

# Highlights From “Data Accuracy: How to Fix Programmatic’s Unsolved Problem”

*Presented by Jake Moskowitz, Head of Emodo Institute, Ericsson Emodo*

This presentation discusses parameters for marketers to make data decisions that mitigate against the tradeoffs that are an inherent part of the data landscape today. It looks at:

1. The Seriousness of Data Accuracy in Programmatic
2. The Potential Points of Failure in Making a Segment
3. Key Questions to Ask Data Vendors
4. How to Calculate the Cost of Bad Data

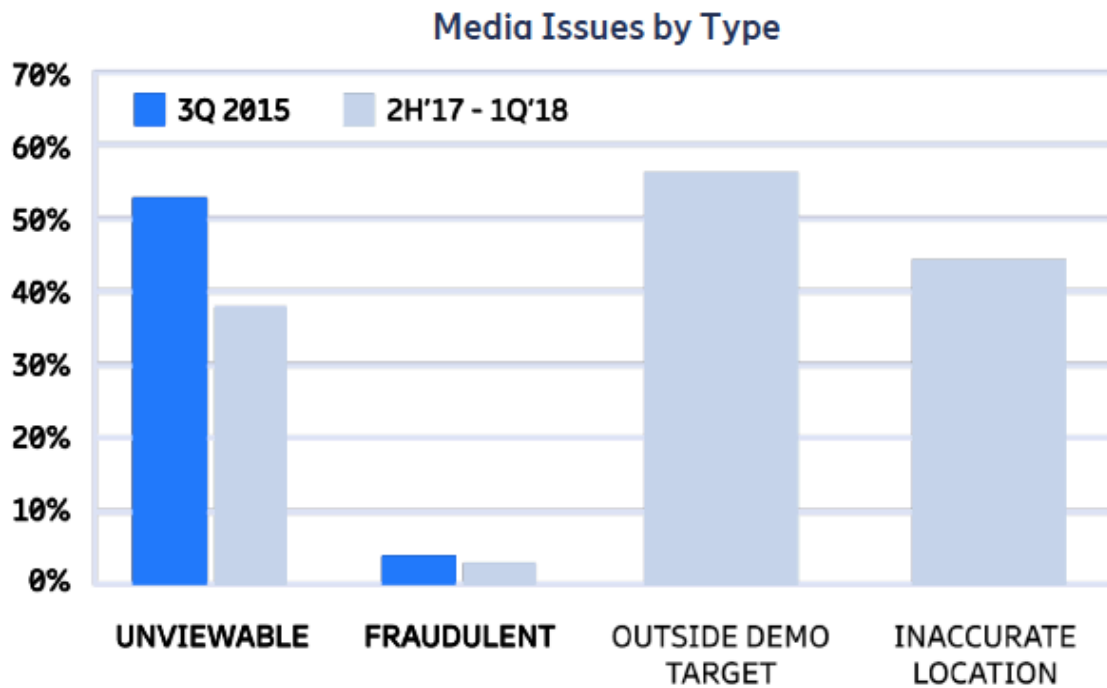
Most marketers make decisions about data based on reach, their trust of the data’s source, the ease with which they can create segments, price, and the segment’s description, but none of these address tradeoffs that start data down the path of being inaccurate. For instance, if the segment description is “people in the market for a car,” how is that being defined?

## **1. The Seriousness of Data Accuracy in Programmatic**

Here are statistics that show the scope of the data market -- and the scope of the problem:

- **\$20 billion:** The amount of money spent in 2017 on third party targeting data, according to the Interactive Advertising Bureau.
- **130,000+:** The number of segments available within leading data stores.
- **9%:** The share of impressions that are fraudulent, according to an Integral Ad Science Media Quality Report.
- **41%:** The share of impressions that are non-viewable, also according to IAS.
- **45%:** The share of location data that is inaccurate based on an analysis by Ericsson Emodo of its carrier data, which was deterministically matched to third-party and SDK and exchange data to verify user location.
- **56%:** The share of demo targeting data that is inaccurate, according to Nielsen Digital Ad Ratings norms.
- **34%:** The reduction in viewability and fraud issues over the last 2.5 years, per IAS.
- **0%:** The number of industry standards, regulations or initiatives to improve data accuracy.

That there’s actually been a reduction in viewability and fraud doesn’t seem to square with the lack of industry initiatives to solve those issues. The reason cited by IAS is that people in the industry ***started to care about it***. According to Moskowitz, the way to fix the problem of programmatic data inaccuracy is the same -- marketers have to start caring. The scope of the data problem compared to viewability and fraud is detailed below:



Sources: IAS Media Quality Reports 3Q'15 and 2H'17 for Direct Publisher Display; Nielsen DAR Benchmarks Reports 1Q'18 for M/F 16-30 years; Ericsson Emodo Verification Benchmarks 2017-18

Inaccurate data falls into three general categories:

1. **Data That Comes From Bad Sources.**
2. **Data That Creates Segments With Too Many Tradeoffs.** At each point in the creation human beings have to decide whether each data point is being used toward accuracy or scale. (The default is usually scale.)
3. **Data That Is Unverified.** Most of the time, when marketers ask vendors to validate data, they are doing it using pattern recognition -- finding regularities in data. Emodo thinks the best way to verify data is to deterministically match data to an accurate truth set.

Solving these issues becomes more difficult in marketing, because within the industry, data science is a veiled science; everything from competition, to rapid sales cycles, to human perception keeps the realities of data hidden in a black box.

## **2. The Potential Points of Failure in Creating a Segment**

The first step to opening the black box is understanding how data points morph throughout the seven stages they go through, ending with being used to target a consumer. Here's more detail:

1. **Occurrence.** A data point is created, maybe as a lat/long that occurs for example, because an individual's device is near a Subway sandwich store.

2. **Categorization.** The data point needs to be categorized, perhaps by matching it to a point of interest database.
3. **Defining.** Even after categorization, a segment needs to be defined. Would a potential Subway shopper be someone who has been in a Subway three times in the last week, or someone who visits less frequently?
4. **Expanding.** This usually happens via lookalikes, and vendors have to decide how to tradeoff between accuracy and scale.
5. **Matching.** Somewhere along the line, every segment is database matched to LiveRamp, Oracle, The Trade Desk or another provider.
6. **Cross-platform.** Marketers want cross-platform segments, but these databases are by definition probabilistic, which is less accurate. Moving data from one platform to use in another also introduces inaccuracies.
7. **Use.** Ultimately, a human being is going to decide which segments to use.

This can map out in prominent data stores in wildly inaccurate. Some examples:

- One vendor has a segment of 60 million devices have visited a Hyundai dealership in the last 30 days, when 17 million new cars were sold in the U.S. during all of 2017.
- Another vendor has a segment that shows there are 128 million devices in the U.S. associated with drinkers of Millstone coffee, a lesser known brand.

### **3. Key Questions to Ask Data Vendors**

So how can marketers feel better about their data vendors? Make sure to ask questions that only the product team can answer; if the marketing team can answer them, they are not good enough questions. Here is a list of eight specific queries, broken into two categories:

#### Bad Data Sources

1. What percent of your data do you throw out?
2. How do you verify accuracy?
3. How do you verify your POI (point of interest) data?
4. What restrictions do you put on your deterministic data due to privacy concerns?

#### Data Tradeoffs

1. What percentage of your data is modeled?
2. What's your data match rate? (Also: Who did you match to? How many matches occurred before I could use this data? What was your match rate at each stage?)
3. What percentage of the data is running on the source platform?
4. Exactly which segments were used to create this segment?

### **4. How to Calculate the Cost of Bad Data**

Once aware of data inaccuracies, it's possible to do a calculation Emodo refers to as an aCPM (actual CPM), which involves taking the original CPM and adjusting it for inaccuracies. For instance, if you were looking at two segments are both defined as BMW Intenders, you could recalculate the cost per thousand impressions based on further knowledge about that data.

In the example below, the group BMW Intender #1, which is priced at an \$0.80 CPM, has a very broad definition of what qualifies as a BMW Intender. The group BMW Intender #2, which is priced at a \$1 CPM, has a much narrower definition, but is also scale optimized and does not use deterministic data. In the final analysis, the cheaper group to use is #1.

	BMW INTENDER #1		BMW INTENDER #2	
PRICE	\$0.80		\$1	
DEFINITION	VISITED THE BMW PART OF A 3 <sup>RD</sup> PARTY AUTO SITE IN THE LAST 90 DAYS	X4	SUBMITTED NAME & PHONE NUMBER FOR A TEST DRIVE	
LOOK-A-LIKE MODELING	NO		YES, SCALE-OPTIMIZED	X3
DETERMINISTIC DATA USED	YES		NO	X3
WHAT % OF IMPRESSIONS WILL BE SERVED TO THE SOURCE PLATFORM?	40%	X2	80%	
		= \$6.40 aCPM		
			= \$9 aCPM	

**Conclusion: The Four Steps to Fixing Programmatic’s Unsolved Problem**

As is with other problems in the advertising ecosystem, marketers are finding that, with focus, data inaccuracy in programmatic can be improved. Follow the four basic steps outlined here.



1. **Prioritize:** Establish data accuracy as a top priority – equal to media quality
2. **Be Aware:** Keep in mind what goes on behind the curtain and how those steps can affect the accuracy of data segments
3. **Ask:** Seek deeper answers to revealing vendor questions
4. **Calculate:** Assess the value of data options by doing a simple aCPM calculation