INSIDE THE DATA-DRIVEN NEWSROOM

REAL-TIME DATA FEEDS AND
ANALYTICS TECHNOLOGIES
ARE RADICALLY
TRANSFORMING THE WAY
EDITORIAL TEAMS WORK.
JO BOWMAN REPORTS

It wasn't that long ago that, for a news editor to judge public reaction to a particular story, they had to wait for the next day's post and gauge it by the weight of Letters to the Editor. Now, it's possible to know within a fraction of a second, not only how many people are reading each story, but how long they spend on it, what they read next, what they were doing beforehand, which ads they look at – and even who their friends are. This is the era of the data-fuelled newsroom, where real-time information on what the world's consumers are doing can help make or break digital news outlets.

For a digital-only news provider like *Buzzfeed*, the news and entertainment site best known for its amusing lists of timely news nuggets, technology is at the very heart of the organisation. Dao Nguyen, VP of growth and data, explains that news sites can be optimised for all sorts of measures; *Buzzfeed* focuses on two metrics for success – the highest possible number of unique visitors, and the highest possible level of story-sharing by visitors – and everything is built around trying to grow those two elements.

Ultimately, Nguyen says, a successful news organisation needs to have great editorial that's appreciated by the people who are drawn to the site, but data helps journalists and marketing teams get the right people to the right content. "The culture at *Buzzfeed* since the beginning has been very data-based," she says.

"We do AB testing of headlines; editors can write a post and test stories on the fly and figure out which ones are working best; they get alerts (showing how stories are performing), and they can learn from this, working out things like the optimal story length and headline structures."

For *Buzzfeed*, which has more than 80m unique visitors a month and is growing fast, 'sticky' headlines are not enough; content has to deliver to consumers on what the headlines promise, and it must be shareable. "Our assumption is lots of people are going online and straight to Facebook and Twitter, not straight to buzzfeed.com," Nguyen says. "Our goal is to get our stories onto those platforms … and give our content the best chance of going viral."

At *The Huffington Post*, each individual has their own dashboard they can customise

Damon Scheleur/The Huffington Post

GOT A DASH?

The Huffington Post, which served as a training ground for several of *Buzzfeed*'s top managers, now has about 50m unique visitors a month in the US and – half-way through an ambitious international roll-out programme – has about 25m a month from elsewhere in the world.

The extent to which *The Huffington Post* is data-driven came as a surprise even to its CEO, Jimmy Maymann, when he took on the post late last year. "Each individual has their own dashboard they can customise, so each editor will know exactly what's going on in their area; they'll know exactly what stories are trending in their area, how their stories are trending, and they'll be able to AB test their headlines in real time. So the amount of data we use and the way we've empowered people is

really sophisticated and is really helping us to drive our business forward." As *HuffPost* seeks to build its audience through social media and search, speed and relevance are essential. "Without those data, that wouldn't be possible," says Maymann.

New York-based Chartbeat, whose clients include the *Wall Street Journal* and Gawker Media, provides data dashboards to news organisations designed for both the editorial and the advertising sides of the business. Their key metric is "engaged time" – understanding what happens after someone makes an initial click onto a site or piece of content.

"The most important pieces of data that newsrooms should be monitoring are often the ones that are most overlooked, actually," says Chartbeat head of brand Lauryn Bennett. "Above all else, I think editorial teams have to understand the return rate of their audience. If the work they're doing is successful – if the content they're creating is what people want – people are going to read more of it and come back the next day to read even more still. Understanding the number of people who do return, and how you grow that returning audience, is crucial to the long-term success of your publication.

"But if you're focused exclusively on getting more and more page views then, odds are, you're chasing after a new audience every day. That's exhausting and oftentimes wasted effort. Instead, editors should focus on doing what they do best – getting people not just to click on but actually read the best content on their site. If you create content they want to consume, they'll come back. In fact, our data science studies have found that those that read for three minutes are much more likely to return to your site than those that read for one. That means yesterday's stories and hard work will pay off the next day. And this kind of loyalty is important to your overall business. It's less expensive to retain users than it is to acquire them, and a loyal audience is one that

subscribes to your newsletters and paywall content, and the one advertisers want to reach."

NEWS ALERTS

Giovanni Giuffrida is CEO of Milan-based Neodata Group, which provides news organisations with alerts and suggested courses of action based on the data their sites are generating. The aim is to better understand readers, increase content consumption by making it more engaging, and grow publisher revenue.

In partnering newsrooms, editors can see which stories are getting the most traction, where traffic for each story is coming from, whether it's internal traffic after people have clicked on other stories, referrals from social media, or external links. Giuffrida says Neodata's Newscurve tool also examines social engagement with stories, tracking the intensity of sharing on Facebook, how many people are commenting on articles and any tweeting about them. Neodata is able to work out a baseline of expected activity for each publication and each kind of story – how much tweeting is expected from a story on Italian politics, for instance, or how many clicks a natural disaster in China might get – and send alerts to editors if stories aren't meeting (or are exceeding) those expectations.

Dao Nguyen

Macey J Foronda/BuzzFeed

Giovanni Giuffrida

Jimmy Maymann

▸ If a low-attention alert is raised, it might come with a recommendation to change the story's location to a better position, or perhaps to remove it. High-attention alerts enable publishers' social media teams to know which stories they should be commenting on, and editors can see which stories are most likely to benefit from the most frequent updates and picture galleries, and can position links to other stories close to the most-read pieces.

The Noozhawk.com site, in California, was able to see, Giuffrida says, that a story on a particular shark attack was performing better than expected; the fact that the victim in this case was a child was generating lots of additional interest. The article was moved to a more prominent position on the site and traffic doubled; extra resources went into deeper coverage and family interviews, and traffic doubled again.

Bennett says newsroom data can help editors cater to their most valuable audiences. "Not all traffic is the right traffic for your site. Oftentimes, when something hasn't been read by many people, it doesn't mean it's a bad piece, it means the right audience is having a hard time finding it," she says.

"On the other hand, if you get a one-off pick-up that's driving a tonne of traffic to one article, but this audience is simply clicking on the headline, staying for a couple of seconds and then bailing, then it's probably not worth it to pull your entire newsroom off what they're working on to focus on this story. The best newsrooms learn from their audience's behaviours – finding out what sources bring you what kind of traffic. For instance, we see that, on average, traffic from Twitter has a 30% higher return rate than traffic from Facebook. That's important to know when seeing traffic come to a story from either source."

SENSE CHECK

But doesn't a colour-by-numbers approach to news risk leading all publishers towards providing more coverage of trivia on minor celebrities and videos of cats? Not if you don't let it, publishers and data managers say. "This concern is the main barrier we encounter when we talk to editors and prospects," Giuffrida says. "They say: 'We don't want a tool to teach us our job: we know what works, and there are certain things we have to cover even if they're not the most popular'. But the tool isn't automatic; it doesn't make the change for you. It's a decision-support tool."

Data should never have last word. It's not there to be followed to the exclusion of human news sense and business intuition. "When blindly followed, data can certainly lead you down the wrong path," says Chartbeat's Bennett. "I think that's a bit of the problem we have when some sites rely on page views as their be-all-end-all metric – it leads to terrible user experiences on the web, link-baity headlines, and crappy content because we lost all sight of what really matters

Newsroom data can help editors cater to audiences

" To be a credible news source we have to report the news, not just what's trending **"**

in journalism: content that people want to read."

At *HuffPost*, the company will occasionally pursue a project despite data signalling that it's not working. "It's a bit like asking people before they know about iPhones: 'Do you want an iPhone?'" says Maymann. "They say no. When you're trying to do giant leaps, then the data can help with trends, but it's not always telling you exactly what you want to know and so you have to go with your gut feeling."

The daily news agenda is one instance where data doesn't always trump instinct; while lifestyle and entertainment verticals are highly driven by what the online community is already talking about, in news there remains a traditional determination to help set the news agenda, as well as to follow it. *The Huffington Post* won a Pulitzer Prize last year for its series on wounded US veterans.

"To be a credible news source we have to report the news, not just what's trending," Maymann says. "If there's a story about 10 cats drowning, that might get millions of views, but it's not as important as Obama making an announcement on Syria, even if only a thousand people would follow that. We need to stay true to that."

THE STORY BEHIND THE STORIES

An invisible network of trackers powers the data-driven newsroom, but publishers are largely in the dark about how much tracking takes place. **Brian Tarran** reports

A lot of things happen when you decide to visit a website. Your browser needs to make a call to a web server to request the web page, which is then transferred to your machine and rendered on screen. This all happens in a matter of seconds. But that's just a fraction of the activity that occurs; the visible bit. Hidden away behind the scenes, a whole host of software tags and scripts are firing off and running; collecting and reporting information to build up a profile of who you are, where you've been and what your interests are.

This type of data is hugely valuable for editors and publishers, helping them to understand their audience.

Most people are unaware of this invisible web of activity – but Andy Kahl sees it. He's the director of data analysis for Evidon, a company whose stated goal is to "reveal the invisible web". Four years ago, Evidon bought a piece of software called Ghostery, a browser plugin that flags all the tracking scripts running on any given webpage. Twenty million people use Ghostery, says Kahl, and 10 million of those people agree to share data on the scripts they bump into to help Evidon keep track of the trackers.

Exponential growth
The number of unique tracking technologies encountered by Ghostery users rose from 645 in the first quarter of 2012 to 987 by the end of the fourth quarter. That's an increase of 53% – and growth shows little sign of slowing down, says Kahl.

The data collected from Ghostery allows Evidon to compile tracker maps, like the one pictured below. This is a node graph, taken from a live scan of HuffingtonPost.com. The small orange and red nodes are

analytics software and media measurement trackers – in this case Google Analytics, ComScore's ScoreCard Research Beacon, Nielsen's NetRatings SiteCensus, and Omniture. In amongst this, there's also Adobe's Test & Target software, which allows for the AB testing of headlines that *Huffington Post* CEO Jimmy Maymann refers to on p60.

Nodes are sized according to prevalence – and the most prevalent tracker on HuffPost is Advertising.com. Like *HuffPost*, Advertising.com is owned by AOL, and its job is to match ad buyers with ad inventory on AOL-owned sites. So, as you'll see from the map, Advertising. com is putting out calls to ad servers and exchanges such as Yahoo's Right Media and Google's DoubleClick.

Leaked information
Of the 27 tracking technologies present on the *HuffPost* site, only

Andy Kahl

13 are called by the site itself, or its subdomain, s.huffpost.com. Thus, 'data leakage' should be a prime publisher concern, says Kahl. Website owners are frequently in the dark about how much tracking is actually going on – and if they don't know what's happening, it's very difficult for them to keep control of their audience data.

Kahl stresses that having a wide array of trackers is not necessarily a bad thing – a publisher with a lot of low-cost ad inventory to fill might want to partner with as many ad servers as possible. Rather, his view is that the number and type of trackers a site has should be rooted in a strategy for what the publisher wants to achieve. ∎

Red = Analytics
Orange = Tracker
Purple = Unknown
Green = Widget
Blue = Ad
Pink = Publisher