



# **Data Transparency Standards**

Compliance Guide

*Version 1.0 | June 2019*

<b>Executive Summary</b>	<b>3</b>
<b>Change Log</b>	<b>5</b>
<b>Certification Process</b>	<b>6</b>
Certification Fees	6
Qualification	7
Geographic Applicability of Certification	7
Certification Testing	7
Dissonance when completing the labelling	7
People & Process validity	8
Technology validity	8
Publication of Certification Status	8
<b>Continued Certification</b>	<b>9</b>
Recertification and Maintaining Compliance	9
<b>Appendix A: Application Information</b>	<b>10</b>
<b>Appendix B: Validation Test Cases</b>	<b>11</b>
Certification Testing	11
Dissonance Control Objective:	11
People & Process Control Objective:	16
Technology Control Objective:	21

## Executive Summary

The IAB Tech Lab Data Transparency standard establishes minimum disclosure requirements for audience data providers. It is intended for:

- Providers that collect, segment, and market data as a standalone product
- Providers that collect, segment, and market data as a coupled / bundled offering along with media
- Data marketplaces that broker data between buyers and sellers and represent the “point of purchase”

These disclosure requirements are intended to establish a baseline level of transparency for data buyers about aspects of data collection, processing, and modeling that inform data quality and applicability, regardless of buyer use case. These standards are not intended to provide a qualitative grade as to the *efficacy* (“this segment performs well”) or *quality* (“this segment is highly accurate”) of the data in question, but simply surface baseline information that buyers can use to make informed decisions regardless of their data use case.

Companies that agree to be part of the program will go through an annual business audit to confirm that the information provided within the labelling is reliable, that the organization has the necessary systems, processes, and personnel in place to sustain consistent label completion at scale, and that a label can be produced for all in-market segments available. Engagements are expected to last anywhere between 2-5 months, depending upon the size and complexity of the company’s business. Organizations that complete the program will:

- Be issued an IAB Tech Lab compliance seal confirming their adherence to transparency principles and best practices.
- Have their labelling data populated alongside other compliant data within a centralized Tech Lab operated search and discovery tool hosted at <http://www.datalabel.org/>. This will take place either via integrations with participating data marketplaces, or direct upload if the provider doesn’t work with a participating marketplace. This repository is only intended for Tech Lab members and will only house descriptive segment labels instead of the segment IDs themselves (and thus can’t be used for platform activations).

This guide describes the requirements and process of Data Transparency Standards Compliance and adoption.

The Data Transparency standard is developed and managed by the [Data Transparency Standards Working Group](#). This group will serve as an ongoing governance body for the Transparency Standard, and is tasked with ensuring that the minimum disclosure requirement standard is continuously refined and updated based on marketplace needs. As such, version 1.0 will likely evolve in the coming years upon working group approval. The Data Transparency Standards working group is open to all Tech Lab members. If interested in participating, please reach out to [membership@iabtechlab.com](mailto:membership@iabtechlab.com).

## Change Log

Revision	Description	Author	Date
Version 1.0	Document Created, First Draft of Standard Released	Benjamin Dick / Jarrett Wold	6/27/2019

## Certification Process

The Data Transparency Compliance is voluntary and certifies a data provider is meeting the minimum segments disclosures to be transparent in accordance with the Data Transparency Standard. The certification is performed by an IAB Tech Lab approved third party independent auditor that validates compliance with the guidelines for the following three components:

- Dissonance when completing the labelling
- People & Process validity
- Technology validity

The precise content of the audit is proprietary and confidential, but includes a comprehensive review of various dimensions of the business, including:

- Client-facing and internal documentation
- Internal procedures
- Compliance with the IAB Tech Lab's Data Transparency Standard
- Technical implementation details
- Live tests and data extractions
- Platform availability

## Certification Fees

There is an annual fee for the compliance program that covers an initial annual certification and intermittent segment sampling / auditing for the remainder of the year. Pricing to support the program is tiered depending on how many segments or unique audiences a company is expected to make available to the marketplace that year so that costs are equitably distributed among large and small providers. You can check for pricing specifics for your organization by reaching out to [compliance@iabtechlab.com](mailto:compliance@iabtechlab.com).

## Qualification

The IAB Tech Lab Data Transparency standard establishes minimum disclosure requirements for audience data providers and marketplaces. It can be applied to three constituencies:

- Providers that collect, segment, and market data as a standalone product
- Providers that collect, segment, and market data as a coupled / bundled offering along with media
- Data marketplaces that broker data between buyers and sellers and represent the “point of purchase”

## Geographic Applicability of Certification

The Data Transparency certification is a global program and applies to all geographic regions.

## Certification Testing

A third party independent audit validates against the three pillars noted above: Dissonance, People & Process and Technology. The audit is of the company as a whole, and while it does look at a representative sample size of individual segment labels produced by the seller in order to make that determination, it doesn't validate individual segments. Instead, it validates all segments produced by that company for the year in question. Below is an overview of the scope of the validation testing within each pillar:

- **Dissonance when completing the labelling**

This component is intended to address the following questions: *Is the provider filling out the labels accurately, completely in a way that makes sense? Are formatting requirements met? Do any fields contradict others?*

- **People & Process validity**

This component is intended to address the following question: *Does the provider have the correct people, processes, and organization structure in place to effectively process and deliver label information at scale?*

- **Technology validity**

This component is intended to address the following question: *Does the company have the requisite systems and technical capabilities in place to source the label information in question?*

Because the validation is at the company level - and establishes that a provider has systematized conversancy with accurately populating and making available minimum data disclosures - the sample size that is evaluated and the overall fees associated with compliance are based on the provider's **expected total in market exposure** (ie: the # of segments or audiences a providers plans to make available that year). Dissonance, process, and technology testing and auditing work is based on a significant sample size based on that declared number. After compliance is established based on this in-market figure, intermittent label sampling will take place to ensure a) ongoing adherence to the labeling requirements across all in market data, and b) that the total in market figure that was tested against is not exceeded. If a provider begins to either incorrectly populate or exceed their declared in market exposure, the use of the compliance seal could be temporarily revoked, as well as trigger a follow up audit engagement (see below).

## Publication of Certification Status

Upon successful completion of the certification testing process and approval of certification:

- The Data Provider is issued a certificate and IAB Tech Lab compliance seal of approval. A certified company may use the seal to publicly communicate its Data Transparency certification
- IAB Tech Lab will list the Data Providers who have been issued certification on our website here: <https://iabtechlab.com/technology-compliant-companies/>. In addition, IAB Tech Lab will also publish a public version of the certification on the website.

- The Data Provider will have their labelling data populated alongside other compliant data providers within a centralized Tech Lab operated search and discovery tool called [datalabel.org](http://datalabel.org). This will take place either via integrations with participating data marketplaces, or direct upload if the provider doesn't work with a participating marketplace. This repository is only accessible by Tech Lab members and will only house descriptive segment labels vs the segment IDs themselves (and is thus not intended to be used as an activation solution for data).

## **Continued Certification**

Data providers and marketplaces that have achieved the Data Transparency certification must maintain the certification on an annual basis for continued use of the seal and certificate. This is due to the ongoing changes to marketplace requirements for the labelling itself.

## **Recertification and Maintaining Compliance**

Upon completion of the first engagement, an annual re-engagement date will be set. To maintain continuous Data Transparency certification, the Data Provider must:

1. Ensure that all dissonance, people / process, and technology checks continue to be sufficient
2. That in-market audience exposure did not exceed declared figures
3. That segments continue to be populated and/or accessible to any buyer upon request

## Appendix A: Application Information

<b>Email address</b>	
<b>Your Name</b>	
<b>Organization Name</b>	
<b>In What Capacity Does Your Company Provide Data Products or Services?</b>	
<b>How Many Unique Audiences Does Your Company Make Available?</b>	
<b>Comments</b>	

## Appendix B: Validation Test Cases

### Certification Testing

Third party independent audit validates against the 3 pillars of Dissonance, People & Process and Technology.

**Dissonance Control Objective:**

Determine the Data Label (DL) accurately, objectively, and completely represents the data provided by the “Data Solution Providers”.

Test No.	Control Area	Tests	Expected Results
1.1	Name of Data Provider	Data provider organization is incorporated / functioning in market as a data seller	Information is accurate
1.2	Provider Name	Data provider has the right to use the name / trademark	Confirmed
2-1	Provider Contact Info	Email address provided is a real / functioning email	Email is real
3.1 & 4.2	Segment Name	Cross-check content in following fields: A. Segmentation Criteria B. Standard Segment name C. ID Type D. Audience Precision Level E. Geography	The five fields are in agreement.
4.1	Standardized Segment Name	Ensure legitimate entry from IAB Audience Taxonomy 1.0	Name matches taxonomy.

5.1	Segmentation Criteria	<p>Check that the criteria described addresses business rules for ID inclusion, and is not overly reliant on peripheral segment description.</p> <p>At a minimum this description needs to address:</p> <ul style="list-style-type: none"> <li>A. Data provenance,</li> <li>B. frequency at which attribute needs to be observed to be associated with ID,</li> <li>C. and how long ago the attribute was observed.</li> </ul> <p>If business rules for inclusion are proprietary / algorithmically driven, describe algorithm behavior as it relates to the above items.</p>	<p>Business rules for ID inclusion are disclosed including data provenance, frequency and length.</p> <p>Algorithm behavior is disclosed.</p>
5.2	Segmentation Criteria	<p>Cross-check content in following fields:</p> <ul style="list-style-type: none"> <li>A. Audience Refresh Cadence</li> <li>B. Source Look Back window</li> <li>C. Data Sources</li> </ul>	All references are in sync.
6.1	Audience Precision Level	<p>Cross-check content in following fields:</p> <ul style="list-style-type: none"> <li>A. Segmentation Criteria</li> <li>B. Standard Segment name</li> <li>C. Segment name</li> <li>D. ID Type</li> </ul>	All references are in sync.
7.1	ID Count	<p>The count is disclosed (Figure is variable depending on seasonality, activation platform match rates, and other factors)</p>	Count is disclosed.
8.1	ID Type(s)	<p>Cross-check content in following fields:</p> <ul style="list-style-type: none"> <li>A. Data Sources</li> <li>B. Segmentation Criteria</li> </ul>	All references are in sync.

9.1	<p>Geography</p> <p>Standard Country list is used.</p>	<p>Cross-check content to the Segmentation Criteria</p>	<p>Cross reference is in sync.</p>
10.1	<p>Privacy Policy</p>	<p>Link is functional and leads to indicated privacy content</p>	<p>Link works</p>
11.1	<p>Data Source(s)</p> <p><b>Selection of Offline Sources:</b> Requires completion of "On-boarder Details" section</p>	<p>Cross-check content in following fields:</p> <p>A. ID Type B. Segmentation Criteria</p> <p>If "Offline Survey", "Public Record", or "Offline Transaction" is stated, the "On-boarder Details" section must be completed and properly formatted.</p>	<p>All references are in sync</p> <p>On-boarder Details section provided with appropriate disclosure.</p>
12.1	<p>Data Inclusion Methodology</p> <p><b>Selection of "Modeling":</b> Requires selection of "Yes" within "Audience Expansion" field.</p>	<p>Cross-check content in the Segmentation Criteria</p> <p>If "modeled" is stated, ensure that "Seed Size" is completed and accurately formatted</p> <p>Ensure that "Audience Expansion" is completed and accurately formatted</p>	<p>Cross reference is in sync.</p> <p>Checked and disclosed.</p> <p>Audience Expansion section is complete</p>
13.1	<p>Audience Expansion</p> <p><b>Selection of "Modeling":</b> Requires selection of "Yes" within "Audience Expansion" field.</p>	<p>Cross-check content in following fields:</p> <p>A. Data Inclusion Methodology B. Segmentation Criteria</p>	<p>All references are in sync</p>
14.1	<p>Cross-device Expansion</p>	<p>Cross-check content in Segmentation Criteria</p>	<p>Cross reference is in sync</p>

15.1	<p>Audience Refresh Cadence</p> <p><b>Selection of Offline Sources:</b> Requires completion of "On-boarder Details" section</p>	<p>Cross-check against content in following fields:</p> <p>A. Segmentation Criteria B. Data Sources - if any "Offline" source is selected, "Continuous" or "Daily" are not acceptable options</p>	<p>All references are in sync Only acceptable options are listed.</p> <p>On-boarder Details section provided with appropriate disclosure.</p>
16.1	Source Look Back Window	<p>Cross-check against content in following fields:</p> <p>A. Segmentation Criteria B. Audience Refresh Cadence</p>	All references are in sync
17.1, 18.1, 19.1, 20.1	<p><b>On-boarder Details:</b> Input ID / Match Key,</p> <p><b>On-boarder Details:</b> Audience Precision Level,</p> <p><b>On-boarder Details:</b> Audience Expansion,</p> <p><b>On-boarder Details:</b> Cross Device Expansion</p> <p><b>Selection of Offline Sources:</b> Requires completion of "On-boarder Details" section</p>	<p>If "Offline Survey", "Public Record", or "Offline Transaction" is stated, the "On-boarder Details" section must be completed and properly formatted.</p>	On-boarder Details section provided with appropriate disclosure.

17.2	<p><a href="#">On-boarder Details</a>: Input ID / Match Key</p>	<p>Cross check with "Audience Precision Level" in On-boarder Details section:</p> <p>If "Individual" is stated the following input/match keys are required: Name AND address, -or- Name AND email, -ior- Mobile ID, -or- Cookie ID, -or- phone number</p> <p>If "Household" is stated, the following input/match keys are required: Address, -or- postal/geo code, -or- lat/long</p> <p>If "Geographic" is stated, the following input/match keys are required: street address, -or- postal / geo code, -or- lat / long</p>	<p>On-boarder Details section provided with appropriate disclosure.</p> <p>Appropriate match keys are disclosed.</p>
18.2	<p><a href="#">On-boarder Details</a>: Audience Precision Level</p>	<p>Verify that selection corresponds with minimum requirements for "Input ID / Match Key" declaration:</p> <p>If "Individual" is selected, the following input/match keys are required: Name AND address, -or- Name AND email, -or- Mobile ID, -or- Cookie ID, -or- phone number</p> <p>If "Household" is selected, the following input/match keys are required: Address, -or- postal/geo code, -or- lat/long</p> <p>If "Geographic" is selected, the following input/match keys are required: street address, -or- postal / geo code, -or- lat / long</p>	<p>Appropriate match keys are disclosed.</p>

**People & Process Control Objective:**

To obtain a full understanding of the data collection system of the Data Solution Provider, including processes and personnel involved along with controls that are in place.

Test No.	Control Area	Tests	Expected Results
	General	<p>Examine documentation to support the online and offline data partner's acquisition, data transfer and quality control processes.</p> <p>Walk through workflow of data logging process.</p> <p>Review system control documentation within the data loading process to prevent loss or corruption of data.</p> <p>Review data aggregation job process.</p> <p>Review controls and test documentation to ensure successful and complete collection and aggregation of log file data.</p>	<p>Documentation supports the online and offline data partner's acquisition, data transfer and quality control processes.</p> <p>Data logging workflow process review successful. No gaps.</p> <p>Documentation for system controls within the data loading process to prevent loss or corruption of data exists, tested and verified.</p> <p>Data aggregation documentation and process review successful. No gaps.</p> <p>Review of controls and test documentation for collecting and aggregating log file data successful. No gaps.</p>
2.2	Provider Contact Info	Email is delivered to designated FTE	Confirm the email address is of the Provider.

6.2	Audience Precision Level	<p>Provide documentation re how identity is resolved: via in house tech, via a partner, or some combination of both in house and partner.</p> <p>If in house: methodology description (DOM) - documentation to be provided outlining systems for data flow, and for resolving identity</p> <p>If via partner: provide a description of partner, SOW with partner, examples of log files / data ingestion points</p> <p>If both in house + provider: see requirements above for both</p>	Documentation supports capability and disclosure made in the data label for the specific process used.
8.2	ID Type(s)	Confirm on segment files that unique ID types are represented	Provision for ID types are found in the system and assigned correctly.
9.2	Geography Standard Country list is used.	Provide documentation for the process used to assign geographic attribute to the ID level record	Documentation supports capability and disclosure made in the data label for geography.

<p>11.2, 12.2, 15.2</p>	<p>Data Source(s)  Data Inclusion Methodology,  Audience Refresh Cadence</p>	<p>Documentation made available of:  Org Chart of team directly involved in collecting, analyzing, onboarding, or otherwise preparing the data for sale to buyers.  Job descriptions of product, engineering, and business development stakeholders within org chart.  Process by which relevant stakeholders interact to collect and process data.  Process by which relevant stakeholders interact to populate the IAB Tech Lab data transparency standard,  and description of responsible, accountable, consulted, informed parties (RACI) for <u>each label field</u></p>	<p>Org chart confirmed and fully lists personnel involved.  Job descriptions checked for engineering and business development. Appropriate disclosure of level of authority is included.  Documentation is present and adequately discloses roles and responsibilities with respect to the processing of data and completing data labels.</p>
<p>11.3, 12.3, 15.3</p>	<p>Data Source(s)  Data Inclusion Methodology,  Audience Refresh Cadence</p>	<p>Review the interaction between people and processes within the organization.</p>	<p>Evidence of an accurate, repeatable process for all current and future data segments and thus, labels is present.</p>
<p>12.4</p>	<p>Data Inclusion Methodology  <b>Selection of “Modeling”:</b> Requires selection of “Yes” within “Audience Expansion” field.</p>	<p>Conditional checks:  If “Modeled” is selected - documentation shall be provided for model input, output, and scores associated with behavior.</p>	<p>Documentation is provided and sufficiently explains the input, output and scoring. Tested and verified.</p>

13.2, 19.2	<p>Audience Expansion</p> <p><b>On-boarder Details:</b> Audience Expansion</p> <p><b>Selection of “Modeling”:</b> Requires selection of “Yes” within “Audience Expansion” field.</p>	<p>Provide documentation of seed ID footprint with suitable number of attributes to enable modeling / scoring, which outlines the types of attributes being used as an input into the model</p>	<p>Documentation is provided and sufficiently explains the seed ID footprint and attributes process. Tested and verified.</p>
13.3, 14.3, 19.3, 20.3.	<p>Audience Expansion</p> <p>Cross-device Expansion</p> <p><b>On-boarder Details:</b> Audience Expansion</p> <p><b>On-boarder Details:</b> Cross Device Expansion</p> <p><b>Selection of “Modeling”:</b> Requires selection of “Yes” within “Audience Expansion” field.</p>	<p>Provide Output file documentation to ensure that resulting ID count is larger than ID count input file (see Input ID / Match Key log file)</p>	<p>Output count check verified.</p>
13.4, 19.4	<p>Audience Expansion</p> <p><b>On-boarder Details:</b> Audience Expansion</p> <p><b>Selection of “Modeling”:</b> Requires selection of “Yes” within “Audience Expansion” field.</p>	<p>Provide documentation of match report / data portrait analysis (ie, demographic profile)</p>	<p>Match report/portrait analysis checked and verified.</p>
14.2, 20.2	<p>Cross-device Expansion</p> <p><b>On-boarder Details:</b> Cross Device Expansion</p>	<p>Provide documentation of consumer (device) footprint with suitable number of attributes to enable modeling and scoring that are tied to the same Input ID</p>	<p>Documentation is provided and sufficiently explains the process of attributing and modeling to the same Input ID. Tested and verified.</p>

15.4	Audience Refresh Cadence	<p>Syndication record documentation / analysis – evaluation of the delta between syndication records to ensure that IDs were added or removed from segment, and that syndication occurred within the declared cadence.</p> <p>Documentation needs to demonstrate:</p> <p>A. 3-5 syndication attempts (if continuous selected, multiple examples of intra-day refresh need to be provided) over the course of at least a year,</p> <p>B. and all syndication records need to be maintained for evaluation a minimum of one month</p>	<p>Syndication record documents frequency matching the declared cadence.</p> <p>Process for adding and removing IDs tested and verified.</p>
16.2	Source Look Back Window	Documentation of: Source event record associated with IDs to ensure date stamp matches segment's declared refresh cadence.	Date stamp coincides with refresh cadence. Tested and verified.
17.3, 18.3	Input ID / Match Key, Audience Precision Level	Provide processing logs and match reports, ensure that inputs and outputs match.	Logs and reports checked and confirm inputs and outputs match.
17.4, 18.4	Input ID / Match Key, Audience Precision Level	Show existing in-house database records that contain data on the same match key for indicated level of audience granularity	In-house database checked.

**Technology Control Objective:**

Controls provide reasonable assurance that:

- A. data obtained from every online data partner's web activity represents relevant human traffic as much as possible.
- B. data sourcing, collection and validation processes produce accurate user cookie data.
- C. demographic segments are collected, analyzed, and aggregated consistently across the platform.
- D. data is logged accurately and consistently, and accuracy is maintained throughout the collection and aggregation process

Test No.	Control Area	Tests	Expected Results
	General	<p>Review architecture of the Data Solution Provider.</p> <p>Review the flow of data through the system: input, processing, at rest, output (in transit).</p> <p>Review log file and data retention protocols.</p> <p>Review system controls within the data loading process to prevent loss or corruption of data.</p> <p>Review data table structure.</p> <p>Analyze controls and tests in place to ensure successful and complete collection and aggregation of log file data.</p>	<p>The architecture of the Data Solution Provider is reviewed and aligns to the support documentation provided under the people and process General objective. No variance between the physical review and the documentation present.</p> <p>The data flow review through the system reflects input, processing, at rest, and output (in transit) as well as corresponds to the support documentation provided under the people and process General objective. No discrepancy between the physical review and the documentation present.</p> <p>The log file and data retention protocols are reviewed and align to the support documentation provided under the people and process General objective. No difference between the physical review and the documentation present.</p> <p>The system controls</p>

			<p>analysis within the loading process is reviewed and align to the support documentation provided under the people and process General objective. No variance between the physical review and the documentation present.</p> <p>The data table structure is reviewed and correlates to the support documentation provided under the people and process General objective. No discrepancy between the physical review and the documentation present.</p> <p>The analysis of the controls and test in place for log file data collection and aggregation. No difference between the physical review and the documentation present.</p>
13.5, 19.5	<p>Audience Expansion</p> <p><b>On-boarder Details:</b> Audience Expansion</p> <p><b>Selection of “Modeling”:</b> Requires selection of “Yes” within “Audience Expansion” field</p>	Analyze modeling code / script / routine that is used to process input, create and validate model, and select look-a-like data	Modeling code / script is reviewed and supports people and process documentation outlined for audience expansion. No difference between the physical review and the documentation present.
13.6, 19.6	<p>Audience Expansion</p> <p><b>On-boarder Details:</b> Audience Expansion</p> <p><b>Selection of “Modeling”:</b> Requires selection of “Yes” within “Audience Expansion” field.</p>	Analyze Internal environment where script is run (cloud service / grid system)	Internal environment is analyzed and supports people and process documentation outlined for audience expansion. No difference between the physical review and the documentation present.

<p>14.4, 20.4, 20.5</p>	<p>Cross-device Expansion</p> <p><a href="#">On-boarder Details</a>: Cross Device Expansion</p> <p><a href="#">On-boarder Details</a>: Cross Device Expansion</p>	<p>If in-house match table is used to expand segment the Github / product specs for match table are provided; methodology associated with ID resolution is described and documented, provide the modeling code / script / routine that is used to associate information over time.</p> <p>If match table used to expand segment is provided by 3<sup>rd</sup> party, provide MSA associated with business partner that does analysis.</p> <p>Provide process documentation of API calls to outside cross-device expansion partner (which would contain log files, reporting and billing information)</p>	<p>Modeling code / script / routine is reviewed against the Github / product specs and supports the people and process outline for cross-device expansion. No difference between the physical review and the documentation present.</p> <p>Where segment expansion is provided by a 3<sup>rd</sup> party, the MSA is reviewed against the match table. No difference between the physical review and the documentation present.</p> <p>Where an API call is used for expansion, the documentation process and production environment are reviewed. No difference between the physical review and the documentation present.</p>
---------------------------------	---	--	---