i-com

GLOBAL FORUM FOR
SMART DATA MARKETING

I-COM
Whitepaper

CURRENT ISSUES IN ATTRIBUTION
& HOW TO OVERCOME THEM

COLLECTIVE WORK

2020

CONTENTS

ABSTRACT

This Whitepaper outlines, and provides recommendations, on some of the most pressing issues in attribution today. Designed as a collection of essays which can be read individually, each chapter reflects the voice of its author.

THE ISSUES OUTLINED IN THE WHITEPAPER INCLUDE:

THE CONSIDERATIONS BETWEEN SCOPE AND METHODOLOGY

Whilst scope and methodology are typically two separate topics, in attribution they are related. This section focuses on three common challenges within methodology: data integration (identity, walled garden), control design and attribution window, and how changes in these lead to different considerations and limitations in methodology.

DATA INPUT QUALITY

The data used to fuel attribution studies, disparate 1st and 3rd party data, media input and behavioral outcomes data, comes in many shapes and sizes. To ensure your results make sense and match your understanding of your business, it's important to have a passing familiarity with the trade-offs and the strengths and weaknesses of the data feeding into attribution models. That way, you can ask questions and ultimately make informed decisions on how to apply model results.

IDENTITY RESOLUTION - ATTRIBUTION'S LINCHPIN

Identity resolution (a.k.a. data matching) is the linchpin of attribution, the technique that knits together all of the siloed measures of media exposure and conversion behavior. It remains the single most challenging aspect of attribution. When it is not done well, attribution models provide biased guidance to marketers that can result in substantial misallocation of funds, underperforming campaigns and low marketing ROI. Informed marketers can avoid such an identity crisis.

CHOOSING THE RIGHT ATTRIBUTION VENDORS

When determining the attribution or measurement vendor you should partner with, you need to ask as many questions of your own organization as you do the vendors. By understanding what you are trying to achieve, what data you have available, and what business questions you need to answer, you are halfway there.

MEASUREMENT OUTSIDE OF E-COMMERCE

Measurement solutions for non-digital conversion outcomes are not unattainable. With a solid approach and proper due diligence, you or your vendors can create tools to help achieve accurate measurement. This section focuses on key questions that can help your investigation of options for conversions that are not as simple as an e-commerce purchase or website visit.

MISUNDERSTANDING OF 'ROI' IN ATTRIBUTION

ROI is complex; there are a huge number of factors that influence the measured outcome and it is based on what has happened, not an absolute indicator that it could never be improved. In this light, there are a few golden rules that can be kept in mind when using marketing effectiveness results in order to get the best out of them:

- Keep the context at the forefront.
- Triangulate across multiple methods for the best results.
- Understand what is truly incremental.

1 | INTRODUCTION TO ATTRIBUTION

Attribution, at its core, is about giving the right credit where it is due. We are at a point with marketing and advertising technology where we can see what content individuals have seen, and then understand – did they take action or not? and if so, what treatment(s) seemed to drive that action?

Right now Attribution is one of those tricky subjects that marketers find themselves disillusioned with.   When the first attribution tools were

MELISSA GRADY
CMO at Cadillac, USA

coming to market, we believed that the holy grail of digital, and data-driven marketing was coming to fruition. However,as we moved into implementation, and the management of these systems, the issues became apparent. The difficulty of implementing these systems and the manpower necessary to do so could be somewhat overwhelming. In addition, while few can afford not to work with the walled gardens such as Facebook or Google, the fact that they are not willing to make their user-level data available in a third party system makes it extremely challenging for us to manage the campaign across the ecosystem and measure the outcomes in the same way as we would do with other publishers. As a result, we often find ourselves modeling on top of the outputs then optimizing results.

Where this leaves us is not in a place where we should abandon attribution, but where we learn how to harness the data and address the challenges that face us.

> This paper walks us through the issues we all face every day – data quality, the landscape of vendors, and how to address data outside of the more easily measured ecosystem. It also looks at how to holistically approach measurement and to understand the role that attribution plays.

There is also the topic of cookies and the move to a cookie-less world that we will not address in this paper. As this very new information is being absorbed, different partners and vendors are coming up with very new and innovative ways to work through and around cookies. What happens in the wake of the industry move away from third-party cookies is still yet to be seen, and will most likely be the subject of a different whitepaper.

Another topic that lies outside the scope of this paper are the challenges specific to the attribution tools within walled gardens. While these large providers can solve the identity within the 'walls' of their platforms there is limited visibility outside the walls, therefore most third party vendors won't be able to measure the outcomes on the user level and provide actionable insights for optimization. There are also questions about the trustworthiness of using the platforms' data to measure their own effectiveness, essentially "grading their own homework". Nonetheless, as more developments in data clean rooms and transparency take place, we believe it deserves its own spotlight.

The tools and methodologies laid out in this paper are actionable today and will help address real business problems in a very tangible way.

2 | THE CONSIDERATIONS BETWEEN SCOPE & METHODOLOGY

**SABLE MI
Chief Research Officer
at NinthDecimal, USA**

In the advertising industry in general, and the measurement of advertising effectiveness in particular, the term "attribution" refers to how credit for a specific conversion is assigned (or attributed) to a collection of media touchpoints that might have influenced the consumer's conversion.

For years, "attribution" has mainly applied to digital campaigns: digital media with digital conversion. The trackability of the digital footprint was seen as the "holy grail" for measuring ad effectiveness. Whether it's first-touch, last-touch or multi-touch attribution, the touchpoint was defined as impression, click, site visitation or digital conversion - in other words, everything that lives in cyberspace.

HERE'S THE REALITY

Even the most digital-centric consumers don't live exclusively in cyberspace. Furthermore, in today's world, it is highly unlikely that any consumer will only be exposed to a single medium, or that a campaign will only use a single platform.

While there might be little dispute to the assumptions above, the complexity of applying an attribution model to a multi-platform campaign could be daunting.

To address the issues in attribution, we start with scope and methodology. Although scope and methodology are not really issues by themselves, as the first in line for consideration they often pose the most challenges when applying attribution to campaigns.

While scope and methodology are two separate topics, in attribution they are related. By "scope," we refer to the parameters or complexity of the project. Marketers must be aware of trade-offs. For instance, should the scope include both online and offline media? How about the success metrics? The combination of online and offline involves both scope and methodology.

Which consideration should come first? Even though scope is the logical choice, it does not exist in a vacuum because the trade-offs for scope often involve methodology.

THREE BASIC PARAMETERS OF SCOPE

In terms of scope, the basic considerations are:

MEDIA EXPOSURE	OUTCOME-BASED METRICS	GEOGRAPHIC COVERAGE
• Single platform or multi-platform? • Does the plan involve walled gardens? • For multi-platform, is it digital only or is it online plus offline?	• What are the KPI(s)? • Does the action take place online, offline, or both?	• Is the campaign in a single country/market or multiple countries? • If the latter, does the advertiser apply a global (centralized) or an international (decentralized) approach?

Each of the above categories would likely result in a change or adaptation of methodology. Often, a project's scope may evolve over time, adding additional layers of complexity to execution. This is especially true when the primary goal of attribution is for (in-flight) optimization vs. (post-campaign) validation. To accommodate the rapid change in technology, data availability, or even regulation, the methodology often needs to build in the flexibility for an additional scope or emerging ideas. This is where scope and methodology needs to be considered together.

A SIMPLIFIED CLASSIFICATION FOR INPUT VS. OUTCOME

While there are different schools in attribution, the basic methodology always includes exposure as input and some kind of performance metric as an outcome. Basically, we can group them into four buckets:

A Digital Media → Digital KPI
Example: desktop campaign to drive online site visitation

B Digital Media → Digital and/ or Offline KPI
Example: desktop + mobile campaign to drive foot traffic to stores

C Digital + Offline Media → Digital KPI
Example: TV + desktop + mobile campaign to drive online site visitation and inquiry

D Digital + Offline Media → Digital and/ or Offline KPI
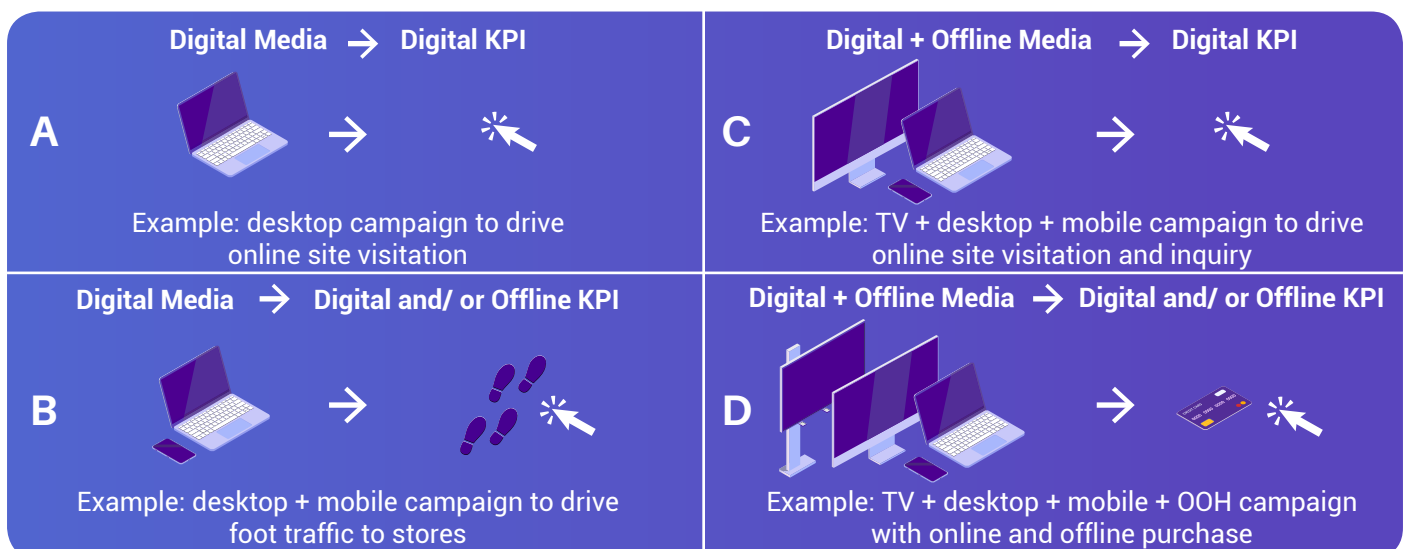Example: TV + desktop + mobile + OOH campaign with online and offline purchase

Figure 1

Each platform has its own data in different formats, collected through different channels. While on the surface these are variation in scopes, how to connect these data points in order to assign proper attribution goes beyond scope and requires an understanding of methodology.

The complexity of attribution methodology has long been a pain point among users. A common point of confusion is between

- multi-channel or multi-platform measurement and
- multi-touch attribution.

Multi-channel/multi-platform measurement simply refers to models inclusive of different impressions or media platforms, but the credit assignment could still be 100% first-touch or last-touch. Conversely, multi-touch attribution not only accommodates different channels or platforms, it assigns fractional credit to all qualified touchpoints (as opposed to single-touch attribution which only assigns credit to one sole touchpoint). Since I-COM has published a paper outlining business cases of different attribution models[1], we will focus on three common challenges within methodology: **data integration (identity, walled garden), control design and attribution window.**

DATA INTEGRATION

Applying the four categories from **Figure 1**, the possible identifiers for each bucket are:

MAID: mobile ad identifier; *UDID: unique device identifier;* *HHID: household identifier*	MEDIA EXPOSURE	
	DIGITAL	**ONLINE + OFFLINE**
OUTCOME — DIGITAL	Cookies/IP addresses MAIDs Emails UDIDs	Cookies/IP addresses, MAIDs Emails UDIDs Names Postal addresses HHIDs
OUTCOME — ONLINE + OFFLINE	Cookies/IP addresses MAIDs Emails UDIDs Names Postal addresses HHIDs	Cookies/IP addresses MAIDs Emails UDIDs Names Postal addresses HHIDs

Figure 2

[1]*For more details in digital attribution, please refer to I-COM Whitepaper: the business case for digital attribution, 2018*

Without properly connecting these data points to the right users, there can be no attribution. In fact, anyone who is deep in identity resolution will tell you that it can be more difficult than connecting hundreds or even millions of jigsaw pieces together - those will at least fit together. Here we are dealing with match rates, many-to-many matching, and the relationship between devices, individuals and households.

The long debate between deterministic vs. probabilistic approaches to identity resolution leads to no satisfactory solution. A deterministic approach relies on observing where a known, often self-reported identifier (match key), e.g., email, connects two other identifiers, such as cookie and IP address. A probabilistic approach connects identifiers through pattern recognition where two identifiers are likely to belong to the same entity, e.g., the same Mobile Ad Identifier is present at the same IP address, in certain hours, with a certain frequency throughout a period of time. The new discussion focuses on how to combine the different approaches and develop the identity graph that can leverage both methods, however to make things more complicated, each provider will often use unique terminology to describe their approach. Instead of choosing an identity solution provider by their marketing material, it is always a good idea to evaluate each provider's approach against the data available in your own study.

Another challenge in data integration is walled gardens. On one hand, walled gardens can provide a viable solution for identity and methodology, as long as the entire campaign stays within the "walls". But if you don't believe in letting the media grade their own homework, or if your campaign goes beyond the "walls", e.g., add OOH (out-of-home) or TV, then you will need a third-party solution that can access data on both sides of the wall.

Furthermore, for any advertisers that operate on a multi-national scale, one of the tough tasks is to determine whether to measure the campaign globally or locally. While the campaign itself could be launched globally, each market is different. First, even within the same media platform, the data format and availability may vary from country to country due to regulations and other constraints. Second, with different media platforms the ability to de-duplicate exposure on the user level is likely to differ. Third, there are few attribution vendors that have the capacity to operate globally with the same scale and efficiency.

On the other hand, if you are measuring each market separately, based on its unique conditions, it is hard to have a holistic view of how your campaign performs or to decide on the best way to optimize it.

Data quality aside, there are likely to be different regulations in different countries or different markets which also impact data availability and compatibility.

As you can imagine, we have barely touched the surface on data integration. More details of the challenges in data integration and how to overcome them will be addressed in depth in the Identity Resolution - Attribution's Linchpin chapter.

CONTROL DESIGN

Another key element in attribution methodology is control design. Because the base of attribution is at the individual level, how to identify the proper control in order to measure incrementality is the essence of the methodology.

Most attribution vendors apply matched control on various attributes including demographic, geographic, behavioral patterns such as TV viewing behavior, previous visitation, or purchase history. To find the proper match, different KPIs require different sets of matching criteria, and as a consequence, the results can vary widely.

A common challenge with multi-platform MTA occurs when the client is interested in comparing a subset of platforms in isolation, e.g., mobile in-app units vs. mobile browser units, regardless of the exposure on other platforms. While the underlying assumption is that the opportunity to be exposed to other platforms is equal across test and control, in reality it is not that simple. As a result it requires a different control set to eliminate cross-contamination for a fair comparison.

Another approach is randomized control design which is widely used in scientific experiments, however this requires extra attention and careful implementation in order to be applied in advertising, especially when it comes to campaigns that involve online and offline media. To properly execute an experiment, you would need to address scale, identity, and budget allocation at once, which few can afford.

ATTRIBUTION WINDOW

For MTA, there are ongoing discussions on what is the most appropriate attribution window: what are the most appropriate start and end points? Should recency be part of the consideration? And for an ongoing campaign, if the exposure takes place in January and the end conversion happens in December, can you still conclude that the January message causes the action in December?

This gets even more complicated with offline media, especially ones used for upper funnel communication. If you only measure digital media on the call-to-action messages, it would likely overestimate digital media's attribution and result in overspending on digital in lieu of upper funnel media potentially resulting in a loss of sales. However, if you over-extend the attribution window, it would likely overestimate advertising impact and result in overspending with lower ROI.

While there is no one-size-fits-all solution, it is important to understand the trade-offs when determining the scope and deciding on the proper methodology for the study. Most importantly, a consistent guideline over time is required in order to establish a baseline for long term comparison.

SUMMARY

While scope and methodology are two separate topics, in attribution they are related.

THE THREE BASIC PARAMETERS OF SCOPE ARE:

MEDIA EXPOSURE: Single platform or multi-platform? Does the plan involve walled gardens? For multi-platform, is it digital only or is it online plus offline?

OUTCOME-BASED METRICS: What are the KPI's? Does the action take place online, offline, or both?

GEOGRAPHIC: Is the campaign in a single country/market or multiple countries? If the latter, does the advertiser apply a global (centralized) or international (decentralized) approach?

The scope is a mix of these three parameters which will then lead to different considerations and limitations in methodology. This section focused on the three common challenges within methodology: data integration (identity, walled garden), control design and attribution window.

3 | ISSUES IN DATA QUALITY FOR ATTRIBUTION

ALICE SYLVESTER
Partner at Sequent Partners, USA

In this chapter, we address the data commonly used in attribution. Unlike data fueling marketing mix models which encompass market-level data, attribution data involves household or device-level data. Since there is no universal database, no single source of media input and sales or other outcome data, attribution modelers adapt and combine disparate data sources. There are limitations to the data they use, and trade-offs to be made. It's far better to understand the nature and quality of underlying attribution data ahead of time than to be surprised later, when you get results. We will delve more deeply into the root causes of lower-quality data later in this chapter, but importantly, you and your team need to know—and approve—the trade-offs that will naturally occur in data selection.

WAITING FOR PERFECT DATA?

We begin this chapter with a simple statement: there is no such thing as "perfect data".

But waiting for perfect data is not a good strategy. You will never begin. The sooner you take the first step in the journey toward deciphering losers from winners, from average content to extraordinary content, from ads that misfire for some reason to razor-sharp messages that resonate and convert, the sooner you will learn the intricacies of data and what works for your business.

That's what has led so many marketers to experiment with attribution and to encourage development and innovation along the way. But because imperfect data looks exactly like good data, it's paramount for marketing executives to recognize trade-offs and the strengths and weaknesses of data feeding into the models to make informed decisions. Data quality is everyone's responsibility.

WHERE DATA QUALITY ISSUES LURK

Data quality is everyone's responsibility. Data quality issues aren't isolated to only one step in the attribution process. All the inputs and outcome data—have strengths and weaknesses. Attribution integrates a wide variety of device-level and aggregate data—first-party CRM data, third-party target enrichment data, digital data, social data, television data—that all have unique characteristics and some flaws that must be acknowledged. It's far better to understand the limitations of the underlying data ahead of time than to be surprised by them when you get results.

THE VALUE OF A COMPREHENSIVE DATA STRATEGY

It is simpler than you think to assess data quality before it shows up in an attribution study, but it takes planning and time. Marketers must take ownership of the data issue and acknowledge the potential impact on timing and model results. Again, it is everyone's responsibility—not a task to be relegated to the agency or attribution model provider alone.

It's possible to score the strength of the data used in an attribution study, but it's better when data assessment is an ongoing process. Data stewardship is not "one and done". A carefully cultivated data strategy ensures availability of strong, high-quality internal and external data. Marketers must invest in data quality for the long haul. We know one U.S. marketer who pulled back from their attribution modeling, went dark, and spent two years evaluating and developing their internal processes and data streams that drive attribution[2].

A couple of things work against a clear picture of data quality. Attribution (and the media world at large) can suffer from excessive jargon. Jargon leads to confusion. Getting everyone on the same page with language and definitions could help. The fact that there are few standard definitions or standards for cleaning and sharing data in attribution makes the task difficult. An industry standard would be immensely helpful.

BASIC DATA REQUIREMENTS FOR MTA

HOUSEHOLD/DEVICE-LEVEL DATA: Attribution is distinguished from marketing mix modeling by the fact that it operates entirely at a household or device level, rather than a market level. Not all media is measured at the household level, and as a result, attribution has skewed toward the digital ecosystem; search, video, and social. More recently however, television and digital out-of-home have been added to attribution data streams.

DIGITAL IMPRESSIONS: Digital impressions are generally captured from server logs or identifiers such as tags, cookies, mobile ad IDs, etc. These are all being challenged by privacy policies and regulations and will be increasingly difficult to obtain. We already know of one U.S.-based e-commerce marketer who has not undertaken digital attribution because of the challenges presented by the State of California's privacy regulations (CCPR), which are similar to GDPR, the European privacy edict.

TV AD OCCURRENCES: TV ad occurrences are another key input variable in attribution. A variety of ad occurrence providers with different tracking methods (AI, automatic content recognition, fingerprinting, encoding) undoubtedly produce different levels of basic ad counts. A study is underway in the United States to determine the degree to which these methods produce different schedules and impact attribution outcomes. A comparison of the schedule of ad occurences with your own internal standard will enable you to assess strengths and weaknesses. You may find

---

[2] *E*TRADE, Attribution Accelerator, New York, 2017.*

ads of certain lengths, or from certain sources, are missing. Or you may find that, while brands are accurately identified, the featured sub-brand or variety is not.

TV EXPOSURES: TV exposure data provides the household or set-level measure of the number of people who saw the ad. Today, television data is available through set-top boxes or smart TVs, in addition to standard measurement panels. Depending on availability, these measures can provide very different ad campaign exposure levels.

MEDIA COSTS: Because attribution is a measure of the value of the media investments, cost plays a critical role in attribution and ROI measurement. Working media dollars? Non-working media dollars? But the bigger issue is negotiated rates. The syndicated sources provide an estimated cost, but each advertiser pays something different. Ideally, the ROI calculation in attribution should be based on the advertiser's own true costs.

OUTCOME MEASURES AT THE CORRECT LEVEL OF GRANULARITY: Attribution outcome measures; online behavior, off-line sales, foot traffic or online traffic, brand metrics, and customer lifetime value, all need to be aligned on a highly disaggregate level for readability in attribution.

CHARACTERISTICS OF QUALITY DATA

It's not difficult to articulate what ideal quality data is for attribution. The challenge comes in understanding the trade-offs of available, but less than ideal, data.

First, attribution data should be complete and comprehensive, with no missing pieces of the consumer journey, the ad buy, or sales outlets. This includes paid, owned, and earned digital and traditional media, plus the data walled off by certain providers. Because this is a bit of a challenge today, it's imperative that the data be representative of whatever universe it purports to cover and cover a significant part of the media investment involved in the study.

The data should be as unbiased and accurate as possible, and it should comply with industry measurement standards, best practices, and privacy standards.

Quality data in attribution should be suitable for the specific business application at hand, and should match the business goal. A marketer who wants to measure immediate response to television needs highly granular television data, and may value speed over full coverage of the geographic footprint.

COMMON ISSUES THAT THWART DATA QUALITY

There are many reasons data can be compromised in attribution studies. Because the data used in attribution is derived from multiple data sources; agencies, internal sources, the media, and other third parties, it can be disconnected and inconsistent. These misalignments typically appear in geographic, timing, and granularity perspectives. The scale of the media datasets might not

match. Brand name and sub-brand taxonomies used by the providers may vary. It's amazing how much confusion there is regarding what brand or brand extension is advertising.

The device graphs and data matching/linking processes inherent in attribution are also a major source of bias. This occurs when providers are only capable of matching back as few as 15%-25% of impressions across the data sources. Furthermore, matching devices to households also introduces bias and the processes vary by data provider. There are few industry standards and very little validation governing this phase of attribution data development.

Discerning the difference between acceptable and unacceptable data for attribution certainly can be a challenge. Many data scientists lack the domain knowledge; they tend not to come out of the media measurement world and lack an appreciation for the intricacies of coverage and comprehensiveness. We've seen significant mistakes made because of a lack of understanding about such fundamentals as the audience "universe" size, advertising carry-over, and the diminishing effects of media investment—the saturation effect where sales increases reach a limit after which each additional advertising dollar has a decreasing incremental effect. Education and awareness of the issue will hopefully close the gulf between the data science discipline and media measurement.

RISKS ASSOCIATED WITH POOR-QUALITY DATA

What happens when attribution results are based on low quality or incomplete data? Interestingly, marketers have had a lot of experience with data issues in attribution, and the consequences are relatively well known and discussed openly.

In truth, they sum up the experience this way: analytics mayhem. What do we mean by that? Model outputs that don't line up with business experience. Multiple versions of the truth from different models. Confusion about what is really driving outcomes. Bad data quality can undermine the success not only of the project, but also the credibility of modeling, the business, and its bottom line. It also undervalues the medium being measured.

HOW CMOs CAN IMPROVE DATA QUALITY SUCCESS RATES

As an industry, we've learned a great deal about the importance of a sound data strategy. Companies that invest and nurture marketing intelligence that is built around their particular business outcomes and aligned with operations and other internal processes tend to achieve greater success in measurement.

But that doesn't happen overnight. It takes management leadership and prioritization. All parts of the organization, from Marketing to IT to Finance, need to understand the importance of the data strategy and policies, which can be challenging in big organizations. CMO's need to champion data acquisition and quality in the C-suite and clear the way for strong marketing data to become a corporate priority. Without that, you're just somewhere in the IT queue.

When you hire agencies and providers, it behooves you to define your expectations and insist the team understand principles of good data hygiene, as you define them according to your business needs. It's important that your tolerance for data imperfections and natural trade-offs are clearly articulated and acknowledged by the team.

In terms of staffing, empower and trust your in-house analytics team and surround yourself with experts. Don't presume outside modeling providers automatically understand your business. Every business is unique, and you and your team must play a key role in shepherding data collection and the modeling process.

We can't emphasize enough how important it is for CMO's to be realistic about timing requirements. It's easy to strain the data collection and quality systems when undue pressure is placed on the providers. It's important to be patient. Today, data collection, cleaning, and matching can take upwards of 3-4 months depending on where you're starting from and the enormity of the task. Be realistic. Insist on enough time to review and evaluate the strength of the data inputs before acting on the attribution results. When speed is of the essence, it's essential to understand the quality trade-offs necessary and their potential impact on decision making.

Finally, it makes good business sense to have an innovation fund and a research agenda to foster new learning and continual improvements. Participate in the industry dialogue and send your team to industry conferences—the field is changing so quickly, it's easy to fall behind in your understanding of what's available and what's strong data. By actively encouraging new learning and improvements, you're sending a clear signal that data quality is important to you and the entire industry.

CONCLUSION

Data used in attribution can be disconnected and inconsistent. It can be misaligned geographically and temporally. Brand names and sub-brand taxonomies may not match. Since the data is derived from multiple data sources; agencies, internal sources, the media, and other third parties, it can be challenging. There are few industry standards and very little validation governing this phase of attribution data development. But asking the right questions and making the right trade offs will help guide you to a successful outcome.

4 | IDENTITY RESOLUTION: ATTRIBUTION'S LINCHPIN

WHY YOU SHOULD READ THIS CHAPTER

Identity resolution (a.k.a. data matching) is the linchpin of attribution, the technique that knits together all of the siloed measures of media exposure and conversion behavior. It remains the single most challenging aspect of attribution. When it is not done well, attribution models provide biased guidance to marketers that can result in substantial misallocation of funds, underperforming campaigns and low marketing ROI. Informed marketers can avoid such an identity crisis; read on …
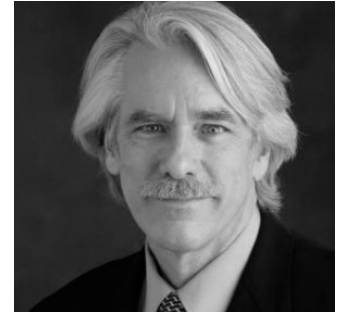
**JIM SPAETH
Partner at Sequent
Partners, USA**

PROLOGUE: ONE ADVERTISER'S STORY

From an interview with a CPG brand marketer, who decided to build their attribution model in-house

When we set out to develop our attribution model, the first step was to align all of our siloed data to represent our consumers' path to purchase. Our digital touchpoints were housed in different parts of the organization and used different tracking technology – server logs, cookies, tracking pixels. Though imperfect, our only choice to reflect our massive television investment was to cobble together various sources of set-top box data and smartTV data using street addresses and proprietary smartTV identifiers. Sales, the outcome variable, could only be represented by retailer loyalty card data identified by street address.

We were like a dad on Christmas morning, who has just opened the box of his kid's very complicated new toy labeled with "some assembly required". How do we align all of these data sets to accurately and completely reflect the consumer's path?

By the time we connected all of the dots, our "big data" sets of tens of millions of consumers boiled down to about two million for whom we felt we had most of their paths assembled. But, compared to our average consumer, these two million were quite a bit more affluent, white, older, suburban, with larger families. A rather distorted lens. For each analysis, we needed to isolate purchase occasions, but for some of our brands, we found only a handful in any one day. And when we aligned those purchases with digital or television exposures, that handful got much smaller.

Data matching, or more properly, identity resolution, was the soft spot in our entire attribution scheme. Big data became not so big data. Slightly biased data became very biased data. We need a better solution now.

THE ROLE OF IDENTITY RESOLUTION IN ATTRIBUTION

As our story shows, identity resolution is the linchpin of attribution. Myriad device or household data streams must be knit together to reflect the consumer's path to conversion. This process is more than data matching, a pairwise relationship between two data sets. The reality is that dozens of data sets need to be connected with the consumer as the hub. That requires an important conceptual shift to resolving identities. With the consumer in the center, how do we best link all of the consumer's data to her?   That said, we have to acknowledge that rarely do we know about the person behind the ID.

It's not essential for a marketer to be intimately familiar with the technologies of identity resolution; it is essential for a marketer to understand the implications of identity resolution. Nevertheless, we will take a moment here to review the basics.

Many data providers offer Identity Graphs: proprietary data sets containing all of the common identifiers, called match keys, for households and sometimes consumers. Like a central directory of match keys with names, addresses, phone numbers, email addresses, cookies, IP addresses, mobile ad IDs (MAID), etc. If your media data is identified by cookie, IP address, email, whatever, they can all be resolved to a single household using any of those match keys. Importantly, these providers can perform the identity resolution in a secure, confidential process that respects all privacy regulations. No Personally Identifiable Information (PII) is ever disclosed. Each household/consumer is assigned a unique and anonymous ID proprietary to the provider. In addition to these keys, data providers have already pulled together a virtual encyclopedia of household or personal characteristics; demographics, credit card and banking data, home ownership, car ownership and on and on.

Many of these data providers originated from direct mail, catalog, magazine subscription, or change of address applications. So they have their feet on the ground - literally. But then other characteristics data and match keys are accumulated. Some more accurately than others.

Using match keys, digital data is resolved to an identity using cookies, IP addresses, email addresses, MAIDs and the like. Each of these have strengths and weaknesses, which we will address in the next section.

For attribution, television exposure data comes from two sources; cable or satellite provider's set-top boxes (STB), or smartTVs. STB data is resolved to a common identity via name and address. SmartTVs have their own IDs for registered users who don't opt-out.

Connecting online identities to offline data has different challenges for different businesses. Your first-party data may offer a strong solution for part of your target audience. A data onboarder can use their identity graph to do a secure name and address match, protecting the PII. But that's not your entire target audience. Third-party data such as loyalty card, credit card, automotive registration data sets provide a broader view of target audiences, but need to be integrated through

identity resolution. In this case, often by matching name and address via the ID graph to your other data in compliance with privacy regulations and policies.

Where store, restaurant, showroom or theater visits are a meaningful KPI, location data providers can fill that need. Since location data is most often gathered through capture of location/GPS data on mobile devices, MAIDs are the natural match key.

DIGITAL EXPOSURES

Laptop/Desktop
Cookie

Mobile Device
MAID

TELEVISION EXPOSURES

MVPD: Cable/Satellite/Telco
STB Name & Address

Connected TV
CTV ID

OOH EXPOSURES
Location Data w/MAID

HOUSEHOLD

OUTCOME DATA

Credit Card Charges

Retail Loyalty Card Purchases

First Party eCommerce Purchases

Name & Address, Phone Number, or Email

Resolving Identity With Match Keys –
Matching Media Exposure Data to Retail Purchase Data Illustration

In each case there is a dependence on two things:

1. The accuracy of the ID Graph, especially the match keys associated with each identity
2. The accuracy of the match key from the source data.

What could possibly go wrong?

NOT PERFECT YET

Attribution is a long road with a huge promise… which makes the long road worthwhile. So where are we today and how does a marketer navigate today's imperfect version of the vision?

One important reality is, that despite the fact that we often talk about consumers, customers or people - for the most part we have data on computers, mobile phones, tablets, cable/satellite set-

top boxes, loyalty cards, auto registrations, credit cards - not people. Which is startling when you think about how much effort marketers put, rightfully, into struggling to deeply understand the hearts, minds, needs and desires of their consumers. There are services that bridge the consumer-computer divide, and good progress is being made, but it's not perfect yet.

As with most data sources, there is a range of accuracy. At one end of the spectrum consider first party data, let's say from a subscription that has your name, address, phone numbers and email. Data linked with those three match keys will resolve to a single identity fairly accurately. Although even deterministic data has flaws. But not all match keys are deterministic. Some are more probabilistic. For example, mobile and desktop behaviors are often linked when a MAID and an IP address are frequently coterminous. But that assumption fails in the case of a frequent visitor. Email addresses can become associated with the wrong identity if someone uses a shared email account for an e-commerce transaction. If you frequently use more than one computer, the IP addresses may not both be associated with your identity. These probabilistic approaches always make sense and often work. But they are not perfect yet.

The difference in data quality from deterministic to probabilistic is not binary; it's a spectrum. For example, while a name and address match may seem air-tight, what if you sometimes use a middle name and sometimes a middle initial and other times no middle name reference at all. What if sometimes you use an apartment number and sometimes not. A certain amount of judgement will apply to classifying these as the same identity, or not.

Along this spectrum, we also face the accuracy versus scale trade-off. Attribution suffers when the data sources can only be partially linked. If we only use deterministic matches, there's a limitation in scale. To get the scale required, we need to venture into probabilistic matches that trade-off accuracy to gain scale.

When we consider the various ways in which data can be linked through identity resolution, there are clear strengths and weaknesses. Cookies were part of the original design of the Internet, they helped your favorite websites remember who you are. But third-party cookies have also enabled advertisers to know that you were shopping for golf clubs last week and haunt you endlessly with golf club ads. Cookie tracking is now history. Apple's ITP (Intelligent Tracking Prevention) blocks all third-party cookies by default, preventing any cross-site tracking. Google's SameSite policy, announced in February effectively does the same thing for Chrome users. This policy has been temporarily rolled-back, but can be expected to return soon. Mozilla's Firefox browser has blocked third-party cookies for some time.   Fresh legal initiatives like the California Consumer Privacy Act (CCPA) have only exacerbated the difficulty. Identity resolution dependent on cookies has become ineffective.

Privacy regulations aren't limited to cookies. CCPA also applies to IP addresses, device IDs, in fact any personal identifier, and even inferences about personal identity. According to the Data Protection Report, CCPA "sets a new precedent with its sweeping definition of personal information." More stringent than any other current privacy regulation, it gives Californians the right to forbid

companies to sell their information to third parties. Chilling and potentially leaving marketers with even less perfect identity resolution.

WHY SHOULD A MARKETER CARE?

Does any of this matter? Nothing is perfect. Marketers have to make judgements all the time, why be so precious about the accuracy of identity resolution. Isn't it just directional anyway?

Yes, all of those observations are true, but remember the marketer's question from our prologue – after the imperfect process of identity resolution, do we have the right data to guide more effective marketing decisions and do we have enough of it? Will my attribution model give me accurate guidance as I optimize my campaign? Or has the lens become so distorted that it guides me to the wrong decisions, wasted resources and poor campaign performance?

Bias is the number one issue to consider because the data used for attribution is almost never a representative sample of the total population. For example, if we are using smartTV data, we know we begin with an upscale bias. When we match the smartTV data to digital data we may find that bias exacerbated due to differential mobility rates. If we are studying a campaign for a brokerage account, we probably have little to worry about. The consumers behind the devices in our matched sets of data are likely to be representative of the market for the service. But if we are studying a campaign for macaroni and cheese, the data feeding the model is likely to be missing a particular kind of consumer of importance to the brand; one that might respond differently to the campaign than others. In such a case, the attribution model might provide the wrong guidance, even directionally.

A recent experiment conducted by Analytic Partners under the aegis of the Advertising Research Foundation[3] simulated different levels of data loss due to possible increases in consumers' cookie opt-out choices. The experiment concluded that when data loss exceeded 50% the attribution models started to become unstable – more than 30% of cases deviated from the original model by 15% of more. When the data loss was concentrated in certain demos (18-24, middle-income, separately) slightly greater model instability was observed with only 25% data loss. While these might sound like extreme scenarios, the 2015 CIMM study cited typical match rates of 40%. So we are already often working in a 60% data loss scenario.

Nielsen's Digital Media Consortium II produced similar findings[4]. When television exposure data used to fuel the models was missing certain portions of the population, the model errors for television increased by a factor of 24, on average across three cases presented. The data loss for television was extreme, but no more extreme than the loss of representativeness experienced when shifting data sources from a nationally representative panel to a specific set of STB data.

Both empirical studies underscore the fact that as data is lost attribution models become unreliable.

---

[3]*Nancy Smith, ARF Data Loss Experiment, November 13, 2019, DataxScience, ARF*

[4]*Ross-boy Link, Learnings From Digital Media Consortium II, ARF Re!Think 2016*

A study by Neustar examined specific implications of data loss due to problems with identity resolution[5]. The authors state, "The study shows that advertisers using media channel-specific (identity) graphs are likely to see over-attribution to the media channel with which the graph is associated, and under-attribution to other channels. In practice, this would result in overspending in that media channel, while under-spending in all others." Taken to a less extreme scenario, we can infer that media that are more readily matched to other data for attribution may be more fully valued than those media less readily matched.

The study also found, "Advertisers using sparse or inaccurate graphs are likely to underestimate the importance of advertising to their business while becoming over-reliant on a perceived high base of ongoing business. Again, this would naturally lead to a reduced spend on advertising, and a resulting loss in sales, resources, and budget."

Even coveted first-party data is not ideal, not when used alone. The Neustar authors state "Advertisers who rely too much on their first-party data for identity resolution and ID integration are likely to overestimate the impact of advertising on their business, resulting in an overspend on media that is not incremental to sales."

We hardly need to add the conclusion, accurate, complete and unbiased identity resolution is essential to attribution modeling. Anything less threatens to produce incorrect guidance for marketing decisions, inefficient media allocations and sub-optimal ROI results in market.

QUESTIONS MARKETERS SHOULD ASK

The world of identity resolution lacks transparency with respect to disclosure of methodologies, best-practices and validation. But, without becoming an expert data scientist, a marketer can ask two things about the data as it stands after identity resolution:

1. Does it look like the world as I know it (representativeness), and,
2. Is there enough data left to provide the granularity of information I need (sufficiency).

In assessing representativeness, your data scientist should be able to provide comparisons between the attribution inputs and your trusted macro data sources. Choose the comparisons most relevant to your business and your goals for attribution, but here are some thought starters:

---

[5]*Robert Stratton, PhD, and Dirk Beyer, PhD, Attribution and Identity: A Codependent Relationship, ARF AudiencexScience, 2019*

OUTCOME METRICS	MEDIA EXPOSURE
In general, the micro-level outcome data used for attribution should be aggregated for comparison to the accepted source of macro-level outcome behavior. These two data sources should align reasonably well for key consumer dynamics. For example, with purchase data the following comparisons would make sense: • Weekly purchase incidence, level and seasonal pattern, especially around events like holidays. • Repeat rate • Brand share, assuming you also have category purchases in the same data set	Similarly, for each medium, key consumer behaviors can be compared, as reflected by the micro-level sources used for attribution and the macro-level sources used for planning and buying: • Gross weekly impression levels and flighting patterns • Four-week reach/uniques and frequency for brand exposures • Daily reach/Unique levels for the total medium

The sufficiency question depends on two things. First, the granularity of insights the marketer seeks. For example, do you want to understand the relative contribution of each creative execution? Or each media placement by: platform, genre, property. Or do you want to understand the context effect of each creative execution in each media placement? The finer the granularity of your question, the thinner the number of exposures of each type in your data set. Then you need to consider the incidence of your purchase, or other conversion event. The exact data requirements will also depend on the analytic approach being used, so we won't go into those details. But you can ask your data scientist for the level of precision around each of these model outputs, given the volume of data available post identity resolution.

Ask before the modeling is done so that you know how much confidence you can have in the results. Or to understand specific biases that will inform how you interpret those results.

SOLUTIONS – IS THERE A ROADMAP?

No. And that's a problem.

The current situation advantages walled gardens. Attribution models within walled gardens overstate their contribution. In digital, generally, identities are more frequently and accurately resolved than is true for offline media. Offline media are likely to be undervalued due to identity challenges. All of these scenarios are problematic for marketers who want to improve the effectiveness and efficiency of their media allocations, strategies, and tactics.

Facing increasing privacy regulation and more privacy protective corporate policies, on top of an already imperfect identity resolution landscape, the industry is scrambling toward a variety of plans for universal IDs: DigiTrust (IAB), Unified ID (TTD), Advertising ID Consortium, ID5, PubCID, to name a few. The World Federation of Advertisers is considering its own solution as well. Meanwhile Tim Berners-Lee is re-inventing the Web to give consumers even more control of their identities.

While we applaud these efforts, in the meanwhile, we advise marketers to proceed with attribution; proceed with caution and a healthy dose of skepticism.

5 | CHOOSING THE RIGHT ATTRIBUTION VENDORS

Having a clear view of what marketing efforts are driving impact to your business is immensely important. However, getting to those insights is often a daunting project with huge mountains of data in some areas, and complete lack of visibility in others. In addition, there are challenges to overcome with technology, internal adoption and modeling. It is no surprise that most experts suggest partnering with a trusted vendor who can help bring sanity and experience to the process of attribution and measurement.

JENN LEIRE
VP, Client Engagement
at Analytic Partners,
USA

1. DETERMINE IF ATTRIBUTION IS WHAT YOU ARE LOOKING FOR

Attribution is a tricky topic. Sometimes the term is tossed about as a synonym for "measurement", or it is used when specifically speaking about digital attribution technology. The first challenge at hand is likely solved by asking some simple questions:

- What am I trying to accomplish?
- Where are most of my sales?
- What data do I have access to?

If you are looking to determine the impact of digital touchpoints, with mainly online sales and online media, MTA digital attribution may be what you're after. If you are looking to assess television, digital, CRM and other media touchpoints consider an integrated approach that leverages user level data where available (e.g. linked customers and exposure data via an identity matching provider) and/or within specific platforms for further tactical learnings. The approach should be customized based on the clients business questions and data availability and Martech capabilities. But remember MTA is only one digital measurement tool in the modern marketer's toolkit. Other options abound, including experimental design and touchpoint analytics. Between walled gardens, blind spots, data limitations and newer challenges like Google's third-party cookie elimination, the standard idea of MTA is being disrupted.

If, however, you are looking to determine ROI of various forms of media, with mainly offline sales and media, you may be better served by assessing a baseline of activity and then measuring with marketing mix modeling (MMM).

And finally – and most commonly – if you are like many companies and want to have a more holistic view of measurement and have either:

- A mixture of online and offline sales and media
- Sales that are offline and media that is mixed
- Sales that are online and media that is mixed

Then Unified Measurement will likely be your solution.

While many assume unified measurement is always a blending of MTA and MMM, it is actually bigger than that. By integrating multiple approaches into a single framework, such as tactical and strategic lenses, or customer-level and business-level views, marketers can leverage a unified measurement approach. This complex approach requires close collaboration and compatible data to succeed and is rarely possible trying to integrate findings from multiple vendors.

2. ASSESS THE RELEVANT EXPERIENCE OF THE VENDOR

Experience matters. One of the reasons you work with a vendor for measurement is to benefit from their experience doing similar work. Experienced vendors will see challenges before they surface and know what questions to ask.

Not only do you want to know how long the company has been performing MMM, MTA and Unified, but you should also ask about industry-specific expertise. This is especially important if your business or sector has a lot of specific challenges. To get an idea of how well the vendor understands your business, ask for relevant case studies and references when you are further along in the discussion.

Another way to gauge experience is to access reports or speak to industry analysts about the measurement providers you are considering. Companies like Forrester and Gartner are very useful in helping you to gain an understanding of the space quickly.

3. DETERMINE A MATCH FOR THE BUSINESS CHALLENGE

Picking a vendor to help your business tackle measurement is also about finding a proper match. This can be looked at in a few dimensions:

- **STRUCTURE OF THE COMPANY:** You may want to consider a company's ownership structure and if they have any connections that would make them biased. Are executives more or less likely to believe results if the company is tied to a specific technology or data solution?
- **LOCAL OR A GLOBAL PROJECT:** Besides data challenges across regions, you will want to work with a vendor that has specific market experience and length of service.
- **DATA SOURCING:** How open is the vendor to using your preferred data sources? Are they able to be unbiased in partnerships?
- **OUTSOURCING:** Some analytics companies need to outsource or subcontract various tasks such as programming. What is their policy? Will that present concerns about data leakage or quality?

4. ASSESS YOUR METHODOLOGICAL DEMANDS

Not every company's measurement can be approached the same way. Each situation has its own challenges and specific questions to answer. First, determine how flexible the methodologies are and if there is transparency in the approach. You should also ask if the approach and models are static or dynamic – meaning, do you need to start from scratch if you want to look at your data a different way?

For specific approaches, such as unified measurement, ask how long the company has supported the technique. How granular is the company's approach? This is especially important when looking at MMM as not every MMM is equal.

If you are looking at MTA, you will need to consider the challenges from privacy laws and the in-flux access to data from the Walled Gardens, companies that do not share data like Google, Amazon or Facebook. Find out how the company handles Personally Identifiable Information (PII) data given various security, GDPR, the European Privacy Regulations and other governance issues. How does the company future-proof your analytics if access to data disappears?

5. ASSESS YOUR PROCESS NEEDS

Depending on the complexity of your program, you need to determine how much customization is available for your business situation and requirements. Is the vendor able to provide consultancy services, in addition to technology, to determine these customizations and drive analytic adoption?

It's important to remember that once your program is rolled out your team will need to interact with the analytics. How sophisticated is your staff; will they need training? Ask about the availability and depth of training and if the training is in-person or online.

You should also consider the level of ongoing account management needs. Do you need a consistent integrated team from the outset of the engagement, through the modeling process and ultimately the delivery of insights and recommendations?

Additionally, depending on how embedded analytics are across your company, you will need to determine what level of change management assistance is necessary for adoption. Ask if the company offers this service.

6. WHAT ARE YOUR TECHNOLOGICAL DEMANDS?

Technology is not always as easy as flipping a switch and getting insights. Make sure you not only understand the scope of the company's platform used for modeling and optimization, but also what your needs truly are. Will your team be working in the technology day to day, or will you simply be reviewing dashboard data?

In the age of the cloud, it is important to align on what cloud environments you prefer and which ones the vendor supports. If you have specific demands and want to work with Google, AWS, Tencent or Microsoft Azure, you may find that analytics providers may not work with all.

Typically, the most complex part of the process is handling the data so that it can be properly assessed. Ask what process the company recommends and what partners they may work with to ease the process. If you have specific data needs, you should understand if the company's product supports real-time or scheduled data ingestion via multiple disparate sources (ex. Amazon Simple Storage Service (AWS S3), Google GCP, SFTP, etc.). Once automated data processes are deployed, do the data ingestion and time requirements decrease significantly?

Ask if they have built-in ETL (Extract, transform, and load) tools that will support data pre-processing as needed and if they use other data ingestion tools, understand what they are and how do they work.

How seriously does the company take security? Do they have certifications like SOC II - an auditing procedure that ensures your service providers securely manage your data to protect the interests of your organization and the privacy of its clients? How do they manage data legislation requirements in the regions you work in?

And of course, speed is very important. Understand how quickly one can optimize and run scenarios on their platform.

7. INSIGHTS: WHAT'S ON OFFER BEYOND THE DATA?

Analytics and measurement are not just about spitting out model results. Try to understand how the company handles insights beyond the modeling work. Can the vendor show you redacted examples of the level of insights that they present in a typical engagement? Is it actionable? Can the company put the results in context and measure both the "what" and "why's" behind the drivers of business performance?

And finally, can you actually access, use and share the insights such as model visualizations exported for use in other applications or via your analytics platform?

WHAT DOES UNIFIED LOOK LIKE?

A financial services company had been increasing online media but had limited visibility into incrementality or efficiency of online activity and addressable media and lacked a clear view into the impact of addressable media on offline channels.

The brand leveraged a Unified Measurement approach that brought together a holistic business and customer view incorporating both addressable and non-addressable media, as well as non-marketing drivers.

This solution incorporated Marketing Mix Modeling (to capture incrementality and quantify the impact of online and offline marketing activities on all sales channels), and Multi-Touch Attribution (to inform tactical digital planning).

The resulting insights allowed the client to take a holistic approach in planning and reporting as well as right-time decisioning through tactical insights. In addition, offline and online projections could be forecasted together, and media plans were rethought to generate increased synergy across all channels. These insights uncovered a 23% increase in media ROI.

SUMMARY

The truth is, when determining what attribution or measurement vendor you should partner with, you need to ask as many questions to your own organization as you do to the vendors. By understanding what you are trying to achieve, what data you have available and what business questions you need to answer, you are halfway there. But don't think you need to have a firm grip on everything before you start. The right partner will act as a guide and educate your team on how to get to where you need to go.

6 | MEASUREMENT OUTSIDE OF E-COMMERCE

RYAN KING
Managing Director, Research at Samba TV, USA

THE ISSUE

In any measurement solution, to estimate lift with certainty, you need to capture all conversions. Since this is not feasible, it's important to ensure that the subset of people or households provided in the data is representative so that the inferences made from the data, can be extrapolated to the general population. If the sample set is not representative, and cannot be appropriately weighted for representativeness, there is a greater need to understand what is available within the data you have, to be able to move forward. You will then need to narrow the insights in order to focus on what portion of the market is captured in the data.

> **For purposes of simplifying this chapter, "conversion" will be used in place of purchase, orders, customer signups, footfall location, or any type of customer acquisition.**

Acquiring customer conversion data for measurement outside of e-commerce that can be used within a measurement solution, is not straight-forward or easy to accomplish. For example, measuring attribution lift for the purchase of a soft drink is not directly linkable to a person or household at the point of sale, at a physical retail location. Furthermore, soft drinks are sold in a number of retail locations, including grocery stores, convenience stores, vending machines, entertainment venues, etc. so any single point of sale measurement solution will not provide insights for the full retail landscape.

IS A MEASUREMENT SOLUTION POSSIBLE IF THE CONVERSION DATA IS OUTSIDE OF E-COMMERCE?

Very much so, but it requires a specific investigation into all the potential conversion data available to assess quality, accuracy, and usability.

Below are a few important questions that need to be answered before proceeding with attribution measurement:

- How is the conversion information captured?
- What is captured and what is missing?
- Is the sample data representative?
- What is the coverage of the purchase data in terms of total conversion/sales?
- If 3rd party, is the data directly accessible or only available behind a wall?
- How is the data stored and is the conversion data matchable to measurement tools?

HOW IS THE CONVERSION INFORMATION CAPTURED?

For offline conversions, there can be multiple methods of capturing and accessing conversion data via 1st or 3rd party platforms. Using the soft drink category as an example again, purchase history can be supplied by vendors that tap into shopper panels or loyalty card data aggregators. These services are a great way to access purchase data in a format that is more easily integrated into measurement strategies. However, any service will have positives and negatives that require an understanding of what is included or excluded (this is addressed in the next question).

Even outside of food, beverage, and CPG goods, other vendors are available for other forms of conversion including automotive sales and registration, pharmaceutical scripts, or location visitation. Whichever method is explored, the question of how the purchase information is captured must be answered in order to move forward and accept any caveats that might arise.

WHAT IS CAPTURED AND WHAT IS MISSING?

Even when you're able to capture and integrate offline purchase into your measurement strategy, you must understand what is included and more importantly, what is excluded from the data. For example, if you're acquiring soft drink purchase data from a 3rd party vendor with access to loyalty card purchase data, you need to understand what stores are covered in the solution and what percent of sales occurs in stores with these sort of programs. Please keep in mind, it is unlikely any dataset will capture all conversion points for a product but this should not deter the effort, or lower the confidence, in obtaining a measurement strategy. Having a clear understanding will help define the methodologies required to overcome any perceived gaps or shortcomings in the collection.

IS THE AVAILABLE SAMPLE DATA REPRESENTATIVE?

As a follow-up to the above question, once the sample sources are established an investigation into the representativeness of the data must be completed. Every data set has the potential for bias which must be identified in order to account for it within the measurement, adjust the insights, or weight the data for full representativeness of a population. Without knowledge of what is contained within the sample set, there cannot be confidence in the results or insights generated from the measurement.

WHAT IS THE COVERAGE OF THE CONVERSION DATA IN TERMS OF TOTAL CONVERSION/SALES?

As a follow-up to the above question, you must understand the estimated percent of conversions/sales that are tracked within the potential data sets. Once the estimated count of consumers and total sales are known, cross check with your measurement goals to ensure the population size is large enough to meet a minimum base size to test for statistical significance. For each industry or product vertical, the percent of consumer coverage required varies. For example, .01% of total soft drink sales is likely enough given the scale of the market but .01% of over-the-counter allergy sales may not be enough for statistically significant or relevant results.

IF TRACKED VIA 3RD PARTY, IS THE DATA DIRECTLY ACCESSIBLE, EXPORTABLE, AND PRIVACY COMPLIANT?

Before designing an attribution tool, this question needs to be answered in order to understand the level of access to the direct data and, if your tool can be built in-house or created by a 3rd party. With robust data privacy policies and legislation adoption around the world, access to exportable and matchable datasets have become increasingly limited. Investigation of privacy compliance is an additional step you will be required to undertake in order to review and approve your potential vendors and data collection services, ensuring the collection and use of the data is compliant both internally and externally.

HOW IS THE DATA STORED?

Where and how the offline purchase data is stored, is important to understand. Since attribution measurement entails connecting multiple data streams to the conversion data, a clear path and process to accessing and merging the conversion data needs to be established. You need a clear understanding of what is contained in the stored data, and to ensure there is a clear method and enough information to match the data with other data streams. Without matchable conversion data, the attribution measurement is not possible. As addressed in the above question - privacy policies and legislation is a big driver in how data is stored. It is important to understand if the storage of the data meets privacy standards and, if that data can be directly merged with other data streams or if it requires additional steps prior to masking certain details.

RECOMMENDATIONS

If your industry has a significant percentage of sales occurring offline, you cannot rely on a digital-only measurement strategy to assess attribution for all sales and generate meaningful insights. Answering the questions in this chapter and following the advice and best practises throughout this paper can help get your attribution tool up to speed and account for the fragmentation in tracking technologies with thoughtful design, transparency and most of all, good science.

7 | MISUNDERSTANDING OF 'ROI' IN ATTRIBUTION

WHY A CHAPTER ON MISUNDERSTANDING ROI IN ATTRIBUTION?

**MATT ANDREW
Managing Director at
Ekimetrics, UK**

Businesses like to have certainty, to reduce performance down to individual metrics and figures that can be easily shared and instantly assessed to be able to judge success. The same is true of understanding marketing performance, especially with return on investment.

However, this desire to create a single number is difficult when it comes to ROI in the marketing performance space. ROI is complex; there are a huge number of factors that influence the measured outcome (objective, channel choice, measurement method, targeting, time of day, and size of brand to name a few) and it is based on what has happened, not an absolute indicator that it could never be improved. Because of this, without holding that context in mind, interpretation can be fraught, therefore reducing ROI to a single number and having that be the focus of discussions can be dangerous.

Ultimately this can lead to the wrong decision being made on the assumption that it came from a 'data-led' perspective. Business leaders can take action based on a seemingly robust figure from statistical modelling, but have a weaker than expected impact on their business because they didn't hold on to more of the information.

In this chapter, we focus on a number of key areas to ensure you approach decisions in the right way. We will cover how ROI is being calculated, the influences of ROI, the level of impact attribution modelling captures, and explore whether a channel or lever is truly incremental within a measurement framework. Together these will help you be better equipped to question your sources of ROI information and lead to a higher impact from the decisions you take because of it.

ARE ALL ROI NUMBERS EQUIVALENT? THE FIRST DANGER AREA

While this may seem simple, there are some important factors to consider when defining ROI. Firstly, there is a lot of flexibility around metrics used to calculate a return against, and 'return' is a broad term that doesn't automatically mean monetary:

$$ROI \text{ (any metric)} = \frac{Incremental \ units \ of \ metric}{Investment \ in \ the \ activity}$$

or

$$ROI \text{ (monetary metric)} = \frac{Incremental \ units \ of \ metric - Investment}{Investment}$$

The general ROI calculation can be used with any metric – you can view a return as leads, traffic, searches, a brand metric; any metric used in the business to assess performance. The interpretation of this can be tricky though, as there is no explicit 'break-even point' as there is with a financial calculation. These will generally not be comparable to each other, only within the same metric, and so are more suitable to in assessing the relative success of a channel or lever in driving an increase in the metric.

The monetary calculation is simpler; any ROI greater than 0 is a net gain, and less than 0 is a net loss. What is important to remember is that if you're working in revenue, an ROI of 0 will ultimately be a net loss at the profit level. Where possible, try to measure monetary returns with all costs considered to get a true reflection on profitability and net gain from a lever.

These considerations around calculation and metric are especially important when sharing results. Any ROI figure less than 1 can quickly be viewed as poor, which may not be the case depending on the metric used. Ensuring clear communication of the ROI definition and interpretation can significantly alleviate this, such as a colour-scale, legend, note-areas, etc. and **no ROI figure should be delivered without a simple and clear legend** to help the reader quickly understand how to frame this positively or negatively, even if this has been explained at the beginning of a report; pages and excerpts can be taken out and fired off very quickly, and the context needs to stay with it.

KEEP THE CONTEXT IN MIND WHEN REVIEWING RESULTS

ROI results are a complex mix of different factors interacting in a market and the context for an ROI measurement is extremely important. At its most basic, there are three main factors that determine the ROI that can be achieved:

1. **OPPORTUNITY:** how big is the market or distribution channel you're operating in? How big is your brand? What is the organic growth in demand for your categories? How much are you able to drive penetration vs. simply harvesting existing demand? These are all questions that are important to keep in mind when you think about the absolute scale of ROI.

   If you're a small brand operating in a niche market then clearly bigger spends - 6 month long primetime TV campaigns, for example – can struggle to make positive ROI returns. But for a large brand in a category that can be 'triggered' to purchase, then those campaigns could be highly profitable. Understanding this scale at the beginning, and what it means for ROI expectations, is fundamental in avoiding disappointment.

2. **CHANNELS:** the channels of investment. For some brands and categories TV might make sense. For others, a stronger focus on digital or social platforms, or video-on-demand will. Others still might respond better to radio. Let's also not forget non-media channels such as activating price and promotion through different distribution channels.

   These channels all have different demographics, usages and dynamics through

the days, weeks, and seasons. This makes them more or less suitable for different brands at different times and is another strong determiner of ROI performance.

3. **EXECUTION:** how activities have been executed across the different channels can have a strong impact on measurement. It is vital to take into account the elements of execution; the objective of the campaign, the creative, flighting, frequency of exposure, optimisation objective, etc. are all important to the context of measurement even if not broken out in the modelling approach.

    In backwards-looking methods, without understanding the executional context of an activity, it is easy to write-off channels that could, with a different execution, be strong performers. This risks a brand being pushed down a narrow path of optimizing against things that have worked in the past. Instead, it is better practice to review the performance of the channel against the objective in the current context and question if or how it could be used differently in today's environment, not yesterday's.

ROI is not a static measurement for a channel and there is no 'absolute' – it is purely a measure of the context of historical opportunity, channel selection and execution. A change to any one of these could see fundamentally different results in the future.

MEASURING WHAT YOU CAN TRACK VS. THE HOLISTIC VIEW

Any measurement approach needs to be understood for its strengths as well as its blind spots and limitations. This is key especially for attribution modelling, where many channels will be missed when giving a highly granular vision of digital effectiveness:

- Non-digital media is generally not covered in pure attribution models as user-level data is required and this simply doesn't exist in a robust, unbiased form, outside of the digital space.

- External impacts such as the negative impact of competitors, macro- and micro-economic effects, weather and seasonality, are also not included. Again, this is due to building on user-level data and tracking interactions through to a conversion action; external impacts are more of a global framing of what overall expected performance can be at any time.

- Some digital channels can also be excluded from the study either because tracking is not in place (attribution can come with a higher technical overhead from those that are not mature in their digital setup), or for publishers and ecosystems that have blocks on user- and cookie-level data.

There has been innovation in the attribution modelling space to close some of these gaps but these too have their own challenges. Some vendors have taken a probabilistic approach to channels that do not have user-level data, modelling TV, out-of-home, and others down to an individual level. Here, the specific approach taken can be different for each vendor and is nearly impossible to actually verify so should be approached cautiously. There have also been data innovations in this

space with stronger data availability for some offline channels like addressable TV, but these are limited in scope and scale to the population they cover.

What this boils down to is that ROI in attribution can be overstating the contribution of digital by ignoring these other impacts, and can create challenges when comparing results with other marketing effectiveness methods that are showing different results.

To combat this, the blind spots and limitations of each analytics method used needs to be kept in mind when comparing results, with each approach being used for its best purpose. In the case of attribution this is generally for the shorter-term optimisation of budget across digital channels only at the hourly, daily or at most weekly level.

UNDERSTAND THE BREADTH OF THE IMPACT BEING CAPTURED BY THE METHOD

The ability of different measurement approaches to capture varying effects is vital to consider when assessing the ROI calculated by each method, and the granularity behind them you should expect and reasonably push for. Marketing effects can happen immediately, such as direct response advertising or paid-search, or they can happen over a longer period of time. This could be through first raising awareness and perception of a brand, triggering a need for a long purchase cycle product, or maintaining a strong position in consumers' minds. Marketing actions can have very different measured responses depending on whether the measurement technique considers long-term or short-term impact.

There have been a huge number of papers and studies on marketing analytics for as long as marketing has been around. One more recent example, which also looks to break down the breadth of impact being captured by common methods, is 'Profit Ability: the business case for advertising' (2018) by Gain Theory and Ebiquity. In this, they estimated that attribution modelling captures 18% of total profit impact vs. 42% for short-term econometrics (Marketing Mix Modelling) and 100% for long-term studies, such as Unobserved Components Modelling.

With this in mind, we can see that attribution modelling is likely to be more favourable to direct response channels vs. brand-building channels; anything that drives the shortest response impacts. This is not a problem, just something that needs to be kept in mind when deciding how attribution modelling should be used, and what decisions to take from it.

Long-term studies are still a method of divergent practice and uptake in the marketing and consultancy community. There is a clear understanding that long-term effects exist, but a number of different approaches seek to capture and explain them. It also gets harder to be granular in channel measurement when you look over longer time periods. Even if long-term effects can be captured, it may not be able to provide the tactical guidance to a brand that shorter-focus approaches can. For a recent view of the state of long-term measurement, see Sequent Partners'

(and co-authors of this paper) 'Evaluating Long-Term Effects of Advertising' (2014), developed for the Council For Research Excellence.

What needs to be understood from this section is that there is no 'silver bullet' method that gives you the full picture; attribution is just one tool in an arsenal to tackle marketing performance and is best used when focused on tactical optimisation at a granular time level.

IN THE PATH DOESN'T ALWAYS MEAN IMPACTING THE PATH

Marketing analytics seeks to understand the actions that drive an incremental purchase, one that would not have happened without that action impacting a consumer. Multi-touch Attribution (as opposed to last-click, or rules-based) works by understanding the touchpoints that consumers have interacted with and, using a mathematical approach, assigning their value or weight in the purchase (and there are a number of different algorithms used in the space). However, one element of measurement that is often overlooked is whether the interaction is truly incremental or just in the path by default.

Take search as an example. Search ads are only served on expression of consumer interest. You search for a product, or a brand, or just inspiration, and your search engine of choice delivers ads and organic results to help you with your query. These ads tend to be highly correlated to sales, as it is a consumer already on a purchase journey and displaying clear intent. Even more, if someone searches for your brand, they are clearly displaying an affinity in purchasing directly from you. The Brand search ad may be in the purchase journey, but its incrementality is a question.

To take another angle, non-tracked channels can drive a significant uplift in search activity, which would then receive the credit in an attribution model. For example, TV spots could drive direct searches for a product, but these TV spots would not be captured by the attribution model itself, only the search returns. Does the search ad get the credit, or the TV ad? We know the answer from a real-world perspective, but in the data available to an attribution model this isn't the case and there is a key challenge around accounting for the incrementality of non-tracked activity that has a synergy impact on tracked channels.

So, beyond making sure you understand the ROI definition used, have the context of the measurement in mind, understand the measurement approach and how it should be used, there is still a need to think hard about whether the activity is truly incremental or just in the path to purchase. This is absolutely critical; if you move activity from something that drives upper funnel movement to something that only works on lower funnel (like brand search) you are likely to see the upper funnel decline over time and your other channels, where you've moved budget to, become less effective.

SUMMARY: UNDERSTANDING ROI IN ATTRIBUTION

To help bring this all together, there are a few golden rules that can be kept in your mind when working through marketing effectiveness studies and results:

KEEP THE CONTEXT AT THE FOREFRONT

- Optimisation tools can be dangerous, as they strip out that context and only rely on what has been achieved in the past, not what could be achieved if things were approached differently.

- Understand how and where things could be done better, rather than just dropped.

- Comparing across different sized business units/benchmarks can be confusing as the scale of expected response can be completely different – focus on your own business and ones of similar sizes.

TRIANGULATE ACROSS MULTIPLE METHODS FOR THE BEST RESULTS

- Attribution can be used best in shorter-term, tactical optimisation.

- Other methods provide different views of ROI that should be interpreted and used in different ways. All quote ROI, but all the figures mean, and come from, different things.

- The best business decisions will come from triangulating responses across different methods to take a total view, and this is where the CMO and marketing team should steer the business from. They can then pull out the key figures and facts as needed to support their decisions to the broader business, where absolute figures are desired.

UNDERSTAND WHAT IS TRULY INCREMENTAL

- Weighting spend towards high ROI, but not truly incremental, channels can lead to an opportunity cost for further growth.

- Tease out true incrementality using the full context keeping in mind where in the purchase journey a channel is predominantly acting.

- Understand that ROI from end-of-funnel activity might not be as important as ROI at top-of-funnel activity and that it is not a straight switch to move budget across this spectrum.

8 | EIGHT ACTION POINTS TO BETTER LEVERAGE ATTRIBUTION FOR IMPROVED BUSINESS PERFORMANCE

SABLE MI
**Chief Research Officer
at NinthDecimal, USA**

JIM SPAETH
**Partner at Sequent
Partners, USA**

As you surely know by now, with the increasing complexity of media platforms and KPIs, attribution can feel overwhelming. To keep it actionable, here are our suggestions for how a marketing leader can master this essential marketing capability:

1. Before jumping into any specific attribution model, check your business KPIs and make sure that your attribution outcome variable aligns with the KPIs that drive your business performance. Attribution, while useful, may not be what you need.

2. Be mindful of the trade-offs in scale and methodologies. The broader the scale, the more complex the data collection, identity resolution, and the possible gaps in methodology will be. There's no one-size-fits-all solution; certainly no perfect solution. Understand the limitations of your data and the attribution method to ensure they are both feasible and meaningful to your business goals.

3. While there's no perfect data, a well-planned data strategy is the first step towards a successful attribution study. Clearly define your expectations and insist your agencies / service providers understand the importance of data hygiene as well as your tolerance for data imperfections and trade-offs. It's equally important to know what data is available as well as what's missing.

4. If the conversion is tracked via third party, make sure the data is directly accessible, exportable, and privacy compliant.

5. Unless your campaign and measurement are based on the same identifier without additional input, e.g., an email message that asked the user to click on a link where the same email is used for registration, it is more than likely that identity will play an integral role in the attribution study. Check how the identity graph is connected: both accuracy and fair representation are equally important. Ask yourself:

   - Does it look like the world as we know it?

   - Is there enough data left to provide the granularity of information I need?
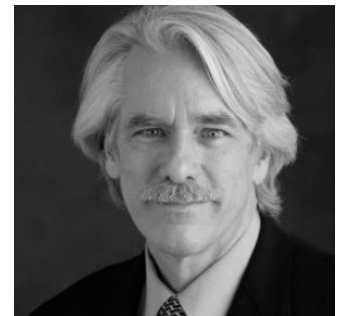
6. When choosing an attribution provider, ask the following questions:

   - Does the provider have the experience and industry-specific expertise to guide you to enhanced in-market business performance fulfilling your needs?

   - Does the provider have the proper coverage and scale to address your business issues for your needs?

   - Can they accommodate your specific data needs from sources to process?

   - Can they execute the study without any bias from either the data or the analytical technique?

   - Is the study measuring true incrementality and allowing you to triangulate the results across different methods for actionable responses?

7. There is no absolute ROI. It simply provides a measure of the relative contribution in the context of historical opportunity, channel selection and execution. Any change or omission in one of these would fundamentally alter the results; different methodology would inevitably impact or even bias the results. It's important to remember that in the path doesn't always mean impacting the path. Have the context of measurement in mind and measure true incrementality for meaningful ROI.

8. When analyzing results, keep in mind that

   - No ROI, or lift, factor should be delivered without a simple and clear legend.

   - Optimization tools can be dangerous as they strip out the marketplace context which could lead to misguided actions or missed opportunities.

   - Understand how and where things could be done better.

   - When comparing results, focus on your own business and ones of similar industries, sizes and goals.

There is no silver bullet in attribution. Understanding the tradeoffs of **scope and methodology**, **data quality**, **identity**, **choosing the right vendor** that can **measure the outcomes beyond e-commerce**, and provide **true ROI** is no easy task. However, when done right, it can help you properly assess your media investment and grow your business in a highly competitive space. We hope this paper has provided some clarity and makes the path to attribution less bumpy and more rewarding.

GLOSSARY

ADDRESSABLE TV	Television technology platform able to serve different targeted ad content to different audience segments watching the same linear TV program on Internet-enabled TVs and set top boxes (STB) in live, playback or VOD mode.
ATTRIBUTION	The statistical method of assigning credit to the media stimuli consumers encounter along the path to "conversion" – taking action, sales, etc. – a "bottom up," consumer- and transaction-level model.
AUTOMATIC CONTENT RECOGNITION (ACR)	Refers to the capability of Smart TVs and mobile apps to identify programs and ads within its proximity using audio signatures, video or a digital image. Based on sampling a portion of the audio or video, the content is processed and compared to a source service that identifies the content by its unique characteristics, such as audio/video fingerprints or watermarks.
CONVERSION	A record of purchase, order, customer signup, physical visit, or any type of customer acquisition that can be tied into a device, user or household.

COOKIES, TRACKING PIXELS	Small text file or HTML code stored by a browser used to identify individual users or track user behavior and conversions.
CREDIT ASSIGNMENT - FRACTIONAL	For multitouch attribution, each touchpoint will be assigned partial credit to a conversion.
DATA MATCHING	The task of identifying, matching and merging records that correspond to the same entities, typically households, or devices, from several databases using methods such as personally identifiable information, IP addresses, etc.
DETERMINISTIC	In the context of identity, deterministic refers to a binary approach where two identifiers are connected through a known (self-declared) match key such as email.
DEVICE ID	A distinctive identifier associated with a smartphone or other connected device. Device IDs are separate from hardware serial numbers. Important in data matching and assigning devices to households or individuals.
DIFFERENTIAL MOBILITY RATES	The rate at which people change addresses (move houses).

ECONOMETRICS

Statistical models that specify the statistical relationship between variables.

ETL TOOLS

Extract, Transform and Load. The process in which Data is extracted from data sources and transformed into a format for storing and use.

FIRST-PARTY DATA

Loosely defined, first party data is information owned and collected by one entity from a group of people with whom they have a relationship. Could be directly from a panel, audience or CRM (customer relationship marketing) data. Typically appended with third party data, publicly available or behavioral data.

GRANULARITY

The level of detail considered in a model. The greater the granularity, the deeper the level of detail and potential for actionable insight. Granularity can also be a solution for collinearity.

IDENTITY GRAPHS

An Identity Graph (ID Graph) is a database that stores all identifiers that correlate with individual consumers. These identifiers could be anything from usernames to email, phone, cookies and even offline identifiers like loyalty card numbers. Identity graphs are the backbone of data matching.

INCREMENTALITY

In modeling, sales that would have occurred without any marketing efforts are considered base sales. Incrementality reflects the sales lift associated with media/marketing stimuli. Important to distinguish between the two to avoid misattributing to a medium, the value of sales that would have occurred naturally. Not measurable, this is a model inference.

IP ADDRESS

A unique string of numbers that identifies each connected device (e.g., smartphones, tablets, television sets) using the Internet protocol to communicate over the Internet. Useful when IP addresses are static, for matching data to households for targeting purposes. But when IP addresses are dynamically assigned, they do not provide persistent identities.

LONG-TERM EFFECTS

Cumulative effect of advertising on consumers' brand choice behavior, lasting over several years. Measures loyalty to a brand or consideration to a set of brands. It can also reflect customer lifetime values. Lacks consistent definition and, in some cases, long-term effects of digital advertising are measured on a "next quarter" basis.

MANY-TO-MANY MATCHING

Aggregate groups are created with the average characteristics of group A being similar to the average characteristics group B, but with no other restriction placed on the size of groups A and B.

MATCH RATE

The percentage of the overlap portion between two datasets that can be found with a common identifier such as MAID or cookie.

MMM

Marketing Mix Modeling
Refer to:
https://www.i-com.org/white-paper-2018

MTA

Multi Touch Attribution

PROBABILISTIC

In the context of identity, identifiers are connected through pattern recognition with various degrees of certainty.

RANDOMIZED CONTROL DESIGN

Popular in digital analytics but an elemental research approach involving creating random test and control groups as a way of determining the behavioral lift (e.g., visits, conversions) associated with exposure to a specific campaign. The estimation of the measured effect is only as good as the controls associated with assigning subjects to each exposed vs. unexposed condition.

SMART TV	A digital television that is an Internet-connected, storage-aware computer specialized for entertainment. Smart TVs are available as stand-alone products, but regular televisions can also be made "smart" through connected devices and set-top boxes that enable advanced functions.
STB	Set Top Box
THIRD-PARTY DATA	Any information collected by an entity that does not have a direct relationship with the user the data is being collected on. Oftentimes, third party data is generated on a variety of websites and platforms and is then aggregated together by a third party data provider such as a DMP.
UNIFIED (IN THE CONTEXT OF MMM, MTA)	In measurement, a unified model integrates different approaches such as strategic marketing mix modeling and tactical digital analytics into a holistic model.

VALIDATION

A measure of the accuracy and precision of modeled results. There are two common and complementary approaches. Goodness of fit (MAPE: Mean Average Percent Error, or R^2: percent of variance explained) describes how well the model replicates the historical data to which it was fit. Predictive validity: the same statistics can be used to evaluate how well a model replicates hold-out or future data not used in the original model fitting process. In essence, the extent to which modeled results are well-founded and correspond accurately to real world results.

WALLED GARDENS

Publishers who are unwilling to share their user-level data with a third-party outside a clean room