

The data lake and unified attribution models: analytical challenges at AAI

Brian Carruthers

WARC

Source: Event Reports, IPA Effectiveness Week, October 2017

Downloaded from WARC

There's more data than ever, but analytics is more complicated and more diverse than it has ever been before; with more being measured, finding out what you need to know is more important than ever.

- With a wealth of data, there are a number of methods that can be used to work through it; three are key: agile, design thinking, and artificial intelligence.
- Elsewhere, attribution continues to be a sticking point – though digital platforms give quite simple effectiveness metrics, finding the whole picture remains a challenge dealt with at a strategy level.
- ROI promises a simple ratio to show to business decision makers, but in reality its calculation with regards to brand building (rather than activation) requires patience and a wide view.

Curiosity and flexibility are essential characteristics of the insights researcher, never more so than today when there are so many new digital tools available to them and the volume of data that can be assessed is growing exponentially.

"The world is changing fast and on the data side of things, it's coming at us not just in terms of change but also in the changing types of data that we're seeing," Simon Edwards, VP/market development and insights at IBM Europe told an audience at the Audience Analytics & Insight Forum (AAI), a satellite event during the IPA's Effectiveness Week (London, October 2017). For example, there are 12 terabytes of Tweets daily – "what the hell do we do with all that data?"

The data lake

If you're going to avoid drowning in the data lake, there are a couple of things you need to consider, he suggested. First of all, come back to the 'why?' question. "Asking why gets to the point of understanding what people are really looking for." Then it is necessary to connect the dots: "how do I bring together all those information sources" – from traditional surveys, to social dialogues to past relationships – in a way that can help "deliver an experience that they individually recognise [as one] of value".

"We've moved from hypothesis-based discovery to correlation-based," he added. "We don't actually know what we're looking for – which bit works, why does it work, how does it work, what's the connection. And we don't know where that starts."

Edwards uses three tools:

- **Agile:** working in sprints and iterating to deliver value quickly but being prepared to move off what you're doing if it's not working.
- **Design thinking:** pick a person, give them a name and bring them into the design process.
- **Artificial intelligence:** you have to invest time and data to train it and think about what you want it to learn. It has to be a specialist in providing you with support and being a personal assistant.

"There is so much going on you have to start using new approaches, both in terms of techniques and in terms of technology," he stated. "Otherwise we can't make sense of the data, at scale, at speed and bring it all together on a global basis."

Peter Grindrod, a professor of mathematics at Oxford University, added that it's important to always be able to get back to the source data. And he offered a couple of examples of how the data lake had successfully been scoured to produce new services, one using tweets, the other phone data.

"We can look live at live events like X Factor [the TV show] and make decisions about the type of people who are second-screening, what sort of things they're saying, what sort of things are discussed," he said. "And the sheer weight of Tweets [more than 40 million] is what gives this authority." But "you've got to work hard to turn this haystack into knowledge".

Having done that. It's then possible to move on to look at how one can use that knowledge. "There's a sort of community around brands [like the X Factor] and that allows us to suggest brands and celebrities who would be appropriate for moving those kinds of people," he said – like Dominos pizza, for example, which created the Pizza Party idea for people watching the show.

Grindrod's own business Cignifi uses two weeks of mobile phone activity to profile the user and place them into one of 200 segments for the purposes of behaviour-based credit scoring of people living in shanty towns who are on no other database. "If I want to lend you money then it's guilt by association all the way down," he said. "The good news is if you change your behaviour, if your mother comes to stay, your credit score can change."

Such real-time analytics are highly differentiating, he said, adding that insights "need to be intuitive and beneficial to both sides: people understand their data is going to be exploited, they know they're going to get a better fit and they think it's cool".

Attribution

The 'why?' question was also important for Gerhard Louw, senior manager/international media & digital transformation at Deutsche Telekom. If media reporting shows what is happening with audience behaviour, media analytics shows why it is happening. "It's a journey and you can't really jump any steps," he said. There's the added complication of different parts of the business having different views on the subject. "The business isn't really that interested in marketing reach and media channels," he noted. "They want to know how to get more with the same assets – more website visits, increased market share – and many people want everything at the same time. Some people want to look backward, some forward and other people want to know 'what if?'"

There's a need to bring more science to answering all these questions and to attribution modelling, he suggested. That in turn begs the question of what tools are going to be used. His approach has been to start with where sales and media are taking place (online, offline, mixed). "Based on those two variables you also look at the level of data you have access to – user level, aggregated level or both – and then come to a suggestion of either multi-touch attribution (MTA), media mix modelling (MMM) or the unified model which is really the two combined."

In the various markets in which Deutsche Telekom operates, the approach has been to start with MTA or MMM – "don't try to do both if you haven't done anything before" – with a medium term aim of developing a unified model a couple of years from now when statistical methods make that possible. "There's a whole load of other analytics that come into play," he added, like brand tracking and A/B testing which can't yet be put into one single model.

So having accepted that you can't get all the answers you want in one step, it is necessary to develop a longer-term plan; his own stretches over four years. Online is a good place to start, he suggested – "data is more available and more granular" – before moving to the offline world where most sales still happen (and anything with a time stamp, like TV and radio, you can attribute, he said). Then one can bring in market share and pricing – the latter an area that is particularly difficult for Deutsche Telekom given the numerous options in terms of mobile phones and long-term contracts. Finally, brand tracking and creative are issues that have to be brought in at a later stage.

"Get some alignment in the business in terms of KPIs," he added, while acknowledging that sharing these across departments is not easy. Supplier management and internal change management are also challenges that will have to be addressed. "It's a long and difficult journey to ROI and the single source of truth," he concluded.

Agencies and platforms have a different take on the attribution question. Anton Kopytov, head of data & analytics at GroupM's [m]Platform, explained how the media agency has developed a "unique single identifier" for a consumer. "Connecting the dots between the different signals that describe an individual, getting data from panel research, from behavioural data on a person's activity online and offline and connecting this to a single identifier can really solve those questions [around attribution]."

"Welcome to the era of the post-demographic consumer," he announced. Media planning, media activation, media buying and outcomes measurement will in future be based on behavioural economics.

"We can get a real picture of what has happened and moreover why it has happened." And getting all that information in one computational engine, he argued, makes it possible to answer questions about the lifetime value of consumers and the influence of brand equity on decision making, while programmatically activating consumers with appropriate messaging strategies.

From Facebook's global market science division, Igor Skokan pointed to the explosion of mobile phone ownership as a tipping point in the attribution debate, as the number of online access points multiplied and the use of multiple browsers further complicated the challenge of stitching together all the data points that can be gathered. And, of course, all this produced one more data silo to be integrated. "People are the only common denominator across all channels and any future measurement system needs to be anchored in people," he said.

Last-click attribution continues to be widely used even though it's outdated, he noted - the error rate is high as 54% – "but it's just one instance of a simplistic science". Adding other touchpoints using a rule-based model is a better approach, but, he argued, the best way to get to "ground truth" is "statistical MTA calibrated with lift".

There is no currently no unified way to do this across the industry, but he expected that a standard would be agreed at some point.

The optimism of advertiser, agency and platform was tempered by the response of a researcher. Donald Hamilton of Research Now observed that true attribution was not possible since media is always changing and the industry is always playing catch-up. "We'll be stuck with a probabilistic point of view of what consumers believe they've consumed," he said. "And consumers themselves will start to push back and say 'you can't have my data' or 'I want some value for my data'," he added.

ROI

Direct Line takes a holistic approach to marketing evaluation. "We try to immerse ourselves in as many different analytical techniques as we can," said Billy Ryan, marketing effectiveness manager at the insurer. "They all have strengths and weaknesses depending on what question you're trying to answer. And for the more complicated questions you're going to have to integrate them and use them together."

Consider, for example, a marketing manager pondering whether to use Martin Lewis, the self-proclaimed money-saving expert, or Hollywood actor Harvey Keitel in his communications. The latter represents the brand proposition of getting things sorted fast and would maybe be seen as the obvious choice. "But if we just look at our basic evaluation methods, we see that the affiliate marketing – the pure short-term stuff- has a much more favourable cost-per-acquisition.

It's necessary to go one step beyond, he said, and ask: are those customers' behaviours any different once we've acquired them? "For us, attribution doesn't tell us enough about the incrementality of our activity. But what we can use it for is to create meaningful and powerful lifetime value segments." So short-term affiliate marketing is better than TV in year one, but the lifetime value profiles show that non-digital customers are more likely to remain and to be open to cross-selling.

Louise Cook, MD of econometrics agency Holmes & Cook, picked up the issue of companies having multiple messages - something that especially the case with supermarkets which can often have different strands of campaigns around value, offers, service and quality. "How does that affect their image measures," she asked. Her example showed service advertising leading to deteriorating perception of value for money – "so we need to be really careful of how we balance these messages".

And it's important not to get so focused on tracking your own activity that you forget about competitors, she added. Another example showed how quality perception can be affected by a rival's advertising. "Some quarters we were ahead, some quarters it was effectively a zero-sum game."

A final piece of advice was to be patient in waiting for data and effects. In her work with mobile operator O2, when it decided to sponsor the O2 Arena in London, it was several months before the effects of this activation were visible among older contract customers as opposed to younger pay-as-you-go consumers. And it was a year before the effects of recommendation filtered through.

About the author

Brian Carruthers
News Editor, WARC



© Copyright WARC 2017

WARC Ltd.

Americas: 2233 Wisconsin Ave NW, Suite 535, Washington, DC 20007, United States - Tel: +1 202 778 0680

APAC: 20A Teck Lim Road, 088391, Singapore - Tel: +65 3157 6200

EMEA: 85 Newman Street, London, United Kingdom, W1T 3EU - Tel: +44 (0)20 7467 8100

www.warc.com

All rights reserved including database rights. This electronic file is for the personal use of authorised users based at the subscribing company's office location. It may not be reproduced, posted on intranets, extranets or the internet, e-mailed, archived or shared electronically either within the purchaser's organisation or externally without express written permission from Warc.

WARC

