Robot News: Responses to data-driven news production

This report is the outcome of a workshop held on 24/2/16 at LSE with participants from a range of organisations including Financial Times, The Times, Reuters Institute, BBC, Storyful, The Guardian, Global Editors Network, CUNY and Paris Est University.

The workshop was hosted by Dr Alison Powell, Assistant Professor and Programme Director of the MSc in Media and Communication (Data & Society) and Prof Charlie Beckett, Director of Polis, the LSE's journalism think-tank.

The ideas outlined below are intended as a starting point for further discussion and research.

Executive Summary

- There is a tension between the democratic/ public function of news and the business imperative of reaching a large audience.

- When using audience metrics, journalists should be mindful of the 'news gap': the preference gap between the supply (journalism) and demand (audience).

- While data can enable new stories to be found, journalists still have to make an editorial judgement on what should be made public.

- There is a difference between openness, (the presence of accessible data), and transparency, where journalists deem it in the public interest to make information accessible.

- There are two types of data-driven journalism: one, exploratory which sees data as the heart of the story, and another, explanatory, which sees data as supporting a story.

- When sampling from social media datasets and online communities to discover stories, journalists need to be aware of the bias inherent to the platform itself.

- To build a positive culture around data in the newsroom, there needs to be a willingness to imagine the future of journalism in collaboration with algorithms.

LSE MEDIA@LSE
Department of Media and Communications

POLIS
Journalism and Society

1. Engaging with the audience

What do the metrics say and are you listening?

The conversation started with a participant pointing to the interesting rhetorical shift in the democratic value of knowing what your audience wants. Looking back to the analogue era, knowing your audience was far from being a priority. Nowadays, journalists cannot afford not to pay attention to their audience. The debate revolved around the following dilemma: has journalism become more democratic, or simply more competitive? Or perhaps both?

Participants widely agreed on the fact that since data about audience preferences is available, it would be a missed opportunity not to use it. The issue of setting limits to the use of metrics however raised some interesting questions about making news for the audience. It was suggested that some journalists believe having too much information can be detrimental to good journalism. There is a potential tension between the democratic/ public function of news and the business imperative of reaching a large audience.

The efficiency of metrics in measuring audience engagement was also questioned. What do likes and attention minutes really tell you? Making sense of metrics implies that the audience is a rational entity, which perhaps isn't always the case. It was mentioned that there is such a thing as the 'news gap', a preference gap between the supply (journalism) and demand (audiences).

According to one of the participants, determining the value of what people want to hear in the news is something which goes beyond journalism. The same tension could be applied more broadly to the field of social sciences: who determines value and in whose interest?

2. Does data-driven journalism improve stories?

Historically the news has faced two key challenges:

- Journalists tend to focus on the powerful and elite
- There is more interest in shallow, event-based stories

Given the changes in technology that allow data-driven journalistic processes, are either of these aspects of the industry altered?

Open data: hidden in plain sight

A number of examples raised by participants pointed to the use of data as an integral part of breaking a powerful public interest accountability story. For example the case of Qatar's alleged bribery of [FIFA officials to win the World Cup 2022 bid](#) was used to show the important role of

data in proving such a story of corruption but the question remains whether the most important aspect of this story was the data, or more traditional source connections to FIFA's chief executives who spoke out on the issue.

Additionally, journalists work to unearth data that has been 'hidden in plain sight' to show the public where the stories are. Often information, for example the salary data on charity CEOs is publicly available but laborious to access.

It was highlighted that completely computational methods can also be used to find stories, for example in the case of the UK's tennis match fixing scandal. However, journalists still have to make an editorial judgement on what should be made public, and in this example the code used to ascertain match fixing was made open-source, even though the names of individuals were not revealed. Journalists are needed to make these judgements, because the impact can be devastating. For example, a project in Los Angeles published the ratings of schoolteachers in the city, and it was shortly followed by the suicide of one of the teachers who had been named.

Openness vs Transparency

One participant posited that there is a difference between openness, as in simply the presence of accessible data, and transparency, where journalists deem it in the public interest to make information accessible.

There was some disagreement over whether it was the responsibility of journalists to curate open data and make decisions about what should be put into the open, or if everything should be 'democratised' for people to access and explore at will.

Somewhere in between these polarities would be interactive data tools on news websites that allow users to find their own stories, rather than be told a particular view through an article or video. It was questioned whether this kind of service falls in the remit of the journalist, or the state. It was suggested the question may be redundant as the overriding aim is to attract traffic to the media outlet's website and keep them on the page.

Is data-driven journalism changing the news agenda?

Are journalists now following stories that they would not have followed before the current technology was available? It was suggested that there are two types of data-driven journalism: one, exploratory which sees data as the heart of the story, and another, explanatory, which sees data as supporting a story.

For example, in a big story such as austerity in the UK, it was suggested that data can add a 'human aspect' like a testimonial of experience but on a much bigger and powerful scale. The

aim of such stories is to enable the audience to relate to issues usually presented to them in impenetrable economic and political language. Data made it easier to relate austerity policies to people's everyday life, which is ultimately what journalism should be aiming for.

Journalistic partnerships

As informants start to bring data to journalists, such as Wikileaks, should journalists be acting as a mouthpiece for these 'whistle-blowers'? It was suggested that perhaps media should partner with open data NGOs on this kind of work as they hold expertise and have resources to support journalistic efforts. However, at the same time, this presents a problem for journalistic impartiality, one which is not entirely new.

3. Journalistic trust in algorithms

Data availability, bias and representation

Participants shared their experience of dealing with a wide range of platforms and insisted on the importance of understanding how these platforms work, developing awareness and expertise. For example, when sampling from social media datasets to create communities around a story, you need to be aware of the bias inherent to the platform itself. If a community of very active journalists start tweeting about the same story, how representative is it of what is going on in the world? And how much will this influence the news agenda? Moreover, participants discussed the ethical responsibility of curating data from social media as users are not necessarily notified when added to a source list.

The problematic bias of existing sources is not necessarily something that can be measured, and yet it is something journalists have to be constantly aware and critical of. When it comes to data representation, are we at risk of further silencing the hidden voices? Participants highlighted that some stories simply don't manifest themselves in the data.

There is arguably a geographical bias, so that data sheds light on the 'statistical centre of the world'. It was also pointed out that some voices are louder than others which leaves some stories untold. What happens to those stories, the human stories which often go unreported, such as stories of rape and sexual assault? Participants grappled with the moral implication in choosing to use data that is readily-available and ignoring what is not.

Making the most of algorithms

It was agreed that building a positive culture around data requires both an ability to see beyond the challenges such as bias, and a willingness to imagine the future of journalism in collaboration with algorithms. As a matter of self-reflexivity for the profession, the discussion

revolved around the following interrogation: what can algorithms do better than journalists? Participants were divided on this issue, there was some scepticism over the verification process and information overload in the newsroom.

Beyond data and leaks, participants acknowledged the use of algorithms as 'signals', also known as 'digital rumours'. Signals are a game-changer for journalists, they reduce time spent on deciding whether or not they should pursue a story. There is still however a necessity for journalists to keep a healthy distance from the signals to ensure professionalism and avoid data-dependency.

It was pointed out that signals are not necessarily a story in themselves, but rather something pointing you to a potential story. Throughout the discussion emphasis was put on the need to add an editorial value judgement to the use of algorithms.

4. Conclusion

Journalism has gone through cycles of change, from offline to online, to web 2.0 and now 'big data'.

In some ways the new uses of data as content, revelation and as information on audience behaviour have deepened journalism and made news media more diverse, rich and effective. But there are also dangers such as using data created by organisations that have their own agendas.

The move towards a journalism of 'facts' which is driven by a belief in data as fact is happening at the same time as a shift towards personalisation, subjectivity and relativism. This tension between the factual and the emotional is not entirely new for journalism, but it is does seem to be particularly complex and radical at the moment, driven as it is by the shift to social, mobile and algorithmic news. As this report shows, it offers challenges but also great opportunities for journalism.

Further reading:

Book Review of 'The News Gap' (LSE Media Policy Project Blog)
Legacy Newsrooms Embrace Innovation but not Cultural Change (Media Shift)
The Journalist-Engineer (Matthew Daniels)
Data Visualisation: What's next? (Signal Noise)

Get in touch:

Charlie Beckett C.H.Beckett@lse.ac.uk
Alison Powell A.Powell@lse.ac.uk