53. What Is Data Journalism For? Cash, Clicks, and Cut and Trys

Nikki Usher

Abstract

The financial incentives and the unintended consequences of commercial data journalism are addressed.

Keywords: journalism, datafication, misinformation, political economy, elections, experimentation

The daily refreshing of *FiveThirtyEight*'s interactive 2016 election map forecasts was all but ritual among my fellow Washingtonians, from politicians to journalists to students to government workers and beyond. Some of this ilk favoured *The New York Times*' Upshot poll aggregator; the more odds-minded of them, *Real Clear Politics*, and those with more exotic tastes turned to *The Guardian*'s US election coverage. For these serial refreshers, all was and would be right with the world so long as the odds were ever in Hillary Clinton's favour in the US presidential election's version of the Hunger Games, the bigger the spread, the better.

We know how this story ends. Nate Silver's map, even going into election day, had Hillary Clinton likely to win by 71.4%. Perhaps it's due time to get over the 2016 US election, and after all, obsession with election maps is perhaps a particularly American pastime, due to the regular cycle of national elections—although that is not to say that a worldwide audience is not also paying attention (N. P. Lewis & Waters, 2018). But until link rot destroys the map, it is there, still haunting journalists and Clinton supporters alike, providing fodder for Republicans to remind their foes that the "lamestream media" is "fake news." Politics aside, the 2016 US presidential election should not be forgotten by data journalists: Even if the quantification was correct to anyone's best knowledge, the failures in mapping and visualization have

become one more tool through which to dismantle journalists' claim to epistemic authority (or more simply, their claim to be "authorized knowers").

Yes, it is unfair to conflate data journalism with electoral prediction—it certainly is far more than that, particularly from a global vantage point, but it sometimes seems that this is what data journalism's ultimate contribution looks like: Endless maps, clickable charts, calculators prone to user error, oversimplification, and marginalization regardless of the rigor of the computation and statistical prowess that produced them. With the second edition of this handbook now in your hands, we can declare that data journalism has reached a point of maturation and self-reflection, and, as such, it is important to ask, "What is data journalism for?"

Data journalism, as it stands today, still only hints at the potential it has offered to reshape and reignite journalism. The first edition of this handbook began as a collaborative project, in a large group setting in 2011 at a Mozilla Festival, an effort I observed but quickly doubted would ever actually materialize into a tangible result (I was wrong); this second edition is now being published by the University of Amsterdam Press and distributed in the United States by the University of Chicago Press with solicited contributors, suggesting the freewheeling nature of data journalism has been exchanged somewhat in return for professionalism, order and legitimacy. And indeed, this is the case: Data journalism is mainstream, taught in journalism schools, and normalized into the newsroom (Usher, 2016). Data journalism has also standardized and, as such, has changed little over the past five to seven years; reviews of cross-national data journalism contests reveal limited innovation in form and topic (most often: politics), with maps and charts still the go-to (Loosen et al., 2020). Interactivity is limited to what is considered "entry-level techniques" by those in information visualization (Young et al., 2018); moreover, data journalism has not gone far enough to visualize "dynamic, directed, and weighted graphs" (Stoiber et al., 2015). Data journalists are still dealing with preprocessed data rather than original "big data"—and this data is "biggish," at best—government data rather than multilevel data in depth and size of the sort an Internet service provider might collect.

This critique I offer flows largely from a Western-centred perspective (if not a US-centred perch), but that does not undermine the essential call to action I put forward: Data journalists are still sitting on a potentially revolutionary toolbox for journalism that has yet to be unleashed. The revolution, however, if executed poorly, only stands to further undermine both the user experience and the knowledge-seeking efforts of news consumers, and, at worst, further seed distrust in news. If data journalism just continues to

look like it has looked for the past five to ten years, then data journalism does little to advance the cause of journalism in the digital and platform era. Thus, to start asking this existential question (“What is data journalism for?”), I propose that data journalists, along with less-data-focused-but-web-immersed journalists who work in video, audio and code, as well as the scholars who poke and prod them, need to rethink data journalism’s origin story, its present rationale and its future.

Data Journalism in the United States: The Origin Story

The origin story is the story we tell ourselves about how and why we came to be, and is more often than not viewed through rose-tinted glasses and filled with more braggadocio than it is reality. The origin story of data journalism in the United States goes something like this: In the primordial pre-data journalism world, data journalism existed in an earlier form, as computer-assisted reporting, or was called that in the United States, which offered an opportunity to bring social science rigor to journalism.

In the mythos of data journalism’s introduction to the web, data journalists would become souped-up investigative journalists empowered with the superior computational prowess of the 21st century who set the data (or documents) free in order to help tell stories that would otherwise not be told. But beyond just investigating stories, data journalists were also somehow to save journalism with their new web skills, bringing a level of transparency, personalization and interactivity to news that news consumers would appreciate, learn from and, of course, click on. Stories of yesteryear’s web, as it were, would never be the same. Data journalism would right wrongs and provide the much-needed objective foundation that journalism’s qualitative assessments lacked, doing it at a scale and with a prowess unimaginable prior to our present real-time interactive digital environment replete with powerful cloud-based servers that offload the computational pressure from any one news organization. Early signs of success would chart the way forward, and even turn ordinary readers into investigative collaborators or citizen scientists, such as with *The Guardian*’s coverage of the MPs’ expenses scandal or the “Cicada Tracker” project of the New York City public radio station WNYC, which got a small army of area residents to build soil thermometers to help chart the arrival of the dreaded summer instincts. And this inspired orchestration of journalism, computation, crowds, data and technology would continue, pushing truth to justice.

The Present: The “Hacker Journalist” as Just Another (Boring) Newsroom Employee

The present has not moved far past the origin story that today’s data journalists have told themselves, neither in vision nor in reality. What has emerged has become two distinct types of data journalism: The “investigative” data journalism that carries the noble mantle of journalism’s efforts forward, and daily data journalism, which can be optimized for the latest viral click interest, which might mean anything from an effort at ASAP journalistic cartography to turning public opinion polling or a research study into an easily shareable meme with the veneer of journalism attached. Data journalism, at best, has gotten boring and overly professional, and, at worst, has become another strategy to generate digital revenue.

It is not hyperbole to say that data journalism could have transformed journalism as we know it—but hitherto it has not. At the 2011 MozFest, a headliner hack of the festival was a plug-in of sorts that would allow anyone’s face to become the lead image of a mock-up of *The Boston Globe* home page. That was fun and games, but *The Boston Globe* was certainly not going to just allow user-generated content, without any kind of pre-filtering, to actually be used on its home page. Similarly, during the birth of the first *Data Journalism Handbook*, the data journalist was the “hacker journalist,” imagined as coming from technology into journalism or at least using the spirit of open source and hacking to inspire projects that bucked at the conventional processes of institutional journalism and provided room for experimentation, imperfection and play—tinkering for the sake of leading to something that might not be great in form or content, but might well hack journalism nonetheless (S. C. Lewis & Usher, 2013). In 2011, the story was of outsiders moving into journalism, but in 2018, the story is of insiders professionalizing programming in journalism, so in five years the spirit of innovation and invention has become decidedly corporate, decidedly white-collar and decidedly less fun (S. C. Lewis & Usher, 2016).

Boring is OK, and it serves a role. Some of the professionalization of data journalism has been justified with the “data journalist as hero” self-perception—data journalists as those who, thanks to a different set of values (e.g., collaboration, transparency) and skills (visualization, assorted computational skills) could bring truth to power in new ways. The Panama and Paradise Papers are perhaps one of the best expressions of this vision. But investigative data journalism requires time, effort and expertise that goes far beyond just data crunching, and includes many other sources of more traditional data, primarily interviews, on-location reporting and documents.

Regularly occurring, groundbreaking investigative journalism is an oxymoron, although not for lack of effort: The European Data Journalism Network, the United States' Institute for NonProfit News and the Global Investigative Journalism Network all showcase the vast network of would-be investigative efforts. The truth is that a game-changer investigation is not easy to come by, which is why we can generally name these high-level successes on about ten fingers and the crowdsourced investigative success of *The Guardian's* MPs' expenses example from 2010 has yet to be replaced by anything newer.

What's past is prologue when it comes to data journalism. "Snow Fall," *The New York Times'* revolutionary immersive storytelling project that won a Pulitzer Prize in 2012, emerged in December 2017 as "Deliverance from 27,000 Feet" or "Everest." Five years later, *The New York Times* featured yet another long-form story about a disaster on a snowy mountain, just a different one (but by same author, John Branch). In those five years, "Snowfall" or "Snowfalled" became shorthand within *The New York Times* and outside it for adding interactive pizzazz to a story; after 2012, a debate raged not just at *The Times* but in other US and UK newsrooms as to whether data journalists should be spending their time building templating tools that could auto-Snowfall any story, or work on innovative one-off projects (Usher, 2016). Meanwhile, "Snow Fall," minimally interactive at best in 2012, remained minimally interactive at best in its year-end 2017 form.

"But wait," the erstwhile data journalist might proclaim. "'Snow Fall' isn't *data journalism*—maybe a fancy trick of some news app developers, but there's no data in 'Snow Fall'!" Herein lies the issue: Maybe data journalists don't think "Snow Fall" is data journalism, but why not? *What is data journalism for if it is not to tell stories in new ways with new skills that take advantage of the best of the web?*

Data journalism also cannot just be for maps or charts, either, nor does mapping or charting data give data journalism intellectual superiority over immersive digital journalism efforts. What can be mapped is mapped. Election mapping in the United States aside, the ethical consequences of quantifying and visualizing the latest available data into clickable coherence needs critique. At its most routine, data journalism becomes the vegetables of visualization. This is particularly true given the move towards daily and regular demand for data journalism projects. Perhaps it's a new labour statistic, city cycling data, recycling rates, the results of an academic study, visualization because it can be visualized (and maybe, will get clicked on more). At worst, data journalism can oversimplify to the point of dehumanizing the subject of the data that their work is supposed to illuminate. Maps of migrants and their flows across Europe take on the form of interactive arrows or genderless person

icons. As human geographer Paul Adams argues, digital news cartography has rendered the refugee crisis into a disembodied series of clickable actions, the very opposite of what it could as journalism do to make unknown “refugees” empathetic and more than a number (Adams, 2017). Before mapping yet another social problem or academic study, data journalists need to ask: To what end are we mapping and charting (or charticle-ing for that matter)?

And somewhere between “Snow Fall” and migration maps lies the problem: *What is data journalism for?* The present provides mainly evidence of professionalization and isomorphism, with an edge of corporate incentive that data journalism is not just to aid news consumers with their understanding of the world but also to pad the bottom lines of news organizations. Surely that is not all data journalism can be.

The Future: How Data Journalism Can Reclaim Its Worth (and Be Fun, Too)

What is data journalism for? Data journalism needs to go back to its roots of change and revolution, of inspired hacking and experimentation, of a self-determined vision of renegades running through a tired and uninspired industry to force journalists to confront their presumed authority over knowledge, narrative and distribution. Data journalists need to own up to their hacker inspiration and hack the newsroom as they once promised to do; they need to move past a focus on profit and professionalism within their newsrooms. Reclaiming outsider status will bring us closer to the essential offering that data journalism promised: A way to think about journalism differently, a way to present journalism differently, and a way to bring new kinds of thinkers and doers into the newsroom, and beyond that, a way to reinvigorate journalism.

In the future, I imagine data journalism as unshackled from the term “data” and instead focused on the word “journalism.” Data journalists presumably have skills that the rest of the newsroom or other journalists do not: The ability to understand complicated data or guide a computer to do this for them, the ability to visualize this data in a presumably meaningful way, and the ability to code. Data journalism, however, must become what I have called interactive journalism—data journalism needs to shed its vegetable impulse of map and chart cranking as well as its scorn of technologies and skills that are not data-intensive, such as 360 video, augmented reality and animation. In my vision of the future, there will be a lot more of the BBC’s “Secret Life of the Cat” interactives and *The New York Times*’ “Dialect Quizzes”; there will be more projects that combine 360 video or VR with data, like Dataverse’s effort

funded by the Journalism 360 immersive news initiative. There will be a lot less election mapping and cartography that illustrates the news of the day, reducing far-away casualties to clickable lines and flows. Hopefully, we will see the end of the new trend towards interactives showing live-time polling results, a new fetish of top news outlets in the United States. Rather, there will be a lot more originality, fun, and inspired breaking of what journalism is supposed to look like and what it is supposed to do. Data journalism is for accountability, but it is also for fun and for the imagination; it gains its power not just because an MP might resign or a trend line becomes clearer, but also because ordinary people see the value of returning to news organizations and to journalists because journalists fill a variety of human information needs—for orientation, for entertainment, for community, and beyond.

And to really claim superior knowledge about data, data journalists intent on rendering data knowable and understandable need to collect this data on their own—data journalism is not just for churning out new visualizations of data gathered by someone else. At best, churning out someone else's data makes the data providers' assumptions visible; at worst, data journalism becomes as stenographic as a press release for the data provider. Yet many data journalists do not have much interest in collecting their own data and find it outside the boundaries of their roles; as *The Washington Post* data editor Steven Rich explained, in a tweet, the *Post* “and others should not have to collect and maintain databases that are no-brainers for the government to collect. This should not be our fucking job” (Rich, 2018). At the same time, however, the gun violence statistics Rich was frustrated by having to maintain are more empowering than he realized: Embedded in government data are assumptions and decisions about what to collect that need sufficient inquiry and consideration. The data is not inert, but filled with presumptions about what facts matter. Journalists seeking to take control over the domain of facticity need to be able to explain why the facts are what they are, and, in fact, the systematic production of fact is how journalists have claimed their epistemic authority for most of modern journalism.

What data journalism is for, then, is for so much more than it is now—it can be for fun, play and experimentation. It can be for changing how stories get told and can invite new ways of thinking about them. But it also stands to play a vital role in re-establishing the case for journalism as truth-teller and fact-provider; in creating and knowing data, and being able to explain the process of observation and data collection that led to a fact. Data journalism might well become a key line of defence about how professional journalists can and do gather facts better than any other occupation, institution or ordinary person ever could.

Works Cited

- Adams, P. (2017). Migration maps with the news: Guidelines for ethical visualization of mobile populations. *Journalism Studies*, 19(1), 1–21. <https://doi.org/10.1080/1461670X.2017.1375387>
- Lewis, N. P., & Waters, S. (2018). Data journalism and the challenge of shoe-leather epistemologies. *Digital Journalism*, 6(6), 719–736. <https://doi.org/10.1080/21670811.2017.1377093>
- Lewis, S. C., & Usher, N. (2013). Open source and journalism: Toward new frameworks for imagining news innovation. *Media, Culture & Society*, 35(5), 602–619. <https://doi.org/10.1177/0163443713485494>
- Lewis, S. C., & Usher, N. (2016). Trading zones, boundary objects, and the pursuit of news innovation: A case study of journalists and programmers. *Convergence*, 22(5), 543–560. <https://doi.org/10.1177/1354856515623865>
- Loosen, W., Reimer, J., & De Silva-Schmidt, F. (2020). Data-driven reporting: An on-going (r)evolution? An analysis of projects nominated for the Data Journalism Awards 2013–2016. *Journalism*, 21(9), 1246–1263. <https://doi.org/10.1177/1464884917735691>
- Rich, S. (2018, February 15). The @washingtonpost and others should not have to collect and maintain databases that are no-brainers for the government to collect. This should not be our fucking job. Twitter. <https://twitter.com/dataeditor/status/964160884754059264>
- Stoiber, C., Aigner, W., & Rind, A. (2015). Survey on visualizing dynamic, weighted, and directed graphs in the context of data-driven journalism. In *Proceedings of the International Summer School on Visual Computing* (pp. 49–58).
- Usher, N. (2016). *Interactive journalism: Hackers, data, and code*. University of Illinois Press.
- Young, M. L., Hermida, A., & Fulda, J. (2018). What makes for great data journalism? A content analysis of Data Journalism Awards finalists 2012–2015. *Journalism Practice*, 12(1), 115–135. <https://doi.org/10.1080/17512786.2016.1270171>

About the Author

Nikki Usher is an Associate Professor at the University of Illinois at Urbana-Champaign and is the author of *Interactive Journalism: Hackers, Data, and Code* (University of Illinois Press, 2016).