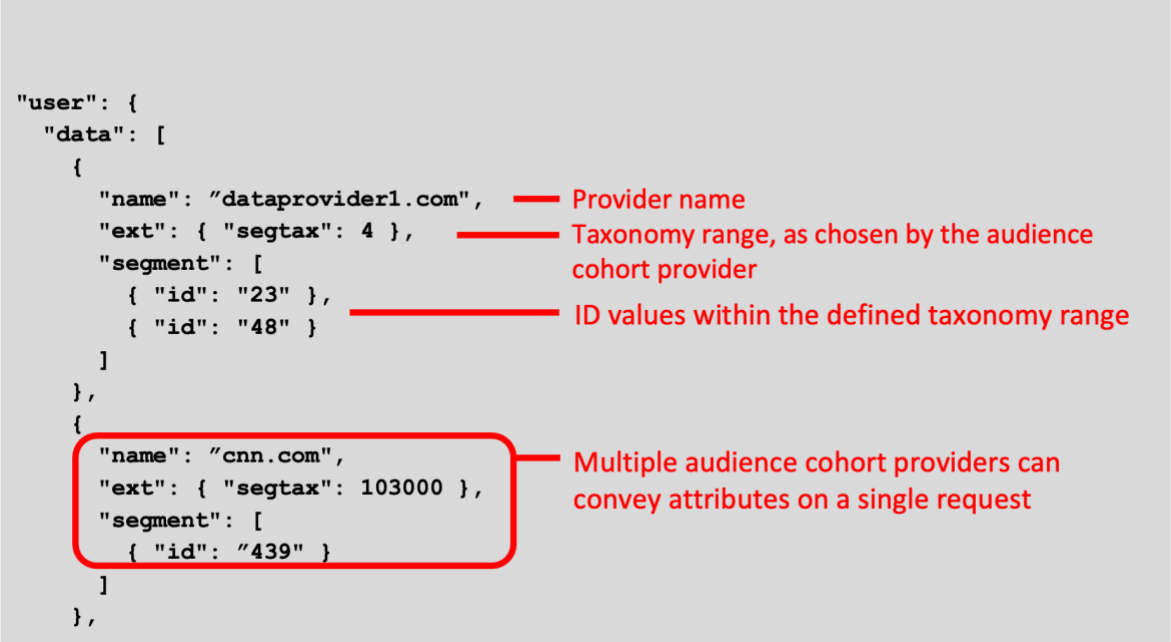
Segment Object: key-value pairs that convey specific units of data. The parent Data object is a collection of such values from a given data provider.

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.

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

- *Multiple Cohort Providers* – publishers that work with multiple DMP/data providers can convey multiple attribute signals per request
- *Multiple Cohort Taxonomies* - cohort providers have flexibility to define different taxonomies (proprietary or standardized) that IDs might be associated with

Below is an example of how the JSON would be structured using "segtax" extension:

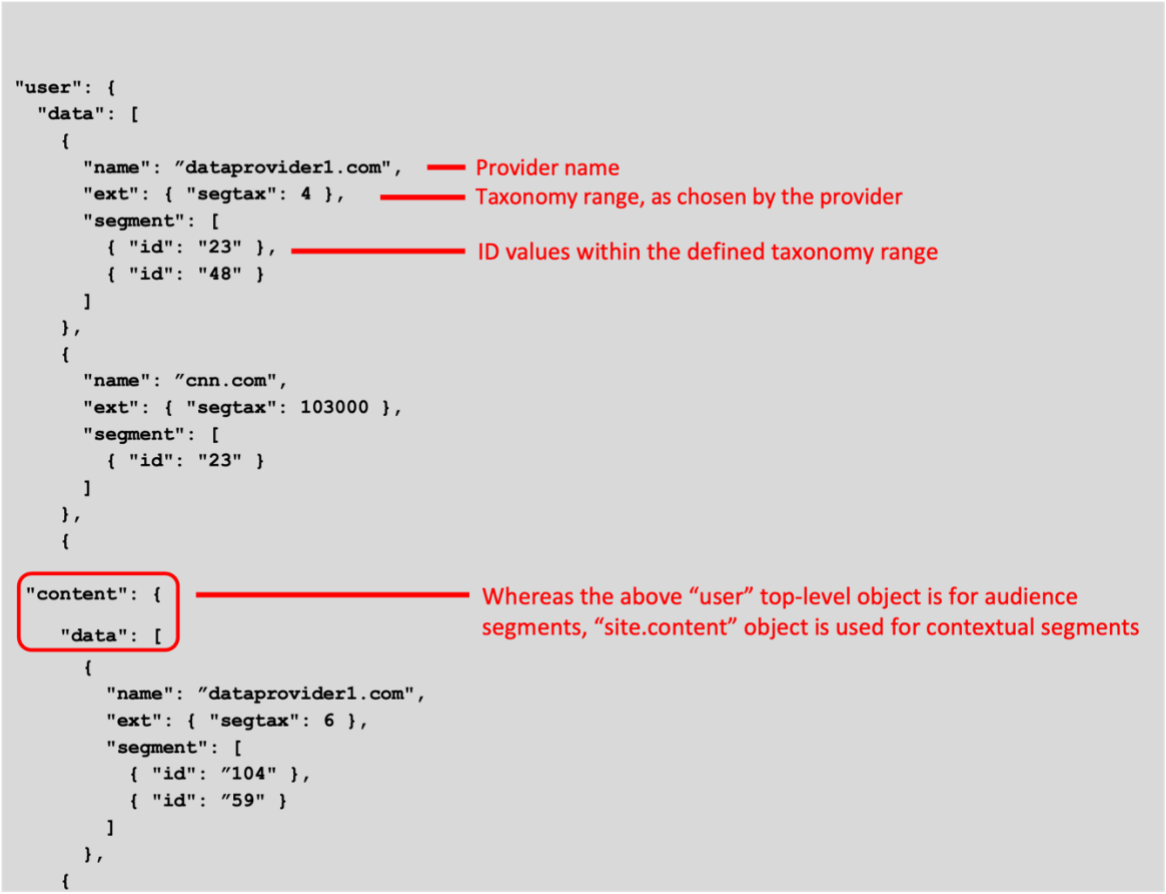


Using Segtax Extension to Pass Context Signals via OpenRTB

IAB Tech Lab 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.

However, in practice, the majority of buy side decisioning relies on signals from third party services that specialize in content categorization via semantic analysis. 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.

Regardless, in this scenario, context signals can use a similar mapping to object values as within the Audience attribute example above, and simply 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):



Using Extended Content Identifier Extension to Pass Content Signals via OpenRTB

While IAB Tech Lab does not manage a taxonomy that standardizes descriptions of episodic content, there are companies/services which act as a clearinghouse or aggregator of such metadata from publishers. These services ingest video content metadata from publishers and assign an ID for each piece of content that is unique within that content data platform. There is a market need to target by video/audio content metadata or content classifications thereof. In such a scenario, the "Extended Content Identifier" OpenRTB community extension should be used, which is a variation of the Segment Taxonomy extension used above.

Below is an example of how the JSON would be structured using the Extended Content Identifier community extension:

```
{
 "site": {
   "content": {
     "data": [
        {
             "name": "iris.tv",
              "ext": {
                 "cids": [
                      "iris_c73g5jq96mwso4d8"
                  ]
               }
          }
        ]
     }
  }
}
```

Prebid Integrations

Integration with Prebid.js is a prerequisite to adopt SDA signaling. The basic workflow is as follows:

1. Publishers declare their first party data using Prebid setConfig("ortb2") and set the attributes that are pertinent per the guidance above. Additionally, publisher's on-page partners can set/augment "ortb2" values via Prebid RTD Module.
2. SSPs will translate publisher FPD to bid requests
3. DSPs will bid on the new bid requests and notify the SSP whether a bid was based on FPD data

Below are high level requirements for publishers, SSPs, and DSPs. Additional questions can be sent to addressability@iabtechlab.com.

Publishers

As the first step in the data flow, publishers or on-page data assemblers set first-party data to Prebid's "ortb2" object via setConfig or as automated by a Prebid RTD Module. The relevant input for *audience taxonomy* signals is defined in the table below. There can be multiple of these 3-point declarations in a single config.

Input	Description	Data Type	Example	Required?
ortb2.user.data.name	Creator of the segment	String	mydmp.com	yes
ortb2.user.data.ext.segtax	Reference to the taxonomy being used.	int	1	yes
ortb2.user.data.segment.id	Object containing the segment ids the user is a member of	String	1	yes

The relevant input for *content taxonomy* signals is defined in the table below:

Input	Description	Data Type	Example	Required?
ortb2.site.content.data.name	Creator of the context classification	String	mydmp.com	yes
ortb2.site.content.data.ext.segtax	Reference to the content taxonomy being used.	int	1	yes
ortb2.site.content.data.segment.id	Object containing the segment ids the user is a member of	String	1	yes

The relevant input for *content identifier* signals (see "…Extended Content Identifier Extension …") is defined in the table below:

Input	Description	Data Type	Example	Required?
ortb2.site.content.data.name	Creator of content classification	String	mydmp.com	yes
ortb2.site.content.data.ext.cids	An array of content IDs, representing one or more identifiers for the video or audio content from the source specified in the "name" field of the "data" object	String	1	yes

On-Page Implementation

setConfig

See additional guidance on Prebid page.

```
pbjs.setConfig({
        ortb2: {
          "user": {
            "data": [
              {
                "name": "hearst.com”,
                "ext": { "segtax": 1 },
                "segment": [
                  { "id": “1001” },
                  { "id": “1002” }
                ]
              }
            ]
          },
          "site": {
            "content": {
              "data": [
                {
                  "name": "cnn.com",
                  "ext": { "segtax": 2 },
                  "segment": [
                    { "id": “2002” }
                  ]
                }
              ]
            }
          }
      });
```

RTD Module

```
"user": {
  "data": [
    {
      "name": "hearst.com",
      "ext": { "segtax": 1 },
      "segment": [
        { "id": “1001” },
        { "id": “1002” }
      ]
    }
  ]
},
"site": {
  "content": {
    "data": [
      {
        "name": "cnn.com",
        "ext": { "segtax": 2 },
        "segment": [
          { "id": “2002” }
        ]
      }
    ]
  }
}
```

SSPs

Publishers or on-page data assemblers set data to Prebid's "ortb2" object via setConfig or as automated via a Prebid RTD Module. With some data set to Prebid.js, the SSP can do its part. For SSPs, there are two stages:

1. Adapter - resolve the data from the Prebid "ortb2" object
2. Exchange - transmit the data into the bidstream applying the same ortb2 fields

Adapter (buildRequests):
Participating adapters must read segment values from the first party data object. To get the FPD object: config.getConfig('ortb2'))

Exchange
SDA establishes a consistent method to communicate segment information with demand. Requests to DSPs are expected to be devoid of any user id / cookie / IP / deal ID information, and should use the ORTB user.data or site.content.data objects.

DSPs

At the demand step in the data flow, DSPs will respond to SSPs based on values within the user.data and site.content.data objects in the bid request. Prior to the DSP data flow, publishers will make data available via Prebid.js so that SSPs can interpret and act. Once the SSP has resolved the data and made it available in the bidstream, DSPs can:

1. Evaluate bid opportunities based on the user.data and site.content.data objects within the SSP's bid request
2. Submit bid responses (standardized formatting is proposed but not final)
3. Log won impression record (standardized formatting is proposed but not final)

Data Labeling Integrations

Publishers, or their data partners, are expected to maintain metadata associated with their audiences - formatted per the IAB Tech Lab Data Transparency Standard - and make that metadata available to DSPs via API integrations with IAB Tech Lab's Transparency Center.

Publishers and Data Providers

Audience metadata should be stored and formatted based on the most recent Data Transparency Standard schema (see appendix), and uploaded to IAB Tech Lab's Transparency Center via API. Here is relevant API documentation and endpoints to upload (note, requires Tech Lab Tools Portal login). Mappings between DTS field names and API field names can also be found in the appendix.

DSPs

DSPs and other interested parties can retrieve audience metadata from the repository within Transparency Center via a "Retrieval API". Here is relevant documentation (note, requires Tech Lab Tools Portal login). The full data set - across all publishers and data providers that have uploaded metadata - can be pulled down for analysis / interpretation in advance of bid requests from supporting inventory sources.

Other Key Considerations

Data Security 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 supply chain participants to minimize commingling of cohort data with other sensitive data types
- Accountability of cohort developers to "sufficiently large" cohort threshold

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

Accountability of cohort developers to the accuracy of self-attested labeling

Cohorts and standardized Transparency Center 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 "games enthusiasts" 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 "games enthusiasts" 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 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 concerns 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 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 – which includes things like user agent information, first party identifiers, probabilistic maps of various IDs, and 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 identifier is a business decision that first parties should control. However, unnecessary commingling of data points should always be avoided. The entities best positioned to validate and curate unnecessary commingling of bidstream data are:

- *Header Wrappers* - these are entities that facilitate unified auctions across multiple exchanges. They are a widely used 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” Servers* - the concept of a trusted server has become a fixture in browser standards conversations. It refers to an external server - usually operated by an organization that does not buy or sell media - that would work with the browser to, among other things, evaluate / anonymize incoming bid requests and output differentially private signals for the ad ecosystem to react to. There are various active proposals on trusted server implementations being debated within industry forums, some of which can be found linked in this document within “Additional Reading and Referenced Documents”

After weighing enforcement scenarios and technologies against factors like ease of

adoption, technical enforceability, and legal complexity, IAB Tech Lab sees independent Trusted Server implementations as the least disruptive and most effective path towards sustainable privacy outcomes that also protect core addressability use cases. However, with the advantaged position of such a proxy service – sitting between a browser and the rest of the ad tech ecosystem - and its potential to disintermediate many downstream entities, its important that any trusted server solution is built and governed based on several core principles to ensure marketplace transparency, trust, and adoption:

- Demonstratable and consistent consumer privacy protections
- Transparent, democratic governance structure that allows for:
    - Equal representation across browser / ad tech / and publisher ecosystem, and other key stakeholders
    - Consistent, well documented processes and decision making
- Backwards compatibility with existing ad tech infrastructure
- Technical design that facilitates open market growth, innovation, and competition
- Supply chain interoperability that supports meaningful addressability, measurement, and other core business use cases

IAB Tech Lab will continue to evaluate specific trusted server implementations via its Privacy Enhancing Technologies Working Group, and make recommendations based on these foundational requirements.

Accountability of cohort developers to a “sufficiently large” cohort threshold

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. First parties or their technology partners 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).

This figure can be used as a directional starting point for industry participants, but can’t be generalized across all first party data sets for a number of reasons: 1) the high

variability in counting methodologies, 2) issues of fairness based on the size of properties and relative visibility of audiences, and 3) fluctuations in re-identification risks based on attribute sensitivity, granularity, and audience precision / intersectionality.

The issue of counting methodology and fairness materializes in several ways. First, there will always be variability in this figure based on the concept of "*unique users*" / average device counts per person. Second, it will also fluctuate based on the concept of a "*lookback window*" established by the first party, which defines the amount of time in the past that devices can be counted within a cohort. Lastly, a static threshold ignores the possibility that a minimum size within any audience segment would *disproportionately impact smaller publishers and brands* because it would take longer to satisfy the threshold benchmark for any given audience.

Privacy and re-identification risks variables that affect "sufficiently large" size thresholds are more pernicious and materialize in other ways. Variability of this threshold will often be driven by the sensitivity of the attribute the cohort is assigned (ie, beliefs, opinions, orientations). Additionally, the granularity of the audience – whether its representative of individuals, households, geographies, etc - will impact reidentification risks. Lastly, it is important to account for the precision of the cohort relative to the size of the entire universe of that audience. A cohort that represents a sizeable % of the total universe puts individuals in that universe at higher identification risk. This becomes amplified when there's intersectionality of cohort designations or other identifiable data on the same impression.

For all of these reasons, to develop static and technically unenforceable guidance for minimum cohort sizes is impractical for cohort developers (especially given that few have a core competency in necessary privacy enhancing technologies). Instead, these cohort developers are best suited by leaning into external ML systems like Trusted Servers, described above, that are specifically designed to handle ecosystem-wide ad request anonymization, noise, and differentially private signaling on behalf of ad tech. As with the "data commingling" problem described above, IAB Tech Lab sees independent Trusted Server implementations as the most expedient path to effectively enforce limits on the identifiability of user cohort information sent in ad requests.

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 optimize in and out of

cohorts 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.

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 additional extensibility of the Transparency Center 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.

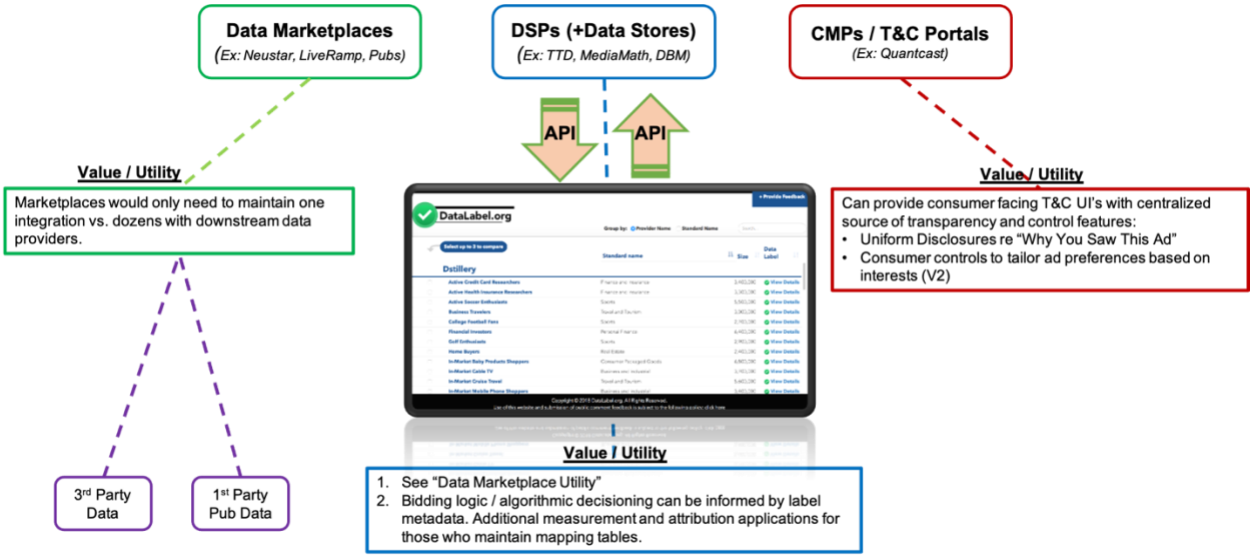
Connective Tissue to Support Consumer Facing Disclosure Information

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 in Transparency Center 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 Transparency Center 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 creative 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.



Appendix

Data Transparency Standard 1.1

See datalabel.org for more information.

Section	API Flat File Key	Field Name	Field Options	Format Requirements	Description
Data Summary	provider_name	Provider Name	Free text	Alpha-numeric: 100 characters	Name of the business entity making the attribute determination.
	provider_domain	Provider Domain	Valid domain	Alpha-numeric: 100 characters	Domain associated with the business entity making the attribute determination.
	provider_email	Provider Contact Email	Free text	Alpha-numeric: 100 characters	Email address where provider can field inquiries about segment / cohort
	audience_name	Provider's Audience Name	Free Text	Alpha-numeric: 100 characters	Provider's descriptive name of audience attribute contained in the segment / cohort
	audience_id	Provider's Audience ID	Free Text	Numeric: 15 characters	Audience segment's unique internal ID as specified by the Provider
	taxonomy_id_list	Standardized Audience ID(s)*	Select from: IAB Tech Lab Audience Taxonomy 1.1	Alpha-numeric: 100 characters	Comma separated list of the standardized IDs that, in combination, best describe audience attribute (as selected from IAB Tech Lab Audience Taxonomy 1.1. Audience taxonomy IDs with "Purchase Intent Classification" modifiers would be delimited via pipe character. Order of Audience IDs should be sequential, and order of modifiers attached to an ID should be alphabetical. Ex: "123|PIFI2|PIPV1,456|PIFI3,789|PIPV1".
	audience_criteria	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_levels	Audience Precision Level	Individual Household Business Device Browser Geography	Multi-select: Dropdown	
	audience_scope	Audience Scope *****	Single domain / App ***** Cross-domain within O&O Cross-domain outside O&O N/A (Offline)***	Alpha-numeric: 1000 characters	The contexts within which an attribute was determined.
	originating_domain	Originating Domain *****	Valid top level domain / app store URL ***** N/A (Undeclared) N/A (Cross-domain, Offline)	Alpha-numeric: 100 characters	Domain of the digital property where the audience originates
	audience_size	Audience Size	Free text	Numeric: 15 characters	Estimated count of addressable units specified within "Audience Precision Level" field.
	id_types	ID Type(s)	Cookie ID Mobile ID Platform ID User-enabled ID	Multi-Select: Dropdown	The ID currencies that were analyzed in order to determine an audience attribute.
	geocode_list	Geography**	Select from: ISO-3166-1-alpha-3	Multi-Select: Dropdown	Pipe separated list of the geographies in which the attribute data was collected.
	privacy_compliance_mechanisms	Privacy Compliance Mechanisms Used	TCF (Europe), USPrivacy, LSPA, NAI Opt Out, DAA, EDAA, DAAC, GPC, Other (Not Listed), None	Multi-Select: Dropdown	Declaration of consumer data transparency and consent tools that provider applies
	privacy_policy_url	Privacy Policy	Free text	Alpha-numeric: 100 characters	Hyperlink to the seller's privacy policy
	iab_techlab_compliant	IAB Tech Lab Compliant	Yes No		Binary declaration regarding whether an organization has completed IAB Tech Lab's Data Transparency Standards compliance audit.
Audience Details	data_sources	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***	Multi-Select: Dropdown	Origin of the raw data used to compile the audience

			Public Record: Other*** Offline Transaction***		
	audience_inclusion_methodology	**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	**Audience Expansion ******	Yes No	Single-Select: Dropdown	Was look-a-like modeling used to include "similar" IDs?
	device_expansion	**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	**Audience Refresh Cadence**	Intra-day Daily Weekly Monthly Bi-Monthly Quarterly Bi-Annually Annually	Single-select: Dropdown	Cadence of audience refresh
	lookback_window	**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* **	**onboarder_match_keys**	**Input ID / Match Key**	Name Address Email Postal / Geographic Code Lat / Long 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
	onboarder_audience_expansion	**Pre-onboarding 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?
	onboarder_device_expansion	**Pre-onboarding 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?
	onboarder_audience_precision_level	**Pre-onboarding Audience Precision Level**	Individual Household Business Geography N/A	Multi-select: Dropdown	The level of granularity to which an audience was resolved before it was onboarded.

Below are outside resources that should be referenced, and conditional requirements depending on selections:

***Standardized Name:** See IAB Tech Lab Audience Taxonomy 1.0 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
*******Audience Scope**: selection of "Single domain / App" requires addition of a valid top level domain / app store URL within the "Originating Domain" field