Data Transparency Standards

Proposal for data transparency framework and automation standards across the data supply-chain

DRAFT FOR PUBLIC COMMENT: THROUGH JULY 16, 2018

Published May, 2018

About the IAB Tech Lab

The IAB Technology Laboratory is a nonprofit research and development consortium charged with producing and helping companies implement global industry technical standards and solutions. The goal of the Tech Lab is to reduce friction associated with the digital advertising and marketing supply chain while contributing to the safe growth of an industry.

The IAB Tech Lab spearheads the development of technical standards, creates and maintains a code library to assist in rapid, cost-effective implementation of IAB standards, and establishes a test platform for companies to evaluate the compatibility of their technology solutions with IAB standards, which for 18 years have been the foundation for interoperability and profitable growth in the digital advertising supply chain.

About the Data Transparency Standards Working Group

Launched jointly by IAB Data Center of Excellence and IAB Tech Lab boards in June 2017, the Data Transparency Standards working group seeks to develop and build technical standards and associated compliance programs - in support of buyers, sellers, and vendors - aimed at making data segment "quality" consistently understandable, transparent, and comparable across vendors. The group focuses on "quality" in terms of the extent to which it accurately assigns demographic, geographic, or behavioral attributes to an identifier–not the extent to which it drives KPI performance or marketing outcomes. At the core of the group's work is an acknowledgement that any determination of quality is relative to use case, and as such focuses on understanding and explaining the underlying determinants of quality primarily associated with data collection, matching, and modeling practices. The group is co-chaired by David Justus, VP, Data Acquisition and Operations, Lotame, Dave Smith, SVP, Monetization and Yield, Pandora, and Alysia Borsa, Chief Marketing & Data Officer, Meredith.

IAB Tech Lab Contact
Benjamin Dick
Director of Product, Data
ben@iabtechlab.com

Working Group Roster

The working group is composed of 113 members representing 75 different companies:

4C Insights	IRI	Pebble Post
4INFO	Jivox	PIVnet
Acxiom	Jumpstart Automotive Media	Placed
Adacado	Kochava Inc.	PowerLinks Media
ADVR	Kr8os	Premion
AOL	Liquidus Marketing	Prohaska Consulting
Blis	LiveRamp	Purch
CBS Interactive	Lonely Planet	Resonate Insights
CDK Digital Marketing	Lotame Solutions	Reveal Mobile, Inc.
Conversant Media	Madhive	RhythmOne
Criteo	Media Japan Network	Salesforce
Cuebiq	MediaMath	Screen6
Dentsu Aegis	Meredith Corporation	Shazam
DMD Marketing Corp.	Meredith Digital	Spectrum Reach
DoubleVerify	MeritDirect	StartApp
Drawbridge	MightyHive	Swoop
Epsilon	NBCUniversal	Synacor
Extreme Reach	NinthDecimal	Teads
Eyeota	NYIAX, Inc.	TenMax
FreeWheel	Oath	The Trade Desk
Fyber	OpenX	Turn Inc.
Grabit Interactive Media	Oracle	Weborama
IBM Watson Advertising	OwnerIQ	Xaxis
Index Exchange	Pandora	YourTango
IPONWEB	Parrable	

Table of Contents

Overview

The following proposed standards establish: 1) a baseline expectation for data sellers regarding the additional information that buyers need to make an informed purchase decision, 2) a standardized audience taxonomy so buyers can compare like segments across vendors, and 3) software to deliver this data to buyers in their platform of choice via an open source API. It does so in accordance with the operational and technical realities of segment development and generation, within a flexible framework that can evolve over time based on marketplace needs.

Note that this document assumes a baseline conversancy with data sourcing and segmentation concepts. For additional glossary terms, concepts and definitions please refer to IAB's Data Techniques and Lexicon Primer (2016).

Proposed Solution

The framework is made up of four component parts. *This document makes available the first two components for public comment–the labeling requirements and audience taxonomy–while the latter two supporting components will be introduced into the market by the end of 2018.* The four components are:

- A set of ID-level labeling requirements for data sellers
- A standardized audience taxonomy / data segment naming convention that's incorporated as one of the labeling requirements
- Supporting open-source API intended to structure and communicate machine-readable label data between supply chain participants
- Supporting compliance programs to acknowledge transparent data sellers

The labeling requirements and audience taxonomy will be available for 60-day public comment through **Monday July 16, 2018.** The labeling requirements can be reviewed in the document that follows, while the associated Audience Taxonomy can be found in a companion document here. Please submit written feedback to ben@iabtechlab.com.

Intended Benefits

When scaled and implemented across the ecosystem, the standards and supporting software would introduce ID-level metadata into the ecosystem for the first time. Very much akin to how OpenRTB largely shifted media buying philosophies from buying publishers as proxies for audiences toward buying individual impressions informed by data, ID-level visibility is expected to refocus data buying from broad audience segments to buying individual ID-attribute pairs informed by metadata.

The framework is expected to not only allow buyers to easily calculate segment composition against data provenance, age and ultimately determine segment applicability at the point of purchase, but also have numerous positive downstream effects, including:

- Improved efficiency and liquidity of the data marketplace
- A shift of human capital to higher value tasks made possible by automated segment composition analysis
- More efficient, effective digital ad campaigns and improved consumer ad experiences
- Growth and differentiation within the data industry at large by removing perceptions of data commoditization and encouraging price premiums / incremental investment for rigorous and accurately defined data segments that play an outsized role in delivering marketing outcomes
- Transformation of metadata into a machine-readable format that will:
  - Facilitate use of calculated fields to understand segment composition and applicability
  - Allow bidding platforms to learn over time which signals have more or less of an impact on KPI performance and naturally skew investment into those higher performing areas
  - Enable more sophisticated application of audience data for campaign efficacy and attribution analysis via application of robust metadata signals

Background

Segmented audience data is ubiquitous in digital media buying and selling, accelerated in recent years by the rise of exchange-based marketplaces and the ability to evaluate and bid on ad impression opportunities in real time. Advertisers, agencies, and publisher audience products frequently rely on audience data to inform purchasing and bidding decisions against consumer demographics, behavioral qualifiers or purchase intent signals. It's also used to perform advanced modelling and analytics to better understand the efficacy of their media investment. This activity has fueled a marketplace now worth an estimated $10 billion.

A large and interconnected group of online and offline data collectors, aggregators, onboarders and modelers has organically grown to support the rising marketplace need for audience addressability within digital media channels. The infrastructure and data supply chain established by these technologies supports dozens of branded data solutions that compete against each other for buyers' attention within robust audience marketplaces–largely accessed within ad serving platforms, DSPs, and DMPs made up of thousands of segments that are searchable by seller, data type, or industry / vertical.

Audience segments are generally accompanied by a descriptive name like "Auto Intender", "Expecting Mother", or "Males 18-24", and include information about the number of unique identifiers that match the description, as well as a corresponding price quoted as either a CPM or percent of media cost. After a segment is selected, it is then "pushed" to and / or applied within a buying platform for inclusion as a targeting parameter, often accompanied by other criteria that buying platforms make available like frequency caps, recency windows, and more. This process of moving data between platform supply chain participants is automated via API integrations.

Although audience attributes are frequently applied as a targeting parameter, it is often done without a thorough understanding by the buyer regarding the extent to which the segment description–like "Auto

Intender"–actually reflects individuals or households that are in the market for a car, how the segment accuracy could be affected by the business practices or technical capabilities of the seller, or the extent to which business rules/logic that govern ID inclusion might differ as data is passed from one supply chain participant to the next. Without a consistent and flexible approach to data organization and labeling throughout the supply chain, the possibility of attribute misclassification increases, producing more intrusive consumer ad experiences, less efficient advertising investment, and diminished monetization opportunities for publishers.

Marketplace Needs

IAB Tech Lab's proposed labeling solution and supporting software attempts to address several fundamental marketplace needs upon its initial release, with flexibility for additional spec features and functionality available in future versions based on marketplace needs. This additional functionality is especially important as the marketplace adapts to and develops consistent mechanisms to capture and communicate consumer consent signals for data collection and processing.

- 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 an ID to be included in a segment?
- Data Taxonomy: when can one data segment be evaluated against another like segment?

While the framework seeks to establish and enforce seller transparency within these areas, it DOES NOT seek to address segment efficacy or accuracy. Specifically it does not establish: 1) a point of view or definition of what "quality" segment data means, given that the notion of quality is subjective and relative to use case, 2) a way of vetting the extent to which segment descriptions reflect audience attributes of users (i.e., the extent to which an "Auto-intender" segment reflects individuals who are in the market for a car), or 3) a mechanism for buyers to evaluate the extent to which audience data will impact KPI performance or marketing outcomes. These functions represent opportunities for individual marketplace participants and / or trade group guidance.

While addressing these needs, the framework also provides a scalable model to understand the contributions of all supply chain actors, and to allow for the analysis of that data at each step in the chain, including by the ultimate user / consumer / buyer of the data.

Transparency Compliance

IAB Tech Lab's compliance initiative–to be rolled out in Q3 2018–will seek to establish that sellers comply with three baseline expectations outlined in the framework:
1. Sourcing of properly formatted qualitative and quantitative metadata labels

2. Use of IAB Tech Lab's standardized audience taxonomy
3. Adoption of IAB Tech Lab's open source API that is intended to facilitate automated delivery of ID-level metadata labels to buyers at the point of purchase within UIs (**to be released**)

As mentioned above, the compliance program is intended to showcase that label information is provided by sellers, is properly formatted, and delivered to buyers in a meaningful / automated way. It is out of scope for the compliance program to make a "quality" determination of seller segments or describe the extent to which seller segment descriptions match the audience characteristics of the associated IDs.

Because this framework introduces new technical and business requirements in order to meet compliance expectations, the framework requirements and associated compliance program are intended to be introduced in a "*crawl, walk, run*" approach. This is representative of version 1.0, representing minimum transparency requirements, as well as a 3-month runway for vendors to react and account for technical and operational updates that may be necessary.

Implementation and Efficacy Benefits of ID vs. Segment-level Disclosures

The proposed disclosures are intended to be applied at the ID-level vs. describing the broader aggregation of segment IDs. This is important for several implementation and framework efficacy reasons.

First, although more granular than segment level disclosures, ID-level disclosures are more easily added to and accrued within a data file given the way that data flows through the supply chain. These labels can simply be included at the moment when unique IDs are added to segment files via additional descriptive ID fields (outlined for review in detail below). Second, the ability to add or remove descriptive file fields also aids with the flexibility of the framework for future iterations. Third, when describing data provenance and age–for compound segments in particular, which tend to be complex aggregations of disparate attribute definitions, source locations, and timestamps–ID-level granularity is necessary to parse out accurate information and understand segment composition along a number of relevant quality dimensions. It is very difficult to generalize about the provenance and age of any given segment. This is especially true when compound segments are made up of multiple other compound segments sourced from multiple providers. And lastly, ID-level visibility un-bundles ID-attribute pairs from the aggregated whole that may be more or less applicable for the buyer's use case.

# Framework Development Process

Beginning in July 2017, the framework went through six checkpoints within the development lifecycle:
1. Developed consensus on primary determinants of data quality inclusive of provenance, age, modeling / manipulation

2. Outlined primary participants involved in the data supply chain, as well as the specific functions / processes vendors perform that affect segment composition
3. Established qualitative and quantitative labels to declare the participants and processes used that govern inclusion of a specific ID-attribute pair in a data segment
4. "Stress tested" group label output against a broad assortment of base and compound segments from a diverse cross-section of seller types. These test results determined the technical and operational rigor associated with compliance and were used to pare down the framework to base minimum label requirements needed to address marketplace needs.
5. Reviewed proposed solution and approach with other leading trade organizations and their members–including the 4A's, DMA, CIMM, and ARF–for additional feedback and proposal refinement
6. Developed a second work stream–housed within IAB Tech Lab's Taxonomy and Mapping working group–to develop a proposed Audience Taxonomy structure

The following material outlines the above referenced data flows and supply-chain mappings and provides descriptions of the specific processes of supply-chain participants use most prominently to govern inclusion of an ID-attribute pair into any given audience segment. *Note that because technology stacks and capabilities vary widely across data sellers and supply chain participants, it is common for individual participants to play multiple roles and perform many of the identified activities that inform ID inclusion. For this reason, the framework's labeling requirements are structured around the specific process used to develop ID-attribute pairs, as opposed to focusing requirements on the entity performing the activity.* Note that while these are the primary activities, there are additional sub-processes not captured in 1.0 that may be helpful to capture in future framework versions.

Visualizing Data Flows and Supply-chain Participants
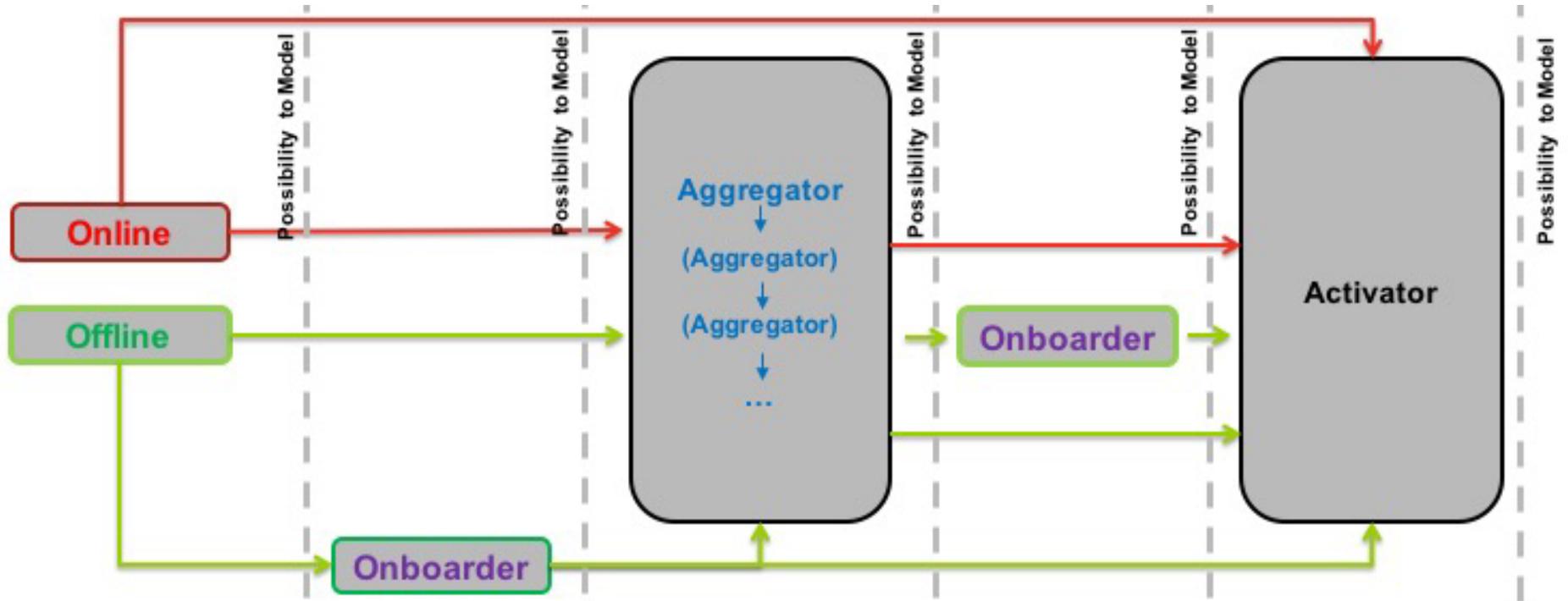
Key Supply-chain Activities That Govern ID Inclusion in a Segment

- o **Online Origination** - ID-attribute pair is determined based on an activity or set of activities performed on one or more <u>owned and operated</u> digital properties (domains or mobile applications)
- o **Onboarding** - ID-attribute pair is determined by first observing an attribute outside digital properties, assigning it to encrypted or unencrypted consumer PII, then associating this with a digital ID (cookie or IFA)
- o **Aggregation** - ID-attribute pairs that have been initially created by an online or offline originator are then organized and assigned attributes that correspond to the ones received
- o **Modeling** (LAL) - ID-attribute pair inferred when comparing characteristics of an ID to those of a seed pool representative of the desired attribute
- o **Cross-device Analysis** - ID-attribute pair determined by associating IDs with other IDs that are known or thought to be used by the same person, household or business
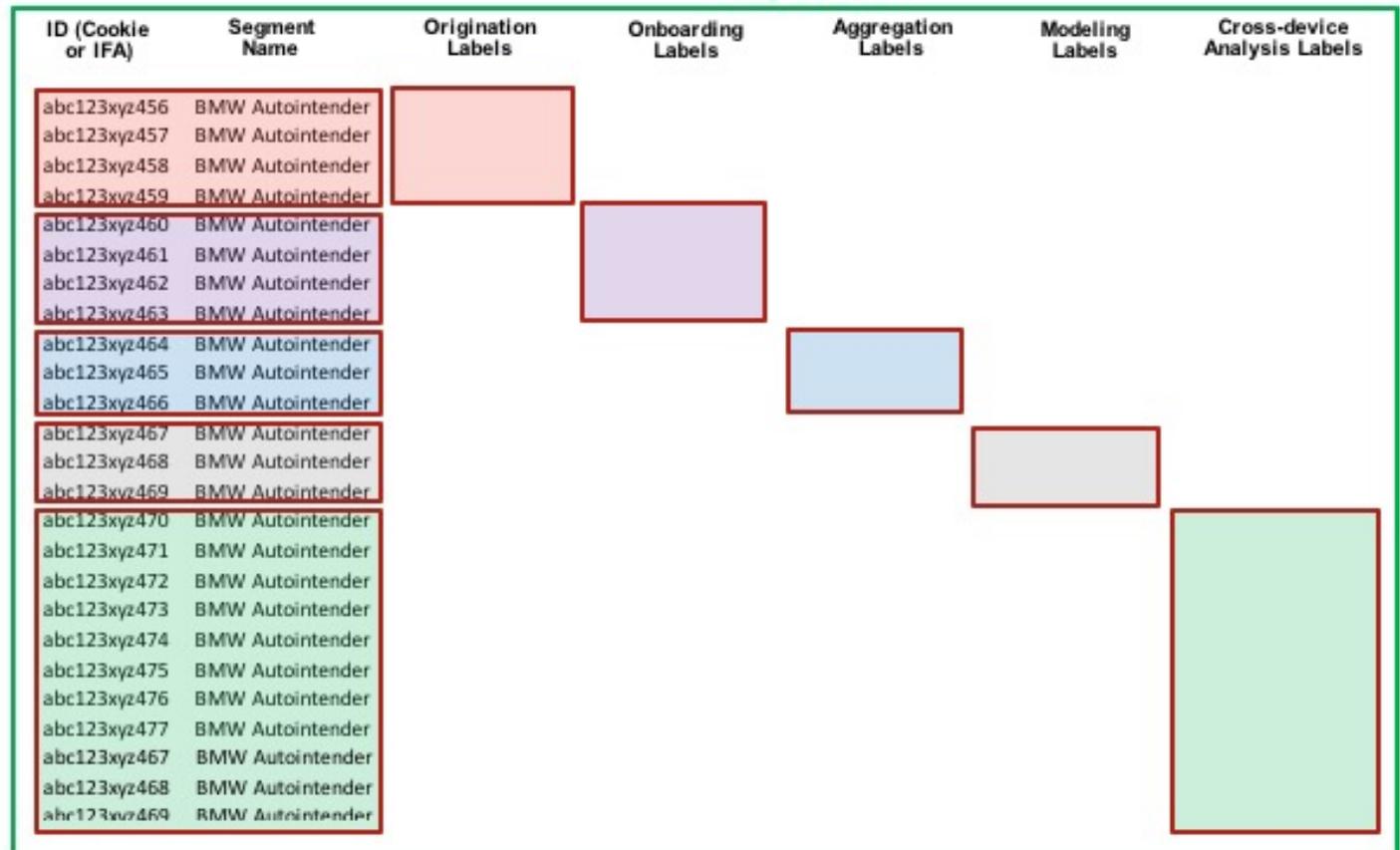
10

Passing ID Disclosures Within a Data Segment File

The above activities are often performed individually or in combination–either within or across organizations–in order to identify many IDs with similar attributes. These are then packaged / sold together as individual *audience segments*. IAB Tech Lab's Transparency Framework outlines ID-level disclosures that are differentiated depending on which activity was used to add the ID-attribute pair to any given segment. They are required to be added and passed by the entity contributing the ID to the segment. The following diagram outlines how the ID level disclosures can be added and passed within a data segment file as it moves through the supply chain–with ID additions added from top to bottom, left to right–and compares this detail with current segment disclosures. This data can accompany any other ID level information–like key value pairs–that a provider may apply for internal use.

Technical and Operational Considerations

In the short term, compliance with the Data Transparency Framework will unavoidably require technical and operational refinements by data sellers and supply chain participants. Over time, resources expended on short term process refinements are expected to be recouped quickly by 1) additional buyer investment in transparency compliant data and 2) storage savings stemming from identification + removal of superfluous ID-attribute data pairs that are currently hard to identify when working with many disparate data partners.

Disk Space

While the ID-level disclosures are intentionally pared down to minimum requirements, the framework nonetheless is expected to require moderate growth of data segment file sizes when information is assembled at the ID-level. The free form text fields–which in some circumstances can contain upwards of 500 alphanumeric characters–are especially resource intensive. IAB Tech Lab suggests that any field requiring free-form disclosures can instead be accommodated by passing a less resource intensive ID that references an internally developed and managed match table containing the text. This match table can be then shared with supply chain partners in order to populate fields. Similar scenarios can be observed with systems like geographic ISO codes within OpenRTB.

However, over time ID-level disclosures will help ease the oft-reported resource drains associated with storing duplicate, non-value add data from many disparate sources by making it easier for data warehousers to identify superfluous ID-attribute pairs.

JSON File Conversion

The Framework's open API under development is intended to create a standardized data schema to communicate segment information between supply chain participants, and in v1.0 will be able to accommodate JSON formats. While JSON is the most common format, in the short-term, supply chain vendors will need to convert .xml or .csv files that they might use to JSON. Future versions of the API will be able to support.xml and .csv conversions.

Human Capital and Process Requirements

Given the sheer volume of audience data and ID level information that flows throughout the ecosystem, it is likely that dedicated data hygiene, taxonomy, and segmentation personnel will be needed to provide systematic oversight of label formatting and completion. The scope and nature of this oversight will largely depend on the technical and automation capabilities of the supply chain participant in question. Additionally, internal processes, controls, and checks + balances will likely need to be developed in order to capture new business requirements associated with compliance.

Appendix: Label and Format Requirements

Online Attribute Origination

Label	Format	Type	Definition
ID	Text field: 100 characters	Required	Unique alpha-numeric string of the cookie or IFA
ID Format	Dropdown (Select from options below) Cookie ID IFA	Required	Type of identifier
Originator Name (Company)	Text field: 100 characters	Required	Name of the organization that is collecting the attribute-ID pair
Date that attribute data was collected	Text field (200 characters)	Required	Date that ID-attribute in question met segmentation criteria and stored by originator
Attribute Definition	Text field (500 characters)	Required	Shorthand, layman's description of the rules applied by the seller that govern inclusion of data points into the audience segment in question. Suggested inclusion of provenance, recency, frequency, logic as necessary.
Online Attribute Source Location	Text field: 100 characters	Required	Root domain or app name of attribute source location
Channel Source Type	Dropdown (Select from options below) Web Mobile Web Browser extension Mobile app Desktop app Email TV/STB Other	Required	Description of the digital channel from which the attribute-ID pair was collected
Content Source Type	Dropdown (Select from options below) Web Article Slideshow FEP Video	Required	Description of the content format from which the attribute-ID pair was collected

Label	Format	Type	Definition
IAB Audience Taxonomy Label	**Dropdown (Select from Audience Taxonomy Options)**	**Required**	**Declaration of the most accurate label from IAB Audience Taxonomy 1.0**
Match Level - Attribute Compilation Granularity	**Dropdown (Select from options below)**	**Required**	**Describes how signal attributes are ascribed to uniques in a segment**
	Browser		
	Device		
Derivation - Relationship of Data Attribute to Originator	**Dropdown (Select all that apply of options below)**	**Required**	**Description of seller's relationship to the audience attribute described**
	Known / Observed		The underlying audience attribute is all directly observed by the seller of the collector of the audience attribute. (Examples of known IDs: first-party cookie, third-party cookie, IDFA, ADID, IP address; Examples of known data: User agent, network level protocols, domain, time stamp, session information, segmentation)
	Declared		The underlying audience attribute is self-reported by the audience members (Examples of declared IDs: social media username, address, email, name, credit info; Examples of declared data: gender, survey responses, clicks/engagement)
	Derived		The underlying audience attributes are computed based on other known or declared fields on record (Examples of Derived ID: Fingerprint ID, Cross-Device ID, Intra-device ID, user ID, Household ID)
	Inferred		The underlying audience attributes are inferred based on business rules or logic, though not necessarily computed (Example of Inferred Data: Look-alike)
	Modeled		The underlying audience attributes are the result of the application of an algorithm developed on the basis of evaluating a seed to a universe (Example of Modeled ID: Stat Fingerprint ID, Stat Cross-Device ID, Stat User ID; Example of Modeled data: truth set, controlled traffic data sets)

Offline Attribute Origination*

*Originator responsible for collecting and passing these to onboarder, which adds to data file along with associated IDs

Label	Format	Type	Definition
Originator Name (Company)	Text field: 100 characters	Required	Name of the organization that is collecting the attribute
Date that attribute data was collected	Text field (200 characters)	Required	Date that attribute in question met segmentation criteria and stored by originator.
Attribute Definition	Text field (500 characters)	Required	Shorthand, layman's description of the rules applied by the seller that govern inclusion of data points into the online audience segment in question. Inclusive of provenance, recency, frequency, logic as necessary.
Offline Attribute Source Type	Dropdown (Select all that apply of options below)	Required	Describes how offline audience attributes are discovered and sourced. Options can include: brick + mortar in-store transaction, online ecommerce transactions resulting in offline data, public record (census record, property deeds, voter files), home address mailing lists (magazines, newspapers, flyers, promotional material), mail-order transactions (catalogues), offline survey data
	Brick + Mortar In-store Transaction		
	Online Ecommerce Transaction (resulting in offline data)		
	Public Record - Census		
	Public Record - Property Deed		
	Public Record - Voter File		
	Public Record - Other (need more depth)		
	Home Address Mailing List (Magazine / Newspaper distribution, other promotional material)		
	Mail Order Transaction (Catalogues, Call-in)		
	Offline Survey		
IAB Audience Taxonomy Label	Dropdown (Select from Audience Taxonomy options)	Required	Most accurate label from IAB Audience Taxonomy 1.0
Match Level - Attribute Compilation Granularity	Dropdown (Select from options below)	Required	Describes how signal attributes are ascribed to uniques in a segment
	Individual		
	Household		
	Business		
	Geographic Area		

Label	Format	Type	Definition
Derivation - Relationship of Data Attribute to Originator	**Dropdown (Select all that apply of options below)**	**Required**	**Establishment of seller's relationship to the audience attribution / information being sold**
	Known / Observed		The underlying audience attribute is all directly observed by the seller of the collector of the audience attribute. (Examples of known IDs: first-party cookie, third-party cookie, IDFA, ADID, IP address; Examples of known data: User agent, network level protocols, domain, time stamp, session information, segmentation)
	Declared		The underlying audience attribute is self-reported by the audience members (Examples of declared IDs: social media username, address, email, name, credit info; Examples of declared data: gender, survey responses, clicks/engagement)
	Derived		The underlying audience attributes are computed based on other known or declared fields on record (Examples of Derived ID: Fingerprint ID, Cross-Device ID, Intra-device ID, User ID, Household ID)
	Inferred		The underlying audience attributes are inferred based on business rules or logic, though not necessarily computed (Example of Inferred Data: Look-alike)
	Modeled		The underlying audience attributes are the result of the application of an algorithm developed on the basis of evaluating a seed to a universe (Example of Modeled ID: Stat Fingerprint ID, Stat Cross-Device ID, Stat user ID; Example of Modeled data: truth set, controlled traffic data sets)

Onboarding

Label	Format	Type	Definition
ID	Text field: 100 characters	Required	Unique alpha-numeric string of the cookie or IFA
ID Format	**Dropdown (Select from options below)**	Required	Type of identifier
	Cookie ID		
	IFA		
Onboarder Name (Company)	**Dropdown (Select from options below)**	Required	Name of the onboarder
	LiveRamp		
	Oracle		
	Experian		
	Epsilon		
	Merkle		
	KBM		
	Neustar		
	Throtle		
	Other		
Date Attribute Data Assigned to Identifier	Text field (200 characters)	Required	Description of process + frequency associated with how attributes are assigned to audience identifiers (cookies or software IDs)
Match Key(s) Between Attribute and Identifier	**Dropdown (Select all that apply of options below)**	Required	Notes the type of data that is used to associate offline data with online ID, as well as the strength of association
	First Name		
	Last Name		
	Postal Address		
	Email Address		
	Phone Number - Mobile		
	Phone Number - Household		
	IP Address		
	Geography		
	Gender		
	Date of Birth		

Label	Format	Type	Definition
Match Level - Attribute Compilation Granularity	**Dropdown (Select from options below)**	**Required**	**Describes how signal attributes are ascribed to uniques in a segment**
	Individual		
	Household		
	Business		
	Geographic Area		

Aggregation

Label	Format	Type	Definition
ID	Text field: 100 characters	Required	Unique alpha-numeric string of the cookie or IFA
ID Format	Dropdown (Select from options below)	Required	Type of identifier
	Cookie ID		
	IFA		
Aggregator Name (Company)	Text field: 100 characters	Required	Name of the company aggregating originators IDs
Date ID Added to Segment	Text field (200 characters)	Required	Date that the aggregator organized originators ID into aggregated segment
Aggregator Attribute Name	Text field: 100 characters	Required	Aggregator's description of segment in question
Aggregator Categorization vs. Originator ID	Dropdown (Select from options below)	Required	How the aggregator infers "Aggregator Attribute Name" based on originator description
	Explicit		Explicit categorizations are captured via information declared in originator description (ex: BMW Intenders -> Luxury car Intenders)
	Implicit		Implicit categorizations are captured via inferences from originator description (ex: BMW Intenders -> Affluent People)
IAB Audience Taxonomy Label	Dropdown (Select from Audience Taxonomy options)	Required	Most accurate label from IAB Audience Taxonomy 1.0
Match Level - Attribute Compilation Granularity	Dropdown (Select from options below)		Describes how signal attributes are ascribed to uniques in a segment
	Browser		
	Device		
	Individual		
	Household		
	Business		
	Geographic Area		

Label	Format	Type	Definition
Derivation - Relationship of Data Attribute to Originator	**Dropdown (Select all that apply of options below)**	**Required**	**Establishment of seller's relationship to the audience attribution / information being sold**
	Known / Observed		The underlying audience attribute is all directly observed by the seller of the collector of the audience attribute. (Examples of known IDs: first-party cookie, third-party cookie, IDFA, ADID, IP address; Examples of known data: User agent, network level protocols, domain, time stamp, session information, segmentation)
	Declared		The underlying audience attribute is self-reported by the audience members (Examples of declared IDs: social media username, address, email, name, credit info; Examples of declared data: gender, survey responses, clicks / engagement)
	Derived		The underlying audience attributes are computed based on other known or declared fields on record (Examples of Derived ID: Fingerprint ID, Cross-Device ID, Intra-device ID, User ID, Household ID)
	Inferred		The underlying audience attributes are inferred based on business rules or logic, though not necessarily computed (Example of Inferred Data: Look-alike)
	Modeled		The underlying audience attributes are the result of the application of an algorithm developed on the basis of evaluating a seed to a universe (Example of Modeled ID: Stat Fingerprint ID, Stat Cross-Device ID, Stat user ID; Example of Modeled data: truth set, controlled traffic data sets)

Modeling

Label	Format	Type	Definition
ID	Text field: 100 characters	Required	Unique alpha-numeric string of the cookie or IFA
ID Format	Dropdown (Select from options below) Cookie ID IFA	Required	Type of identifier
Modeler Name (Company)	Text field: 100 characters	Required	Name of the company adding look-alike modeled IDs
Date that Seed Pool was Modeled	Text field (200 characters)	Required	Date that the seed pool was modeled
Seed Pool Description	Text field: 100 characters	Required	Description of the seed pool of originator IDs used to train look-alike model
# of IDs in Seed Data File	Text field (200 characters)	Required	Number of unique IDs used to train look-alike model
# of IDs in Modeled Output	Text field (200 characters)	Required	Number of unique IDs that were captured in model
Modeled Attribute Name (Output of Model)	Text field: 100 characters	Required	Audience description that the Modeler assigns internally to the ID-attribute combinations resulting from look-alike modeling
IAB Audience Taxonomy Label	Dropdown (Select from Audience Taxonomy Options)	Required	Most accurate label from IAB Audience Taxonomy 1.0
Match Level - Attribute Compilation Granularity	Dropdown (Select from options below) Browser Device Individual Household Business Geographic Area	Required	Describes how signal attributes are ascribed to uniques in a segment

Label	Format	Type	Definition
Derivation - Relationship of Data Attribute to Originator	**Dropdown (Select all that apply of options below)**	**Required**	**Establishment of seller's relationship to the audience attribution / information being sold**
	Known / Observed		The underlying audience attribute is all directly observed by the seller of the collector of the audience attribute. (Examples of known IDs: first-party cookie, third-party cookie, IDFA, ADID, IP address; Examples of known data: User agent, network level protocols, domain, time stamp, session information, segmentation)
	Declared		The underlying audience attribute is self-reported by the audience members (Examples of declared IDs: social media username, address, email, name, credit info; Examples of declared data: gender, survey responses, clicks / engagement)
	Derived		The underlying audience attributes are computed based on other known or declared fields on record (Examples of Derived ID: Fingerprint ID, Cross-Device ID, Intra-device ID, User ID, Household ID)
	Inferred		The underlying audience attributes are inferred based on business rules or logic, though not necessarily computed (Example of Inferred Data: Look-alike)
	Modeled		The underlying audience attributes are the result of the application of an algorithm developed on the basis of evaluating a seed to a universe (Example of Modeled ID: Stat Fingerprint ID, Stat Cross-Device ID, Stat user ID; Example of modeled data: truth set, controlled traffic data sets)

Cross-device Analysis

Label	Format	Type	Definition
ID	Text field: 100 characters	Required	Unique alpha-numeric string of the cookie or IFA
ID Format	Dropdown (Select from options below) Cookie ID IFA	Required	Type of identifier
Name of Company	Text field: 100 characters	Required	Name of the company adding cross-device IDs
Date of Cross-Device ID Match	Text field (200 characters)	Required	Date that the ID was most recently associated with a user, household, business, or geographic area
Date ID Added to Segment	Text field: 100 characters	Required	Date that the cross-device company added ID into aggregated segment
Match Level - Attribute Compilation Granularity	Dropdown (Select from options below) Individual Household Business Geographic Area	Required	Describes how signal attributes are ascribed to uniques in a segment
Derivation - Relationship of Data Attribute to Originator	Dropdown (Select all that apply of options below)	Required	Establishment of seller's relationship to the audience attribution / information being sold
	Known / Observed		The underlying audience attribute is all directly observed by the seller of the collector of the audience attribute. (Examples of known IDs: first-party cookie, third-party cookie, IDFA, ADID, IP address; Examples of known data: User agent, network level protocols, domain, time stamp, session information, segmentation)
	Declared		The underlying audience attribute is self-reported by the audience members (Examples of declared IDs: social media username, address, email, name, credit info; Examples of declared data: gender, survey responses, clicks / engagement)
	Derived		The underlying audience attributes are computed based on other known or declared fields on record (Examples of Derived ID: Fingerprint ID, Cross-Device ID, Intra-device ID, User ID, Household ID)
	Inferred		The underlying audience attributes are inferred based on business rules or logic, though not necessarily computed (Example of Inferred Data: Look-alike)
	Modeled		The underlying audience attributes are the result of the application of an algorithm developed on the basis of evaluating a seed to a universe (Example of Modeled ID: Stat Fingerprint ID, Stat Cross-Device ID, Stat user ID; Example of modeled data: truth set, controlled traffic data sets)

Suggested Calculated Fields

The framework allows for easy calculation of ratios and analyses of segment composition. These are several examples the working group developed, with opportunities for many additional depending on preferences of UI in question.

Field		Format	Definition
Count of Unique IDs in Final Segment		**Text field (Range: 1-10,000,000,000)**	**The number of IDs in final segment being offered by seller–after all modelling takes place–associated with differentiated uniques (individual, household).**
Count of Unique Attribute Originators (sources)		**Text field (Range: 1-10,000,000,000)**	**The number of unique sources where attributes / IDs were identified. If sourced online, source = root domain. If offline, source possibilities listed in above requirements.**
Count of Originator Attribute Types			**Quantifies relationship to the audience attribution / information being sold**
	Known / Observed	Text field (Range: 1-10,000,000,000)	
	Declared	Text field (Range: 1-10,000,000,000)	
	Derived	Text field (Range: 1-10,000,000,000)	
	Inferred	Text field (Range: 1-10,000,000,000)	
	Modeled	Text field (Range: 1-10,000,000,000)	
Average Age of Attribute (Delta been date pushed to activator and date attribute collected)		**Text field (Range: 1-10,000,000,000)**	**Elapsed time between attribution collection date and date segment pushed to activator**