48. Genealogies of Data Journalism

C. W. Anderson

Abstract

This chapter takes a critical and historical look at data journalism.

Keywords: genealogy, data journalism, history, uncertainty, computational journalism, precision journalism

Introduction

Why should anyone care about the history of data journalism? Not only is "history" a rather academic and abstract topic for most people, it might seem particularly remote for working data journalists with a job to do. Journalists, working under tight deadlines and with a goal of conveying complicated information quickly and understandably to as many readers as possible, can be understandably averse to wasting too much time on self-reflection. More often than not, this reluctance to "navel-gaze" is an admirable quality; when it comes to the practices and concepts of data journalism and computational reporting, however, a hostility towards historical thinking can be a detriment that hampers the production of quality journalism itself.

Data journalism may be the most powerful form of collective journalistic sense making in the world today. At the very least, it may be the most positive and positivistic form of journalism. This *power* (the capacity of data journalism to create high-quality journalism, along with the rhetorical force of the data journalism model), *positivity* (most data journalists have high hopes for the future of their particular subfield, convinced it is on the rise) and *positivism* (data reporters are strong believers in the ability of method-guided research to capture real and provable facts about the world) create what I would call an empirically self-assured profession. One consequence of this self-assurance, I would argue, is that it can also create a Whiggish assumption that data journalism is always improving and improving the world. Such an

attitude can lead to arrogance and a lack of critical self-reflexivity, and make journalism more like the institutions it spends its time calling to account.

In this chapter I want to argue that a better attention to history can actually improve the day-to-day workings of data journalism. By understanding that their processes and practices have a history, data journalists can open their minds to the fact that things in the present could be done differently because they might have once been otherwise. In particular, data journalists might think harder about how to creatively represent uncertainty in their empirical work. They might consider techniques through which to draw in readers of different political sensibilities and persuasions that go beyond simply stating factual evidence. They might, in short, open themselves up to what science and technology studies scholars and historians Catherine D'Ignazio and Lauren Klein have called a form of "feminist data visualization," one that rethinks binaries, embraces pluralism, examines power and considers context (D'Ignazio & Klein, 2020; see also D'Ignazio's chapter in this book). To accomplish these changes, data journalism, more than most forms of journalistic practice, should indeed inculcate this strong historical sensibility due to the very nature of its own power and self-assurance. No form of history is better equipped to lead to self-reflexivity, I would argue, than the genealogical approach to conceptual development pioneered by Michel Foucault and embraced by some historians of science and scholars in science and technology studies.

"Genealogy," as defined by Foucault, who himself draws on the earlier work of Nietzsche, is a unique approach to studying the evolution of institutions and concepts over time and one that might be distinguished from history as such. Genealogical analysis does not look for a single, unbroken origin of practices or ideas in the past, nor does it try to understand how concepts developed in an unbroken and evolutionary line from yesterday to today. Rather, it focuses more on *discontinuity* and *unexpected changes* than it does on the presence of the past in the present. As Nietzsche noted, in a passage from the *Genealogy of Morals* quoted by Foucault*:*

> The "development" of a thing, a practice, or an organ has nothing to do with its progress towards a single goal, even less is it the logical and shortest progress reached with the least expenditure of power and resources. Rather, it is the sequence of more or less profound, more or less mutually independent processes of overpowering that take place on that thing, together with the resistance that arises against that overpowering each time, the changes of form which have been attempted for the purpose of defense and reaction, as well as the results of successful counter-measures. Form is fluid; the "meaning," however, is even more so. (Foucault, 1980)

A "genealogy of data journalism," then, would uncover the ways that data journalism evolved in ways that its creators and practitioners never anticipated, or in ways that may have even been contrary to their desires. It would look at the ways that history surprises us and sometimes leads us in unexpected directions. This approach, as I argued earlier, would be particularly useful for working data journalists of today. It would help them understand, I think, that they are *not* working in a predefined tradition with a venerable past; rather, they are mostly making it up as they go along in ways that are radically contingent. And it would prompt a useful form of critical self-reflexivity, one that might help mitigate the (understandable and often well-deserved) self-confidence of working data journalists and reporters.

I have attempted to write such a genealogical account in my book, *Apostles of Certainty: Data Journalism and the Politics of Doubt* (Anderson, 2018). In the pages that follow, I want to summarize some of the main findings of the book and discuss ways that its lessons might be helpful for the present day. I want to conclude by arguing that journalism, particularly of the datafied kind, could and should do a better job demonstrating what it does not know, and that these gestures towards uncertainty would honour data journalism's origins in the critique of illegitimate power rather than the reification if it.

Data Journalism Through Time: 1910s, 1960s and 2010s

Can journalists use data—along with other forms of quantified information such as paper documents of figures, data visualizations, and charts and graphs—in order to produce better journalism? And how might that journalism assist the public in making better political choices? These were the main questions guiding *Apostles of Certainty: Data Journalism and the Politics of Doubt*, which tried to take a longer view of the history of news. With stops in the 1910s, the 1960s, and the present, the book traces the genealogy of data journalism and its material and technological underpinnings, and argues that the use of data in news reporting is inevitably intertwined with national politics, the evolution of computable databases and the history of professional scientific fields. It is impossible to understand journalistic uses of data, I argue in the book, without understanding the oft-contentious relationships between social science and journalism. It is also impossible to disentangle empirical forms of public truth telling without first understanding the remarkably persistent progressive belief that the publication of empirically verifiable information will lead to a more just and prosperous world. *Apostles of Certainty* concluded that this intersection of technology and professionalism

has led to a better journalism but not necessarily to a better politics. To fully meet the demands of the digital age, journalism must be more comfortable expressing empirical doubt as well as certitude. Ironically, this "embrace of doubt" could lead journalism to become more like science, not less.

The Challenge of Social Science

The narrative of *Apostles of Certainty* grounds itself in three distinct US time periods which provide three different perspectives on the development of data journalism. The first is the so-called "Progressive Era," which was a period of liberal political ascendancy accompanied by the belief that both the state *and* ordinary citizens, informed by the best statistics available, could make the world a more just and humane place. The second moment is the 1950s and 1960s, when a few journalism reformers began to look to quantitative social science, particularly political science and sociology, as a possible source of new ideas and methods for making journalism more empirical and objective. They would be aided in this quest by a new set of increasingly accessible databases and powerful computers. The third moment is the early 2010s, when the cutting edge of data journalism has been supplemented by "computational" or "structured" journalism. In the current moment of big data and "deep machine learning," these journalists claim that journalistic objectivity depends less on external referents but rather emerges from within the structure of the database itself.

In each of these periods, data-oriented journalism both *responded to* but also defined itself *in partial opposition to* larger currents operating within social science more generally, and this relationship to larger political and social currents helped inform the choice of cases I focused on in this chapter. In other words, I looked for inflection points in journalism history that could help shed light on larger social and political structures, in addition to journalism. In the Progressive Era,[1] traditional news reporting largely rejected sociology's emerging focus on social structures and depersonalized contextual information, preferring to retain their individualistic focus on powerful personalities and important events. As journalism and sociology professionalized, both became increasingly comfortable with making structural claims, but it was not until the 1960s that Philip Meyer and the

---

1   In the United States the time period known as the "Progressive Era" lasted from the 1880s until the 1920s, and is commonly seen as a great era of liberal reform and an attempt to align public policy with the industrial era.

reformers clustered around the philosophy of Precision Journalism began to hold up quantitative sociology and political science as models for the level of exactitude and context to which journalism ought to aspire. By the turn of the 21st century, a largely normalized model of data journalism began to grapple with doubts about replicability and causality that were increasingly plaguing social science; like social science, it began to experiment to see if "big data" and non-causal forms of correlational behaviouralism could provide insights into social activity.

Apostles of Certainty thus argues implicitly that forms of journalistic expertise and authority are never constructed in isolation or entirely internally to the journalistic field itself. Data journalism did not become data journalism for entirely professional journalistic reasons, nor can this process be analyzed solely through an analysis of journalistic discourse or "self-talk." Rather, the type of expertise that in the 1960s began to be called data journalism can only be understood *relationally*, by examining the manner in which data journalists responded to and interacted with their (more authoritative and powerful) social scientific brethren. What's more, this process cannot be understood solely in terms of the actions and struggles of humans, either in isolation or in groups. Expertise, according to the model I put forward in *Apostles of Certainty*, is a networked phenomenon in which professional groupings struggle to establish jurisdiction over a wide variety of discursive *and* material artefacts. Data journalism, to put it simply, would have been impossible without the existence of the database, but the database as mediated through a particular professional understanding of what a database was and how it could be deployed in ways that were properly journalistic (for a more general attempt at this argument about the networked nature of expertise, see Anderson, 2013). It is impossible to understand journalistic authority without also understanding the authority of social science (and the same thing might be said about computer science, anthropology or long-form narrative non-fiction). Journalistic professionalism and knowledge can never be understood solely by looking at the field of journalism itself.

The Persistence of Politics

Data journalism must be understood genealogically and in relation to adjacent expert fields like sociology and political science. All of these fields, in turn, must be analyzed through their larger conceptions of politics and how they come to terms with the fact that the "facts" they uncover are "political" whether they like it or not. Indeed, even the desire for factual

knowledge is itself a political act. Throughout the history of data journalism, I argue in *Apostles of Certainty*, we have witnessed a distinct attempt to lean on the neutrality of social science in order to enact what can only be described as progressive political goals. The larger context in which this connection is forged, however, has shifted dramatically over time. These larger shifts should temper any enthusiasm that what we are witnessing in journalism is a teleological unfolding of journalistic certainty as enabled by increasingly sophisticated digital devices.

In the Progressive Era, proto-data journalists saw the gathering and piling up of quantitative facts as a process of social and political enlightenment, a process that was nonetheless free of any larger political commitments. By collecting granular facts about city sanitation levels, the distribution of poverty across urban spaces, statistics about church attendance and religious practice, labour conditions, and a variety of other bits of factual knowledge—and by transmitting these facts to the public through the medium of the press—social surveyors believed that the social organism would gain a more robust understanding of its own conditions of being. By gaining a better understanding of itself, society would improve, both of its own accord and by spurring politicians towards enacting reformist measures. In this case, factual knowledge about the world spoke for itself; it simply needed to be gathered, visualized and publicized, and enlightenment would follow. We might call this a "naïve and transparent" notion of what facts are—they require no interpretation in and of themselves, and their accumulation will lead to positive social change. Data journalism, at this moment, could be political without explicitly stating its politics.

By the time of Philip Meyer and the 1960s, this easy congruence between transparent facts and politics had been shattered. Journalism was flawed, Meyer and his partisans argued throughout the 1950s and 1960s, because it mistook objectivity for simply collecting a record of what all sides of a political issue might think the truth might be and allowing the reader to make their own decisions about what was true. In an age of social upheaval and political turmoil, journalistic objectivity needed to find a more robust grounding, and it could find its footing on the terrain of objective social science. The starting point for journalistic reporting on an issue should not be the discursive claims of self-interested politicians but rather the cold, hard truth gleaned from an analysis of relevant data with the application of an appropriate method. Such an analysis would be *professional but not political*; by acting as a highly professionalized cadre of truth-tellers, journalists could cut through the political spin and help plant the public on the terrain of objective truth. The directions this truth might lead, on the other hand,

were of no concern. Unlike the earlier generation of blissfully and naively progressive data journalists, the enlightened consequences of data were not a foregone conclusion.

Today I would argue that a new generation of computational journalists has unwittingly reabsorbed some of the political and epistemological beliefs of their Progressive Era forbearers. Epistemologically, there is an increasing belief amongst computational journalists that digital facts in some way "speak for themselves," or at least these facts will do so when they have been properly collected, sorted and cleaned. At scale, and when linked to larger and internally consistent semantic databases, facts generate a kind of *correlational excess* in which troubles with meaning or causality are washed away through a flood of computational data. Professionally, data journalists increasingly understand objectivity as emerging from within the structure of the database itself rather than as part of any larger occupational interpretive process. Politically, finally, I would argue that there has been the return of a kind of "crypto-progressivism" amongst many of the most studiously neutral data journalists, with a deep-seated political hope that more and more data, beautifully visualized and conveyed through a powerful press, can act as a break on the more irrational or pathological political tendencies increasingly manifest within Western democracies. Such, at least, was the hope before 2016 and the twin shocks of Brexit and Donald Trump.

Certainty and Doubt

The development of data journalism in the United States across the large arc of the 20th century should be seen as one in which increasingly exact claims to journalistic professional certitude coexisted uneasily with a dawning awareness that all facts, no matter what their origins, were tainted with the grime of politics. These often-contradictory beliefs are evident across a variety of data-oriented fields, of course, not simply just in journalism. In a 2017 article for *The Atlantic*, for instance, science columnist Ed Yong grappled with how the movement towards "open science" and the growing replicability crisis could be used by an anti-scientific Congress to demean and defund scientific research. Yong quoted Christie Aschwanden, a science reporter at *FiveThirtyEight*: "It feels like there are two opposite things that the public thinks about science," she tells Yong.

> [Either] it's a magic wand that turns everything it touches to truth, or that it's all bullshit because what we used to think has changed. . . . The truth

is in between. Science is a process of uncertainty reduction. If you don't show that uncertainty is part of the process, you allow doubt-makers to take genuine uncertainty and use it to undermine things. (Yong, 2017)

These thoughts align with the work of STS scholar Helga Nowotny (2016), who argues in *The Cunning of Uncertainty* that "the interplay between overcoming uncertainty and striving for certainty underpins the wish to know." The essence of modern science—at least in its ideal form—is not the achievement of certainty but rather the fact that it so openly states the provisionality of its knowledge. Nothing in science is set in stone. It admits to often know little. It is through this, the most modern of paradoxes, that its claims to knowledge become worthy of public trust.

One of the insights provided by this genealogical overview of the development and deployment of data journalism, I would argue, is that data-oriented journalists have become obsessed with increasing exactitude and certainty at the expense of a humbler understanding of provisionality and doubt. As I have tried to demonstrate, since the middle of the 20th century journalists have engaged in an increasingly successful effort to render their knowledge claims more certain, contextual and explanatory. In large part, they have done this by utilizing different forms of evidence, particularly evidence of the quantitative sort. Nevertheless, it should be clear that this heightened professionalism—and the increasing confidence of journalists that they are capable of making contextualized truth claims—has not always had the democratic outcomes that journalists expect. Modern American political discourse has tried to come to grips with the uncertainty of modernity by engaging a series of increasingly strident claims to certitude. Professional journalism has not solved this dilemma; rather it has exacerbated it. To better grapple with the complexity of the modern world, I would conclude, journalism ought to rethink the means and mechanisms by which it conveys its own provisionality and uncertainty. If done correctly, this could make journalism more like modern science, rather than less.

Works Cited

Anderson, C. W. (2013). Towards a sociology of computational and algorithmic journalism. *New Media & Society*, *15*(7), 1005–1021. https://doi.org/10.1177/1461444812465137

Anderson, C. W. (2018). *Apostles of certainty: Data journalism and the politics of doubt*. Oxford University Press.

D'Ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT Press.

Foucault, M. (1980). *Power/knowledge: Selected interviews and other writings, 1972–1977*. Vintage.

Nowotny, H. (2016). *The cunning of uncertainty*. Polity Press.

Yong, E. (2017, April 5). How the GOP could use science's reform movement against it. *The Atlantic*. https://www.theatlantic.com/science/archive/2017/04/reproducibility-science-open-judoflip/521952/

About the Author

C. W. Anderson is Professor of Media and Communication at the University of Leeds and author of *Rebuilding the News: Metropolitan Journalism in the Digital Age* (Temple University Press, 2013) and *Apostles of Certainty: Data Journalism and the Politics of Doubt* (Oxford University Press, 2018).