



UNIVERSITY OF LEEDS

This is a repository copy of *Between the Unique and the Pattern*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/127477/>

Version: Accepted Version

Article:

Anderson, CW (2015) *Between the Unique and the Pattern*. *Digital Journalism*, 3 (3). pp. 349-363. ISSN 2167-0811

<https://doi.org/10.1080/21670811.2014.976407>

(c) 2014 Taylor & Francis. This is an Accepted Manuscript of an article published by Taylor & Francis in *Digital Journalism* on 20/Nov/2014, available online:
<https://doi.org/10.1080/21670811.2014.976407>

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

BETWEEN THE UNIQUE AND THE PATTERN

Historical Tensions in our Understanding of Quantitative Journalism

C. W. Anderson

This article proposes that the underlying ideas of data journalism are not new, but rather can be traced back in history and align with larger questions about the role of quantification in journalistic practice. This article sketches out a theoretical frame (assemblage theory) in which quantitative journalism is best understood by examining the objects of evidence that journalism mobilizes on its behalf. The article illustrates this perspective by outlining three historical tensions in notions of quantitative journalism: tensions between records and reports, individuality and social science, and isolated facts and broader patterns.

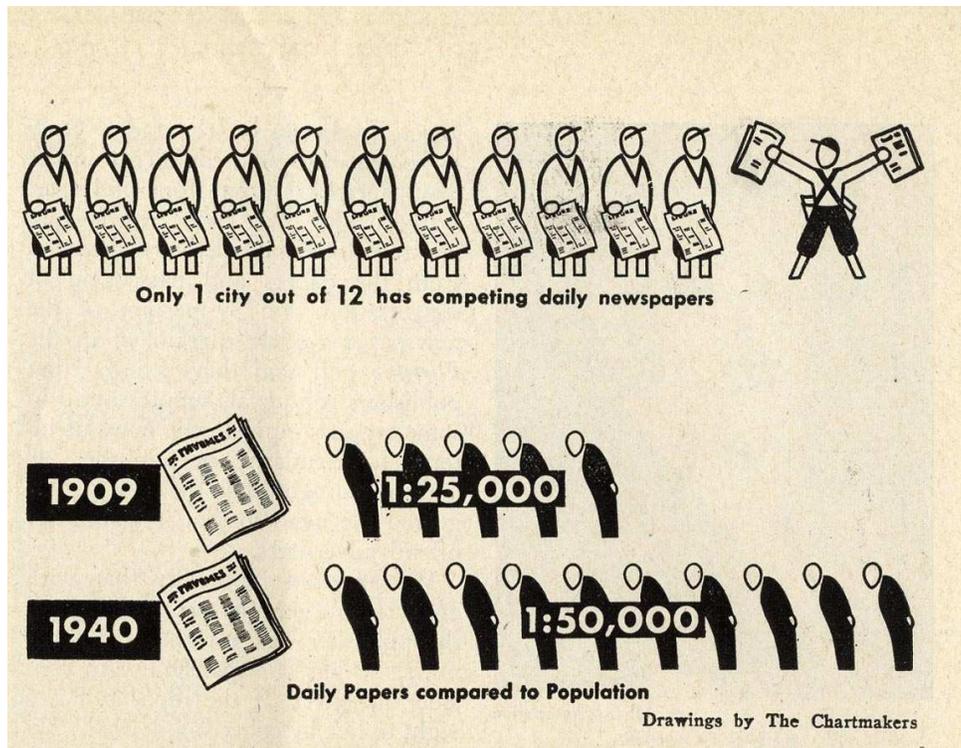
KEYWORDS: assemblage, data journalism, materiality, social science

Introduction

The year was 1946, and newspaper journalism in the United States was in trouble. Such, at least, was the opinion of Kenneth Stewart, writing in the illustrated monthly magazine the *Survey Graphic*. Among the problems for news outlined by Stewart were monopoly ownership, the growth of newspaper chains, a reliance of too many news outlets on information from the Associated Press and other wire services, and collusion on the part of American Newspaper Publishers Association to fix prices. Worst of all, Stewart lamented, was the fact that so many cities relied on only one source of information for their news:

We have lost one thousand dailies and over three thousand weeklies in the past few decades. Only 117 cities continue to have competing newspaper ownership. In more than one-hundred areas the only newspaper owns the only radio station. Five movie producers dominate the screen of the nation, through ownership of key theaters. On top of all this, some trade unions, by feather-bed rules, have placed undue and uneconomic burdens on free enterprise in the mass-media fields. (Stewart 1946, 453)

Accompanying this litany of statistics was a large informational graphic, titled “Newspapers” and produced by someone known only as “The Chartmakers.” The graphic was designed to visualize the arresting statistic that “only 1 city out of 12 has competing daily newspapers.” (Stewart 1946, 453)



This type of illustration was common in the publication appropriately titled *The Survey Graphic* and which, as part of its mission, attempted to make what its editors called “social work concerns” accessible to a broader public. The *Survey Graphic* had existed for more than 25 years, ever since it was decided to bolster the professional journal *The Survey* by publishing a more general-interest magazine designed less for social workers and more for everyday people who cared about social issues. Most importantly for the purposes of this essay, *The Survey Graphic* was designed to stand at the intersection of the world of research, the world of data, and the world of journalism; it attempted to “carry forward swift first hand investigations with a procedure comparable to that of scientific research . . . interpret the findings of others . . . employ photographs, maps, charts, the arts in gaining a hearing from two to twenty times that of formal books and reports” (Finnegan 2003). The 1946 article on newspaper circulation, discussed above, is representative of this drive to meld statistics, contextual journalism, and striking illustrations in order to make a point. What I am calling “quantitative journalism,” following the example of Mark Coddington (2015 [this issue]), seems at first glance many decades old.

* * *

In some ways, this brief trip back to the 1940s is a comforting one. The use of data to tell stories, it would appear, is not some recent imposition by statisticians and computer scientists on journalism’s more humanistically inclined

spirit. Data aggregation and information visualization have been part of journalism for a long time. Perhaps there is even something like a human need for data and for pictures built around data, a need that journalists, with the advent of new technologies, can finally satisfy. One way to understand quantitative journalism in 2014 is thus to find traces of its past in its present and of its present in its past; perhaps we can then even speculate intelligently about its future.

But in other ways the story of quantitative journalism in the United States is less one of sanguine continuity than it is one of rupture, a tale of transformed techniques and objects of evidence existing under old familiar names. I think, in other words, that the continuity perspective mentioned above is deeply and fundamentally flawed. The producers of the *Survey Graphic* understood the meaning and purpose of data very differently than most reporters and editors working today. The varieties and forms of data that journalists have available to them in the early 21st century are different, as are the techniques journalists use to gather, process, and display that data. Thinking about the history of quantitative journalism in this way—as transformation and rupture—is less likely to offer us comfort and more likely to unsettle us. But perhaps it will also help us reframe our current understanding of data and journalism as a *contingent understanding*, less inevitable and timeless than it is the product of deliberate choices intersecting with historical structures and filtered through an intellectually precarious profession. What I am proposing to sketch out in the pages that follow, in other words, is a genealogy of quantitative journalism. There is no entirely unproblematic understanding of quantitative journalism, I hope to argue below; its history is not determined by its purpose and use, but rather its purpose and use are enabled through its tangled history. As Nietzsche noted, in a passage from the *Genealogy of Morals* that held particular resonance for 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. (Nietzsche 2006: 51)

This article proposes that we ought approach the question of data journalism by studying its history, and by conducting the kind of history that Foucault called “genealogy.” That is, we ought to study the process by which journalism engages in a form of public “world building”: the way that a variety of processes, technologies, and evidentiary objects contribute to this crafting of publics and public issues; and the manner in which these materialities and discourses have been loaded with different meanings at different times. From this

perspective, “big data” might be seen as another object of evidence that enters the journalistic bloodstream at a particular moment or moments. This genealogical perspective can, in turn, shed light on why the relationship of journalism to big data matters in a more normative sense. I want to note, at the outset, that this overview is part of an in-progress monograph on the socio-cultural history of quantitative journalism. While the paper primarily draws on a wide variety of secondary literature, it is also informed by extensive archival research, including in the Franklin H. Giddings archive at Columbia University and the Paul Kellogg archive at the University of Minnesota. Although rarely directly cited in this overview article, the time I have already spent in the archives has played a key role in helping me formulate the typologies and overall direction of this paper.

In the first section of this article, I want to make the case for conducting a genealogy of quantitative journalism in a particular fashion, what I call an “objects of journalism-oriented” approach to studying data and news (Anderson and DeMaeyer 2014). To understand the operation of data within the news production process, we need to understand both how that data was embodied in particular social and material “objects” like databases, bound survey reports, and even paper documents, as well as the way that journalists understand their own professional relationship to those objects of evidence. In the second section, I want to move from the question of what an object-oriented approach might offer to the study of journalism to the reverse question: how analyzing the occupation of journalism might shed light on aspects of both our current “big data” moment as well as on general processes of fact-building within a variety of knowledge-oriented disciplines. The final section of the paper will be devoted to sketching four “critical moments” in journalism history, moments whose nuances might contribute to the genealogical understanding of quantitative journalism I propose above.

Material, Cultural, and Practice Perspectives on Journalistic Assemblage

When burrowing into quantitative journalism’s past, it is helpful to think broadly about journalism as a process of *assemblage*. Thinking about journalism as assemblage directs our scholarly attention to the interlocking *material, cultural, and practice-based* underpinnings of data journalism—perspectives that simultaneously broaden the notion of data beyond simply bites, bits, and digital traces to include a variety of evidentiary objects such as documents, maps, surveys, informational graphics, opinion polls, variable-based social science formulas, and computational techniques. Data journalism, from this perspective, has a history—but it is a more a history of jagged edges and discontinuities of practice rather than a continuing unfolding of ever more sophisticated quantitative work. Thinking of journalism as assemblage allows us to see journalism as just one among many knowledge-building process in which heterogeneous “fact-fragments” are assembled into a variety of journalistic products. News stories are assemblages of social and material artifacts

(documents, interviews, data, links, etc.), as are the very organizations that do this assembling (constructed as they are out of human workers, specific technologies, office buildings, work routines, content management systems, and so forth). This perspective on newswork, which itself draws upon Latour (2005) and others working in the tradition of actor-network theory (Anderson 2013; Domingo and Weiss 2010; Hemmingway and van Loon 2011) should be seen as less of an explanatory theory about the world and more of a particular orientation towards news production that allows us to ask interesting questions in new ways. These questions include: How does knowledge get produced? Do the changing material conditions of evidence change the mechanisms through which facts are generated and verified? And do the changing material substrates of knowledge generate different professional understandings of which facts matter, and why?

This assemblage perspective on journalistic work thus demonstrates that there are several lenses through which might choose to historicize journalistic practice, including the practices we now subsume under the label of “data journalism.” We can examine the materiality of the objects that journalists use as evidence, the meaning attributed to these material artifacts by journalists and others, and the actual work process through which these various fact fragments are assembled, and the way that materiality and culture intersect with those practices. In other words, a materialist-assemblage perspective on digital journalism would ask, first, what evidentiary facts do journalists assemble, and what facts do they ignore (what is the materiality of data journalism, and how has that material substrate changed)? Second, *how are* these fact-fragments assembled (what are the practices and routines of data journalism, and how have they evolved)? And finally, how do journalists *understand* what they are doing when they assemble certain facts in certain ways (is there a culture of data journalism, and what is the history of that culture)?

The focus on which evidentiary facts get assembled by journalists is, inevitably, partially a material focus, and I see this analysis of quantitative journalism as contributing towards the recent “material turn” in communication studies. As Diana Coole and Samantha Frost summarize in *New Materialisms: Ontology, Agency, and Politics* (2010: 2-3):

everywhere we look, it seems to us, we are witnessing scattered but insistent demands for more materialist modes of analysis and for new ways of thinking about matter and processes of materialization. We are also aware of the emergence of novel if still diffuse ways of conceptualizing and investigating material reality. This is especially evident in disciplines across the social sciences.

Has there been a similar material turn in the study of journalism? This is a particularly fraught question for communication scholars insofar as many of discussions of technology and communicative processes are met with standard warnings against “technological determinism,” usually accompanied by the invocation of the specter of Marshall McLuhan and a warning that while ordinary folk and political leaders speak of technology as driving history, more informed

and nuanced social scientists understand that most technologies are really just a social construction. As John Durham Peters has convincingly argued, however, “‘technological determinism’ or more recently ‘technodeterminism’ is a notion in desperate need of a critical intellectual history and reappraisal. It is a doctrine more often attributed than advocated” (Peters 2011). This paper tries to follow Peters’ advice and consider one particular case of the material underpinnings of communication—the role of data in the construction of journalistic knowledge—without paying ritual homage to the notion that “the social” stands as the ultimate arbiter of all materialist processes and affordances. Indeed, the field of communications and journalism studies has its own set of canonical theorists who have provided it with a rich set of writings on the relationship between media and materiality (Packer and Wiley 2012)—among them the much-maligned but much-cited Marshall McLuhan (1962), as well as Eric Havelock (1963), James Carey (1989), Harold Innis (1951), and Walter Ong (1958). Of these, it is perhaps Innis (the Canadian political economist) and Havelock (the theorist of Greek orality) who have received the least sustained attention from the broader set of new materialist theorists working in the social sciences; both of them, however, devoted particular attention to the complex intersection between orality and inscription, as well as to the tripartite relationship between materiality, spatial coordination, and temporal power. Materiality, space, and time are all deeply implicated in both the journalistic *understanding of* and *occupational relationship to* data, and thus the insights of Innis and Havelock may be particularly helpful when thinking through the full history and genealogy of quantitative journalism.

A focus on materiality alone, of course, is not enough to do justice to the full complexity of quantitative journalism’s genealogical trajectory. Material artifacts must be assembled into facts by news workers, and in this sense, materials are filtered through, though they are not reducible to, journalistic practices and routines. How do facts get *built* in journalism, both historically and in this new world of data? Are the procedures through which journalists build facts different now than they were in earlier historical eras? For contemporary practices, we can turn to ethnographic research; for past eras we must turn to the analysis of news content, the memoirs of different communication workers and social activists, and interviews with theorists and practitioners of quantitative journalism. Practices add nuance and texture to what might first appear to be a blunt, deterministic focus on materiality by showing how journalists, sociologists, and activists made use of particular technological affordances (Graves 2007) to shape discrete news products.

Finally, the third lens on the practice and history of quantitative journalism turns our attention to issues of how journalists *understand* the meaning of the material they assemble, the reasons why they draw on some kinds of facts and not others, and the way the material underpinnings of those facts affect this conception of “what counts” as important or valid evidence. Journalists could, after all, construct all of their news stories by conducting a daily séance and transcribing the voices and paranormal knockings into a coherent view of the latest news. They obviously do not, and there are important reasons why this is so. Less facetiously, American journalists have moved from valuing documents

as a record of the goings on of a distant government to embracing, in the mid-19th century, the then-alien concept of the interview (Schudson 1995) and relegating material documents (and with them, perhaps, the notion of “data”) to a minor role in news reporting. Combining these three lenses, the materiality of particular facts intersects with the routines that make use of them and cultural understandings of what facts matter, when they matter, and why they matter. By starting from an “objects of journalism” approach to the study of data, and by unspooling that approach historically, across time, we can gain deep insights into our present communicative and journalistic moment.

Periodizing Quantitative Journalism

In his article in this special issue, Coddington (2015) argues that we ought to periodize quantitative news production—dividing it into the ideal types of computer-assisted reporting, data journalism, and computational journalism—and further argues that each of these types is distinguished by a particular species of practice, an epistemological orientation, and a particular vision of the audience. The remainder of this paper can be seen as both an expansion and narrowing of Coddington’s focus—an expansion because I seek to place quantitative journalism within a longer historical trajectory that grapples with the very meaning of “the quantitative” for the production of knowledge, and a narrowing insofar as I focus primarily on the *epistemological* dimensions of these quantitative practices. In particular, this analysis tries to bring the material, cultural, and practice-oriented perspectives on journalistic assemblage to bear on developing a history of quantitative journalism.

This section sketches three historical tensions underlying journalists’ use of data, filtered through a lens that sees this cultural understanding of data as related to different journalistic attitudes toward a wide variety of material journalistic objects. First, I discuss the contrast between the idea of a journalistic “record” and journalistic “report” in the early 19th century. Second, I explore the tension between journalism as individualized narrative and more of a social science, a tension that found real, material embodiment in the use of documents in journalism over the course of the 20th century. Third and finally, I discuss how focusing on the material objects of journalistic evidence helps glimpse an emerging tension between data collection as a process by which hidden facts are brought to light, versus data collection as a process in which previously known information is organized into comprehensible patterns.

Records and Reports

What were the objects undergirding most journalism in the United States in the decades before the rise of the Penny Press in the 1830s? Primarily, ensembles of journalistic evidence before the mid-19th century were centered on documents. Although it has rarely been framed in quite this fashion, it is possible to argue that the major evidentiary shift in journalistic processes and procedures

in the middle of the 19th century consisted in transition of the form from a written to an oral form of knowledge production. Although the journalistic use of documents before the 1830s may seem remote from our understanding of the journalistic use of data in the 21st century, a materialist perspective on journalistic work can help us understand that the affinities between documentary evidence and data-oriented evidence go beyond the perceived transience and permanence of the different evidentiary forms.

Discussion of this transformation in the journalism studies literature tends to focus on *what was new* about the penny press in the 1830s: the interview. The news interview, according to a definition in the *International Encyclopedia of Communication*, is defined by its focus “on matters related to recent news events, its highly formal character, and its [management] primarily through questions and answers” (Clayman 2008: 2510). In Schudson’s (1995: 24) language, “interviewing, all but unknown in 1865, had become a common activity for reporters in the 1870s and 1880s, was widely practiced by 1900, and was a mainstay of American journalism by World War I.” Not only does the interview have a chronological history, however; it can also be examined cross-culturally. Høyer and Pöttker’s (2005) overview of the “diffusion of the news paradigm” between 1850 and 2000 is perhaps the most ambitious and wide-ranging example of this type of analysis, but credit for launching this strain of comparative research belongs to Chalaby (1996), with his work on the “Anglo-American” origins of journalism and his argument that “the interview is an Anglo-American invention.” More than simply a historical and cultural artifact, the interview also has a material and technological basis. Lee (2008), along with other scholars working broadly in the tradition of science and technology studies, has documented the importance of technological artifacts and recording technologies in the emergence of the interview in sociological research.

But what did the interview replace as the center of the journalistic evidentiary ensemble? What was the nature of the document as evidence in the colonial and early Federalist newspaper era, and what changed after the rise of the Penny Press? Around the time of the emergence of the Penny Press, journalists began to articulate a new and shifting relationship towards material evidence, an attitude that helped reorder the evidentiary value of the human beings, paper documents and news reporters that together assembled in the journalistic story. Following some brief but intriguing remarks by Hazel Dicken-Garcia (1989) in her monograph *Journalism Ethics in the 19th Century*, I contend that it is helpful to think of this shift from “paper to people” as part of a broader shift in American journalism from thinking of news products as *records* to thinking of them as *reports*. The key shift here lies at the nexus between materiality, time, and the larger structures and cultures of newsgathering; it occurred between the invention of the Penny Press and the conclusion of the Civil War. In Dicken-Garcia’s (1989: 54) words, “if information is thought of as a record, its value is principally the same whether it is a week or a year old, and this value may, in fact, increase with time. But if it is regarded as a report, recency is its most valuable quality.” The insatiable demand for information about the bloody Civil War, she argues, shifted the public appetite for news irrecoverably in the

direction of demanding news reports. This demand, in turn, activated a variety of latent potentials in the news production process. Techniques such as the eyewitness observation, the interview, and the cultivation of army officers as sources all emerged from the battlefield campgrounds of the war. Technologies also played a role, though primarily a reactive one, with already existing techniques such as photography and the telegraph assuming a more prominent role in the assemblage practices of newspapers.

While Dicken-Garcia discusses the role played by technology in news production as a largely reactive affair, it seems clear that macro-changes in material infrastructures also played a role in the transformation of the journalistic record into the journalistic report. In that light, *Journalistic Standards in the 19th Century* makes a subtle argument that the modern newspaper marked the culmination of a lengthy process of *documentary disenchantment*. When printing techniques were expensive, laborious, and time-consuming, “the tedious work of recording information confined printing to the absolutely essential; anything beyond was required to be of a nature that elevated and ennobled mankind ... a predominant view saw the press as the keeper of the record of human kind and civilization’s store of knowledge” (Dicken-Garcia 1989: 117). The phrase “disenchantment of documents” points to a radical change, one in which print was used as much to convey the report (with its trivialities, its eyewitness accounts, and insider gossip) as for a record of the activities of government, foreign events, and profit-generating market news.

Framing this shift from record to report in terms of the role and status of the document, vis-à-vis the role and status of the news interview, allows us to leaven these vaguely Havelockian musings with a more granular focus on materiality and the actual data that underlies the work of reporting the news. What kinds of documents were actually used to generate news as a record? How did these documents diffuse across the United States in the 18th and 19th centuries? What was the goal of government record-keeping? How did what Garvey (2012) has called “scissors journalism” (the cutting and pasting of already existing news reports and their interpolation into new publications) actually work? Given the fact that these cutting-and-pasting techniques lasted well into the 19th and even early 20th century, were documents really centered on record-keeping in Dicken-Garcia’s sense, or were they used in more trivial ways even from an early stage?

This tension—between paper and people, and between journalism as a record versus journalism as a report—can open our eyes to some of the questions that haunt the use of “big data” in news reporting well into our current era. One of the key questions in the use of data for journalistic purposes is whether it is extending the time horizon of reporting, shifting the focus from chronicling the daily or even hourly event in a de-contextualized fashion versus looking at longer trends over time, embedded in a variety of thickly nuanced places and times. Is there, in other words, a relationship between the materiality of the document and the idea of journalism as a record, versus the orality of the interview and the decontextualized notion of journalism as a report? In some ways, this is the implication of Dicken-Garcia’s historical analogy, and it is an

open and interesting question as to its resonance insofar as it relates to the current digitization of news.

Individuality and Social Science

These arguments relate to a second tension surrounding the use of data and documents in practices of news production: the question as to whether journalism ought to be seen as a narrative telling the story of individuals, or as a more structural mapping of trends *a la* formal social science. These fault lines crisscross journalistic history across the entire 20th century. It is possible to tease from that history a few constitutive moments (Starr 2008) that shed light on the role played by documents and data in evolving journalistic practices.

The “precision journalism” movement of the 1960s (Meyer 2012) is the natural end-point of this line of analysis, marking, as it did, the most explicit fusion of social science, data documentation, and journalistic practice. By founder Philip Meyer’s own accounting, we also need to consider the origins of precision journalism against the backdrop of other journalistic reform movements that had begun percolating in the 1960s, particularly the “new journalism” of Tom Wolfe, Truman Capote, and others. Precision journalism, which spawned computer-assisted reporting (CAR) and a variety of related technological news practices, was less about computers than it was about applying social science methods, such as sample surveys, statistical methods, and hypothesis testing, to journalism. It was also, at least initially, seen as existing in opposition to narrative reporting of the “new journalistic” variety: “The narrative journalists ... are subjective to a degree that disturbs conventional journalists and horrifies precision journalists. In essence, all the other new journalists push reporting toward art. Precision journalists push it toward science” (Meyer 2012). “For decades,” Meyer recalls,

as a precision journalist I considered narrative journalists my natural enemies. It didn’t help that the early practitioners sometimes got caught making things up. For example, Gail Sheehy wrote an article for *New York* magazine in 1973 that described in great detail the sexual and financial escapades of a prostitute in New York who was called “Redpants.” Then *The Wall Street Journal* revealed that there was no “Redpants.” Sheehy had used a composite of several different prostitutes to provide the dramatic compression needed to give her story the pace and depth of fiction. (Meyer 2012)

The rise of precision journalism was hastened by technological developments (such as advanced computing capacity and the ability to more easily tabulate and analyze large data sets) but not by technology alone. Along with the general mood of reform signaled by the new journalism and the rise of professionalization in the 1960s, we should also consider the increase in publicly accessible government record-keeping as helping to create a new way of understanding the

document as evidence, an understanding that was also tangled up in shifting technologies of information storage, duplication, and retrieval.

In some ways, however, precision journalism only marks the culmination of a number of other document/data-based trends that would appear, disappear, and reappear over the course of the modern era. Phillip Meyer can only call for journalism to become more like a social science if a historical argument has already concluded it is not such a science. Precision journalism can only emerge after the *stabilization* of the boundary between professional journalism and social science; indeed, in the early decades of the 20th century, the line dividing reformist sociology, muckraking, and academic quantification was far from clear. The early 1900s, for instance, would see the rise of the so-called survey movement, itself tied to the emergence of the progressive movement and concomitant with the growth of new techniques for collecting and visualizing social data. While this article can only gesture at the history and importance of the social survey, it relates to our genealogical excavation insofar as the progressive-era reform movement marked one of the first attempts to pioneer new mechanisms of document and data collection to understand and visualize poverty, and also cultivated ties with the local press in order to press them for favorable articles about the movement and the reprinting of striking data visualizations.

In the 1930s, as the increasing institutionalization of academic sociology led to the marginalization of progressive-era elements of academic inquiry like the survey movement, these earlier reformist impulses and visualizations found a new outlet in long-form magazines like *The Survey Graphic*. While the steady occupational differentiation of social science from social work is the primary narrative thread around which most standard histories of sociology turn, a focus on the materiality of data can also lead to a consideration of the relationship between social science and journalism, in part by examining the shifting discourse about journalism in the mainstream sociological journals, as well as the complex relationship between these journals and muckraking magazines. In particular, scholars of data journalism need to consider the manner in which those journals discussed, or did not discuss, the materiality of journalistic practices, as well as the different ways that professional journalism began to draw boundaries around its occupational role vis-à-vis both social science and reformist politics (see further discussion in Anderson 2015). In essence, while we cannot deny the relevance of precision journalism and the importance of organizations such as NICAR to the current big data movement in journalism, these developments are themselves crisscrossed by a number of additional tendencies, false starts, and aborted attempts to think through the relationship between social science and news reporting.

Arguments about the use of data to make sense of the world, the relationship between data and the paper document, and the question of how the “story” of that data should be told, are arguments at least a century old. How and why did journalism come to embrace social scientific methods, even as a minority practice? How did journalists come to incorporate a variety of socio-material practices—the regression equation, the census, the database, the public opinion

poll—into their news practices? These questions are more relevant today, in this era of quantitative journalism, than they ever have been, and understanding the history of the relationship between the profession of journalism and the profession of social science can help us gain insight into today's empirical practices.

Hidden Facts and Building Patterns

In his 1980 article “Morelli, Freud, and Sherlock Holmes: Clues and the Scientific Method,” the historian Carlo Ginzburg compares the evolution of two distinct empirical methodologies. The first centers around the discovery and analysis of the *clue*, a practice, he writes, of empirical diagnosis common to art historians like Giovanni Morelli, psychoanalysts like Sigmund Freud, and fictional detectives like Sherlock Holmes. “In all three [of these] cases,” Ginzburg notes, “tiny details provide the key to a deeper reality, inaccessible by other methods” (11). The procedural focus of this method is on “the qualitative, the individual case or situation or document itself” (15). The second evidence-building method, one that Ginzburg attributes to a more abstract scientific culture, “use[s] mathematics and the experimental method ... to measure and repeat phenomena” (15). “In the first decades of the 17th century,” Ginzburg writes, “the influence of this [second], Galilean model ... would lead towards the study of the typical rather than the exceptional, towards a general understanding of the workings of nature rather than particularistic, conjectural knowledge” (20). It would lead, in short, to the displacement of the clue and the individual detail in favor of mathematical models of regularized phenomena.

Ginzburg contends that a reversal of this generalizing tendency—a re-animated focus on characters and clues—began in the 19th century. This rebirth of the “clue-method” thus existed alongside, though it did not displace, simultaneous attempts to understand collectives as generic, law-governed aggregates. “The time at which these [clue-based] principles came after so long to fruition was perhaps not altogether random,” Ginzburg notes. “It coincided with the emergence of an increasingly clear tendency for state power to impose a close-meshed net of control on society, and once again the method used ... involved attributing identity through characteristics which were trivial” (24). The rise of the clue, Ginzburg concludes, coincided with a rapidly emerging state interest in understanding people and objects as *individual and unique*.

While Ginzburg discusses the emergence of the idea of “the clue” in psychology, political economy, art criticism, and detective fiction, historian David Paul Nord (1990) analyzes the invention of “reportorial empiricism” in early journalism itself. Examining the overlap between 17th century Puritan sermons, with their simultaneous focus on bizarre events and the placement of those events within a general scheme of divine history, and an early New England journalism that did much the same thing, Nord probes the roots of the “eclectic, reportorial method of inquiry” (28). He thus describes a method of knowledge construction that

bears a remarkable similarity to those described, in a different context, by Ginzburg:

By reportorial empiricism, I mean that the teleological news literature of seventeenth-century New England was highly empirical, but the style was eclectic and reportorial, not systematic and scientific. The methodology was essentially what we today call “news reporting”: the routine collation and citation of statement and sources. The sources ranged widely, from leading scientists to folklore to average people with stories to tell. The role of the writer was not to conduct systematic empirical research, but rather to report the empirical statements of others. Such a methodology was empiricism without science. It was, in a word, journalism. (Nord 1990: 38)

Nord’s historical research—while fascinating in its own right—also points us towards the contemplation of some thoroughly modern questions: Why has journalism placed so much value in the unique, the bizarre, and the individual rather than in the regular, law-like, and ordinary? How did it come to pass that such an important instrument of knowledge came to be dominated by procedures that strike most social scientists, and many outside observers, as empirically inadequate, to say the least? What do changes in these procedures—a greater reliance on data and algorithms, for instance—portend for the future of news reporting?

A more current way to consider these questions would be to compare the role played by the Pentagon Papers with the huge cache of files published by WikiLeaks known as the Iraq War Logs. The Pentagon Papers, in Ginzburg’s terms, were a “clue”—a previously hidden evidentiary fragment that provided insight into a larger social, economic reality. The Iraq War logs, while arguably hidden sources of information, did not obtain their value from the scarcity or their uniqueness, but rather the way that they worked as a pattern-creating ensemble of evidence that “made sense” out of data that was already known. Just as Watergate and the resignation of Richard Nixon became a central cultural touchstone for an increasingly powerful class of professional Washington reporters, so too the Pentagon Papers would mark a key moment in which a particular vision data and the document would be ratified within the journalism’s collective memory. The remembered story of the Pentagon Papers would crystallize a particular journalistic understanding of information as both *scarce and hidden*, with the document serving as a particularly powerful materialization of this hidden, yet certified, knowledge. This understanding of documentary evidence becomes even clearer when we turn to *Custodians of Conscience*, an in-depth examination of the practices of investigative journalism and their connection to various styles of moral and social inquiry (Ettema and Glasser 1998). Ettema and Glasser discuss the previously noted ontological status of the document at some length in their work, particularly the manner in which the accumulated “weight” of documentary evidence represents a key inflection point for ascertaining controversial truth claims about secret behavior.

The idea of the document as a signal flare revealing the central aspects of a hidden truth stands in contrast to an alternate idea of the document as only one scrap of evidence that goes into making up a larger—and not necessarily hidden—world of information. Herein lies the third tension in our genealogy of data journalism. On the one hand, we can see data, and the documents that underlie that data, as a hidden fragment of information waiting to be uncovered. On the other hand, we might envision data as a thing that is both massive and already known, where the journalistic value-added lies not in unmasking of a hidden truth but in putting overwhelming torrent of information into patterned context. Neither of these cultural belief systems is necessarily truer than the other, but both are subject to professional negotiation, and both represent different ways of configuring technologies, institutions, and organizational practices in the digital information era.

Conclusion

This article proposed that we historicize the relationship between journalism and big data in order to subject that relationship to the genealogical gaze. It has further argued that we need to think of data as a particularly material procedural substrate—that is, we need to consider the material objects (whether interviews, documents, human observations, or other objects) that underlie journalistic processes. The payoff of thinking about data in this way is that it allows us to productively probe the tensions that lie at the heart of the journalism–data nexus, tensions that include, but are not limited to, thinking of journalism as a record versus a report, a social science versus a narrative about individuals, and a hidden source of secret information versus an already known stream of un-patterned knowledge.

I promised early in this article that an additional payoff for thinking about new forms of journalism genealogically might be that it could help provide a normative handle on data journalism, or at least answer the “so what” question about why this kind of journalism matters. To some degree, of course, it is dangerous to draw the implications of data journalism out too strongly; it is far from the dominant form of journalistic work today, as several other articles in this special issue make clear. Nevertheless, while Michael Schudson’s (2013) trenchant observation that “journalism is neither all about data nor all about stories” is certainly accurate, it is also impossible to ignore the fact that the center of gravity between journalism-as-data and journalism-as-story *does* shift, and means different things at different times. That there is a difference in emphasis between the New Journalism of Tom Wolfe and the Precision Journalism of Philip Meyer seems beyond dispute, as does the fact that “audience data” as filtered through Chartbeat and audience dialog filtered through the lens of public journalism are dramatically different ways of understanding “the public.”

Will the tension between “story” and “data” still be meaningful in a dozen years? To posit an initial answer to that question, by way of a conclusion, some perspective drawn from the past decade of journalism history might be helpful. In

2003, the idea that the “people formerly known as the audience” could contribute something to journalistic production was a radical notion, whether that argument was advanced by journalism scholars, technologists, activist journalists, or members of the “blogosphere.” Today, in a world dominated by social media and the perpetual, occasionally newsworthy status update, it seems hard to imagine a journalistic world where the audience did not have something to contribute to the production of news. What is radical about journalism now, in a world of digital media, may be its very *conservatism*: the fact that it exists at all as a relatively professionalized cadre of public information producers whose agenda is not entirely determined by the wishes of the audience. What was once a conservative liability has now become a useful pushback against other trends in the digital world.

Ten years from now, we may look back on the big data debate as a similarly quaint exercise. It is highly likely that the world we live in will be filled with data of various kinds, with an increased focus on measurement, outcomes assessment, and increasingly narrow news production tailored to the consumption of niche audiences. In that world, what matters about journalism may not be the degree to which it embraces big data, but rather the ways in which it reminds us of *other* forms of information that are not data, *other* types of evidence that are not quantitative, and *other* conceptions of what counts as legitimate public knowledge. These other forms of knowing are possible because they once existed, and thus, they can exist again.

C.W. Anderson, Department of Media Culture, College of Staten Island (CUNY).
USA. christopher.anderson@csi.cuny.edu

REFERENCES

Anderson, C. W. 2015. Forthcoming.

Anderson, C. W. (2013). *Rebuilding the News: Metropolitan Journalism in the Digital Age*. Philadelphia, PA: Temple University Press

Anderson, C.W., and Juliette DeMaeyer. 2014. "Introduction: Objects of Journalism." *Journalism: Theory, Practice, Criticism*. [Forthcoming]

Carey, James W. 1989.

Chalaby, Jean (1996). Journalism as an Anglo-American invention: A comparison of the development of French and Anglo-American Journalism, 1830s-1920s." *European Journal of Communication* (11): 303-326.

Clayman, Steven E. (2008) 'Interview as Journalistic Form.' In Wolfgang Donsbach, ed. *International Encyclopedia of Communication*. Oxford: Blackwell: 2509-2513

Coddington, Mark. 2015. Clarifying journalism's quantitative turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism*.

Coole, Danah, and Susan Frost (eds). 2010. *New Materialism: Ontology, Agency, Politics*. Raleigh, NC: Duke University Press.

Dicken-Garcia, Hazel. (1989). *Journalistic Standards in Nineteenth-Century America*. Madison, WI: University of Wisconsin Press.

Ettema, James. S., and Theodore L. Glasser. 1998. *Custodians of conscience: Investigative journalism and public virtue*. New York: Columbia University Press.

Finnegan, Cara. 2003. "Social Welfare and Visual Politics: The Story of Survey Graphic." Online at <http://newdeal.feri.org/sq/essay.htm>, accessed 6/13/2014

Garvey, Ellen. 2012. *Writing with Scissors: American Scrapbooks from the Civil War to the Harlem Renaissance*. New York, NY: Oxford University Press

Graves, Lucas. 2007. "The Affordances of Blogging: A Case Study in Culture and Technological Effects." *Journal of Communication Inquiry* 31(4): 331-346

Havelock, Eric. 1963. *Preface to Plato*. Cambridge, MA: Harvard University Press.

Hemmingway, Emma, and Joost van Loon. 2011. "We'll always stay with a live, until we have something better to go to...": The chronograms of 24-hour television news." *Time and Society* 20(2): 149-170.

Høyer, Svennik, and Horst Pöttker (eds.). 2005. *Diffusion of the News Paradigm 1850-2000*. Göteborg, Sweden: Nordicom.

Innis, Harold. 1951. *The Bias of Communication*. Toronto: University of Toronto Press.

Ong, Walter. 1958. *Ramus, Method, and the Decay of Dialogue: From the Art of Discourse to the Art of Reason*. Chicago: University of Chicago Press.

Packer, Jeremy, and Sean Wiley. 2012. "Strategies for Materializing Communication." *Communication and Critical/Cultural Studies* 9(1): 107-113

Latour, Bruno. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. New York, NY: Oxford University Press.

Lee, Raymond M. 2008. "David Riesman and the Sociology of the Interview." *The Sociological Quarterly* 49(2): 285-307.

McLuhan Marshall, 1962. *The Gutenberg Galaxy: The Making of Typographic Man*. Toronto: University of Toronto Press.

Nietzsche, Frederich. 2006 [1887]. *On the Genealogy of Morality and Other Writings (Cambridge Texts in the History of Political Thought)*. Cambridge, UK: Cambridge University Press.

Nord, David P. 1990. "Teleology and the News: The Religious Roots of American Journalism." *The Journal of American History* 77(1): 9-38

Peters, John Durham 2011. "Two Cheers for Technological Determinism." *Annenberg Research Seminar*, September 19, 2011.

Schudson, Michael. 1995. *The Power of News*. Cambridge, MA: Harvard University Press.

Schudson, Michael. 2013. "A Spotlight, Not a Truth Machine." *Neiman Journalism Lab*. Online at <http://www.neimanlab.org/2013/12/a-spotlight-not-a-truth-machine/>, accessed August 17, 2014.

Stewart, Kristen. 1946. "The Freedom to Read: Newspapers." *The Survey Graphic* 39(December 1946): 452-456.

Weiss, Amy Schmitz, and Domingo David, 2010. Innovation processes in online

newsrooms as actor-networks and communities of practice. *New Media and Society* 12(7): 1156–1171.