

The four big forces conspiring to ruin one's analytics

Received (in revised form): 1st July, 2021



Jeff Greenhouse

Vice President of Subscriber Growth, AMC Networks, USA

Jeff Greenhouse is Vice President of Subscriber Growth at AMC Networks. He is a veteran marketer and analytics expert with 25 years' experience. He founded an award-winning digital agency in 1995 that he sold in 2009, and co-founded a pioneer in viral marketing and promotions that was acquired by About.com in 2000.

E-mail: jeff@jeffgreenhouse.com; Website: <http://www.jeffgreenhouse.com>

Abstract Despite major advances in tracking, attribution and data augmentation over the past decade, or perhaps because of them, several forces now threaten to degrade the ability of marketers to measure their campaigns and understand their audience. A growing privacy movement is resulting in government regulation and technical restrictions from 'big tech'. Walled gardens are growing larger and more numerous, forcing marketers to rely on limited data without third-party verification. Consumers are becoming increasingly fragmented across a sea of connected devices, and the minimum viable product development philosophy is creating a legacy of 'data debt' for many companies. This paper explores these four threats and discusses strategies that marketing and analytics teams can use to attempt to counter them.

KEYWORDS: privacy, attribution, tracking, compliance, retargeting, data quality, analytics strategy

INTRODUCTION

Marketing is both an art and a science, but the science part only works properly if marketers have access to — and the ability to analyse — the right data.

The world of marketing analytics has come a long way over the past 25 years, and marketers now have more information at their disposal than ever before. At the same time, however, the life of a marketer is not all roses and rainbows. The dream of perfect attribution and zero wasted spend keeps dancing into the future. It comes closer with new technologies, but then gets pushed further away with new challenges.

Looking ahead in 2021, there are a number of major forces impacting the

marketplace and making it harder to optimise campaigns, marketing spend and products. This paper looks at four of the big forces that, either on their own, or in conjunction with the others, are limiting how effectively marketers can plan, execute and measure their campaigns.

PRIVACY CONCERNS

Probably the single, most important concept that enables marketing analytics is the ability to attribute results to campaigns and other marketing activities (content included). Knowing which user was exposed to which campaigns and how they convert to value for the marketer (whether by purchase,

engagement or other methods) is core to the very concept of data-driven marketing. Unfortunately, that need to know who the audience is at every touch point has run headlong into a major public movement to protect consumer privacy.

In the Bible, Adam and Eve are banished from the Garden of Eden after eating the fruit of the tree of knowledge, despite being forbidden to do so. As a result of their failure to resist temptation, they — and all their descendents — are punished for evermore.

This is the story of the online data privacy movement.

In the early days of the web, there were just simple HTTP cookies and web server logs. The cookie allowed websites to remember something about a visitor, and the server logs could be parsed and analysed, largely to provide aggregated statistics. There were no rules or laws governing these technologies, but each website was an island, and the data tended not to move around.

As the web grew in size, complexity and utility, however, the technologies powering it exploded into a byzantine labyrinth of server calls and interconnections. It was a gold rush in true Wild West fashion, with software developers coding and re-coding in search of the holy grail: perfect knowledge about each individual user.

The idea that a marketer could track, store and analyse every purchase, web search, social media comment and website visit at the individual level and then use the data to deliver the right advertisement to the right person at just the right time was tantalising. This was the forbidden fruit in the garden of the web, and it proved far too tempting for marketers to resist.

To reach ‘the apple’, the industry developed sophisticated software systems to gather, aggregate and utilise consumer data. It also developed business models to incentivise the collection and sale of data.

At first, the data was largely used for onsite personalisation, customer relationship management and for brands to target

their customers in owned media. By the beginning of the last decade, these technologies expanded further out into the wilds of the web and the concept of retargeting emerged. Consumers started to notice the almost eerie way that advertising began following them around the web. A 2013 survey by MarketingLand¹ revealed that 75 per cent of consumers were aware that they were being actively retargeted.

As this ecosystem of advertisement targeting and delivery mechanisms was evolving, something else was developing that would have a major impact on the landscape. A public increasingly addicted to social media and sharing all aspects of their lives were unknowingly sharing with more than just their friends. Facebook’s app platform offered developers the power to tap into the social networks of people using their apps. Some games harmlessly used this power to pull in more players and generate in-app item sales, but other developers began using these tools to harvest user data. Quizzes and personality tests were secretly building demographic and affinity profiles that could be used to target and influence these users.

This shadowy trend came to a head with the Cambridge Analytica scandal,² when the UK-based company exploited a weakness in Facebook’s Open Graph interface to extract the data and interests of not only the app’s users but all of their Facebook friends. This extracted data was then used by several political campaigns to exert influence over the 2016 US election and the Brexit referendum.³

The revelation that people’s private data had been harvested and used to manipulate them reverberated around the world, adding fuel to a growing mistrust of the advertising and ad-tech industry.

Even before the Cambridge Analytica scandal, pressure had been growing to protect consumer privacy in Europe. The culmination of this effort was the enactment of the EU’s General Data Protection Regulation (GDPR) in April, 2016.⁴

This tough new regulation imposed significant compliance requirements on any company capturing personal data from any EU citizen. With the risk of major fines (18 per cent of companies responding to a 2018 Veritas survey⁵ feared noncompliance could drive them out of business completely), compliance with these new regulations required significant changes to many companies' business practices and came with significant costs running into the millions of dollars.

With GDPR officially enacted, and the public anger over Cambridge Analytica, it was almost inevitable that privacy legislation would also appear in the USA. Sure enough, in 2018, the State of California passed the California Consumer Privacy Act (CCPA), giving companies that collect and use personal data another law to comply with.

While the CCPA was similar to GDPR, it was not identical. Just two years later, the state passed the California Privacy Rights Act (CPRA), modifying and expanding the scope of the CCPA.

These laws do more than just mandate stricter security and more limits on the ways that the personal information can be used. They actually begin to eat away at what companies can gather and store. For instance, the CPRA prohibits businesses from retaining personal information 'longer than absolutely necessary'. This is a very tricky phrase to interpret. As a marketer, how long is it 'absolutely necessary' to retain the fact that a customer searched for 'size 8 jeans'? Every integration and data pipeline will carry the risk of a slip-up and hefty fine.

This movement is not stopping at California either. In March 2021, the State of Virginia passed the Consumer Data Protection Act, modelled in part on legislation being hashed out in Washington State.

Several other states are also debating data-privacy legislation, with some of the current proposals even more restrictive than the laws currently in effect. If this trend continues at

the state level (rather than federal), then the costs of compliance will continue to rise as companies have to tiptoe through a tangle of similar (yet different) rules and regulations.

The legal and regulatory changes are not the only manifestation of the privacy movement. The industry is now facing several significant technological moves being forced by the big tech companies in an effort to limit the potential for bad behaviour. Apple is bringing the issue to the forefront with iOS 14.5's App Tracking Transparency,⁶ forcing app developers to directly ask users for permission to track them outside of the individual app. Aside from the impact this may have on advertising networks, it also serves to further educate the public on their right to control their own data, and this mindset will likely bleed over into other areas of data collection.

Along similar lines, the major browser developers will all be ending support for third-party cookies by 2022,⁷ essentially removing the primary method that software systems use to track consumers from site to site. On the analytics side, while this change will not affect data collection on a marketer's own properties, it may severely impact the third-party data often used to augment first-party data.

The consumer will continue to gain more power and knowledge in relation to their data and who has access to said data. The privacy revolution is not finished yet, and businesses must adjust to these new realities. To do that, marketers and their analytics teams need to:

- *Review the personal information they are collecting, how they store it, where they send it and how they use it:* Among other things, this might entail:
 - maintaining a document that lists all of the data being collected, where data are being stored and who is responsible for maintaining and safeguarding said data;

- talking to business intelligence, IT and data teams and clearly communicate current and anticipated use cases for any personal information — this can prevent unexpected disruptions and help everyone prepare for future changes; and
- discussing the potential impact if personal information were no longer available and prepare a 'Plan B'.
- *Evaluate their reliance on third-party data:* Marketers must challenge their data providers to speak to how they will continue to provide these services in the post-cookie world.
- *Look for ways to improve their first-party data:* This often involves strengthening relationships with visitors and customers and building a level of trust with them. Marketers must remember that everything is a transaction, and consumers will expect value in exchange for their information.

WALLED GARDENS

The entire advertising industry is based around the concept of access to audience. Being able to offer that access, in as focused and controlled a way as possible, is what allows publishers and platforms to draw revenue out of that audience.

While early systems were mostly web-based and 'open', allowing outside services to run advertisements and measure the exposure of those advertisements directly, this began to change in the past decade as major platforms like Facebook and Google began to lock out external services in favour of their own internal advertising platforms. Within these 'walled gardens',⁸ the platforms exercise complete control over data collection, advertisement scheduling, delivery and measurement.

This is not necessarily a bad thing. These internal platforms offer tremendous control in advertisement targeting and audience segmentation, and they generally guarantee

that advertisements are going to appear and behave in the correct way. That said, advertisers find themselves totally at the mercy of the platforms when it comes to the timing, methodology and accuracy of reporting, with potentially significant impact. In 2016, for example, it came to light that Facebook had been inflating its 'average time spent watching' metric⁹ for its video advertisements by 60–900 per cent for nearly two years before rectifying the issue and notifying the public. Arguably, had the platform provided any level of outside measurement on these advertisements, that mistake would have been spotted very quickly by the millions of Facebook advertisers.

These platforms also face limitations in conversion optimisation due to their tunnel vision on their own participation in the conversion funnel. In most cases, conversion optimisation is based on placing the platform's pixel on one's conversion success page, signalling back to the platform that a conversion has occurred. Each of these pixels operates with its own particular attribution window, independent of any other platform. With most marketers running campaigns across multiple platforms, often touching the same target audience many times leading up to a conversion, it is common for the same conversion to be given full credit by more than one platform.

Increasingly, publishers are playing the 'walled garden' card,¹⁰ creating their own tightly controlled environments to maximise the value of their audience. They are using paywalls and registration walls to build up an authenticated audience that increases their ability to target cross-device and build valuable user profiles. The recent crackdowns on third-party cookies have only accelerated this trend.

The majority of marketers will not be able to have their own walled garden. Rather, they will have to play in the gardens of others. Analytics teams must get creative to try to validate the campaign and

conversion data these platforms provide. They are all motivated to put the most ROI-positive spin on the results their platforms generate, but marketers should do everything they can to confirm the true value of their spends there.

For example, consider creating a segment of visitors with conversions attributable to one platform (Facebook, for example), then examining that segment to see how often those same visitors also came through with a campaign tag from other paid platforms (eg Google Paid Search). This method allows the marketer to adjust their own calculation of performance for each platform to get a truer sense of cost per acquisition and ROI.

PROLIFERATION OF PLATFORMS AND DEVICES

The world was arguably a much simpler place when everyone had one computer that was plugged in, and everything happened through an owned-and-operated website. Those days are now far in the rear-view mirror — US households now average 9.5 installed and connected devices, up from 8.5 in 2020.¹¹ These devices range from smartphones and tablets to computers to connected television devices to game consoles. Most of them support native apps as well as web browsing. Consumers are becoming increasingly adept at switching between devices, syncing bookmarks, Airdropping links and using multiple devices simultaneously.

What this means for marketers is multiple points of failure in the most critical aspect of analytics: the ability to match behaviour to advertisement exposure and attribute results to campaign delivery.

Simply put, if my behaviour (the cumulative total of ‘me’) is spread out across all of these different devices, then the analytics of the sites and services with which I interact can reflect only fragments of me, thus degrading the quality of the data.

The magnitude of this issue is hard to measure (if it could be measured easily, it would not be such a problem). Anecdotally though, one can look at the cross-device behaviour of known, authenticated users to get a sense of it. At AMC Networks, the streaming services typically involve web and potentially five or more native apps deployed on different platforms. While a majority of users tend to stream on the same platform they used sign up, a meaningful percentage of them also stream on additional platforms. As a large portion of AMC’s marketing efforts take place on web, e-mail and social media (platforms that favour usage on certain devices more than others), it is possible to observe that a significant quantity of signups on streaming devices are driven by campaigns that are not connected to the actual conversion event.

Finding a solution to this problem is a thorny challenge. Both Adobe¹² and Google¹³ offer types of cross-device tracking; however, these are most effective when dealing with known (ie logged-in) users. The challenge for marketers is that they are most often concerned with anonymous prospects.

What is more, solutions to this challenge tend to run contrary to the privacy movement discussed previously.

Anonymous users will continue to push to remain anonymous unless brands provide some incentive for them to identify themselves. Analytics teams must look towards a future that is more ‘person-based’ than ‘device-based’, and they must work with marketing and product teams to make that future a reality.

Features such as watch lists, favourite lists, reminders and even content hidden behind a registration wall all provide value in exchange for the user identifying themselves and allowing the brand to connect the various devices they use to access that content. It is important to remember to always provide extra value to offset any friction that registration/login adds to the user experience.

THE PHILOSOPHY OF MINIMUM VIABLE PRODUCT

The term ‘minimum viable product’ (MVP) was coined in 2001 and has become increasingly popular in the software and product development world. It has come to be favoured largely because it allows organisations to get a product to market faster and prove out their assumptions about consumer demand before committing a major budget to a full development process. In its purest form, it is really intended to test out a core assumption at the very start of a new product, after which it is decided whether or not to turn it into a full release. But much in the way that billion-dollar, decade-old companies are now still called ‘startups’, and the way that Google kept Gmail as a ‘beta’ for five years,¹⁴ many companies have now defaulted to the MVP development approach even for projects they are 100 per cent sure they will be releasing and maintaining.

At the beginning of the MVP process, the various stakeholders come forward with their desired features and functionality (marketing and analytics support included), and the product development team takes a hatchet (or a chainsaw) to the list until it is whittled down to what they believe is the bare minimum required to prove that the product works, and the target audience accepts it. Once that has been proven, they circle back around to define the next iteration of development and map out when (and if) all of those other initial features will be incorporated.

It is this initial paring down and subsequent rebuilding of the feature set that can spell disaster for the analytics programme. Changes to the information architecture or underlying data structure between the MVP and the finalised product can create mismatched and incompatible data in one’s analytics platform, making long-term reporting and comparison difficult or impossible.

Further, the ‘MVP’ approach to analytics implementation may leave significant gaps in one’s capacity to analyse marketing campaign results. These initial decisions in the first build can inadvertently lock the organisation into structures that are far from optimal and not future-proofed (the Y2K bug being a case in point).

The product development team is focused on the quality of the present and future of the products, but the analytics function is heavily reliant on looking at the *past*. The product team can always improve the future product, but they can never recreate data that were not captured in the first place.

So, what can be done to counter this threat? There are two main actions that the marketing and analytics teams need to take when the organisation is embarking on an MVP development process:

- Map out the entire evolution plan for the underlying data structure of the website or app. Discuss the approach to unique IDs and attributes that will be assigned to pages, screens and events, and make sure they provide a clear path to the kind of reporting and analysis that is planned.
- Push as hard as possible for a complete analytics implementation in the initial release of the product.

WHERE DOES THIS ALL LEAD?

If there is a single thread that ties these four topics together, it is that brands must see data as a privilege, not a right. Marketers need to build relationships with their customers and gather data with full transparency. Data and business intelligence teams need to carefully manage and protect harvested data to ensure compliance and maintain consumer trust. Analytics teams need to turn a critical eye towards the data before relying on said data. They must ask the tough questions of their data providers and work to validate any outside data. Product teams must ensure that accurate and

complete data collection is a required aspect of any MVP, rather than a ‘nice-to-have’ to be added later. Marketing analytics will continue to advance, even if it is ‘two steps forward, one step back’.

References

1. Marvin, G. (2013) ‘Survey: 3 out of 4 consumers now notice retargeted ads’, available at: <https://marketingland.com/3-out-4-consumers-notice-retargeted-ads-67813> (accessed 4th May, 2021).
2. Wikipedia (n.d.) ‘Facebook–Cambridge Analytica data scandal’, available at: https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal (accessed 4th May, 2021).
3. Cadwalladr C. and Graham-Harrison, E. (2018) ‘Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach’, available at: <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election> (accessed 4th May, 2021).
4. Wilson Sonsini Goodrich & Rosati (2016) ‘New EU data protection regulation is now enacted’, available at: <https://www.jdsupra.com/legalnews/new-eu-data-protection-regulation-is-81870/> (accessed 4th May, 2021).
5. Veritas (2018) ‘Organizations worldwide fear GDPR non-compliance could put them out of business’, available at: <https://www.uk.insight.com/content/dam/insight/EMEA/blog/2017/06/GDPR-Infographic-design-final.pdf> (accessed 4th May, 2021).
6. Gartenberg, C. (2021) ‘Why Apple’s new privacy feature is such a big deal’, available at: <https://www.theverge.com/2021/4/27/22405474/apple-app-tracking-transparency-ios-14-5-privacy-update-facebook-data> (accessed 4th May, 2021).
7. Deloitte Digital (2020) ‘What the end of third-party cookies means for advertisers’, available at: <https://www.deloittedigital.com/us/en/blog-list/2020/what-the-end-of-third-party-cookies-means-for-advertisers.html> (accessed 4th May, 2021).
8. de Poulpique, P. (2017) ‘What is a Walled Garden? And why it is the strategy of Google, Facebook and Amazon Ads platform?’, available at: <https://medium.com/mediarithmics-what-is/what-is-a-walled-garden-and-why-it-is-the-strategy-of-google-facebook-and-amazon-ads-platform-296ddeb784b1> (accessed 4th May, 2021).
9. Welch, C. (2018) ‘Facebook may have knowingly inflated its video metrics for over a year’, available at: <https://www.theverge.com/2018/10/17/17989712/facebook-inaccurate-video-metrics-inflation-lawsuit> (accessed 4th May, 2021).
10. Joseph, S. (2020) ‘In the absence of third-party cookies, publishers are building walled gardens of their own’, available at: <https://digiday.com/marketing/absence-third-party-cookies-publishers-building-walled-gardens/> (accessed 4th May, 2021).
11. NPD (2021) ‘US households now have over one billion computers, mobile, and TV-connected devices’, available at: <https://www.npd.com/news/press-releases/2021/u-s-households-now-have-over-one-billion-computers-mobile-and-tv-connected-devices/> (accessed 25th July, 2021).
12. Adobe (2021) ‘Cross-device analytics’, available at: <https://experienceleague.adobe.com/docs/analytics/components/cda/overview.html#cda> (accessed 4th May, 2021).
13. Vu, C. (2021) ‘Cross-device user tracking in Google Analytics’, available at: <https://medium.com/sparkline/cross-device-user-tracking-in-google-analytics-108aac4373f1> (accessed 4th May, 2021).
14. Lapidis, J. (2009) ‘Why did it take Google so long to take Gmail out of “beta”?’, available at: <https://slate.com/news-and-politics/2009/07/why-google-kept-gmail-in-beta-for-so-many-years.html> (accessed 4th May, 2021).