

Alexander von Humboldt
INSTITUT FÜR INTERNET
UND GESELLSCHAFT

HIIG Discussion Paper Series

Discussion Paper 2012-01

# Notes Towards an Analysis of Computational Journalism

26 October 2011

**C.W. Anderson**
Christopher.anderson@csi.cuny.edu
City University of New York,
Department of Media Culture
College of Staten Island,

Electronic copy available at: http://ssrn.com/abstract=2009292

## Abstract

This article advances a critical research approach to computational journalism. By "computational journalism" the article refers to the increasingly ubiquitous forms of algorithmic, social scientific, and mathematical forms of newswork adopted by many 21st century newsrooms and touted by many educational institutions as "the future of news." By "critical approach," the article endorses a research model that brackets, at least temporarily, many of the current industry concerns with the practical usability of newsroom analysis. The bulk of the article outlines a series of six lenses through which such a critical approach to computational journalism might be carried out. Four of these lenses are drawn from Schudson's classic typology of the sociology of news—economic, political, cultural, and organizational approaches. In addition, the author adds Bordieuean field approaches and technological lenses to the mix. In each instance, the author discusses how particular approaches might need to be modified in order to study computational journalism in the digital age.

Electronic copy available at: http://ssrn.com/abstract=2009292

# Contents


1. Introduction 1

2. Internalist Tendencies in Computational Journalism Research 2

3. Six Approaches to a Sociology of Computational Journalism 7

3.1 Politics and Public Policy 7

3.2 Economics 9

3.3 Institutions and Fields 11

3.4 Organization-Level Dynamics 13

3.5 Cultural History 14

3.6 Technology and the News 15

4. Conclusion 17

5. References 19


Electronic copy available at: http://ssrn.com/abstract=2009292

# 1. Introduction

Understanding the interaction between large-scale data collection, algorithmic analysis, computational practices, and the production of public knowledge is one of the central methodological and philosophical challenges of our time (Manovich 2011). The possibilities of so-called "big data" can be glimpsed in the breadth of its potential application; not only have computational practices helped facilitate the growth of the academic subfield known as the digital humanities (Berry 2011; Dalbello 2011; Hayles 2011; Rosenbloom 2010; Slauter 2010,) these practices have also given rise to popular terms like "culturomics" (Michel et. al 2011), terms whose diffusion into mediated discourse (Cohen 2010) have been in part facilitated by the increasing availability of quasi-public analytical tools. This paper is an attempt to come to terms with the scope of the algorithmic revolution in knowledge production by narrowing our focus and examining one particular knowledge-generating activity: the field of journalism and the production of news. Specifically, I analyze what scholars and practitioners have labeled "computational journalism"— "the combination of algorithms, data, and knowledge from the social sciences to supplement the accountability function of journalism" (Turner and Hamilton 2009, 2). As it reviews the relatively scant literature on computational journalism that already exists, and offers a potential framework for future analysis, this paper is synthetic and prescriptive rather than empirical. But, as a critical perspective on the burgeoning field of digital humanities in general (and on the practices of computational journalism in particular) can only be maintained through empirical research, I hope this overview will serve as an impetus for such future empirical work. Finally, insofar as it adopts four long-standing sociological frameworks for the analysis of news and also supplements these frameworks with two additional approaches, this paper makes a more discipline-oriented argument about the proper ways to study journalism and news production in the digital age.



Electronic copy available at: https://ssrn.com/abstract=2009292

## 2. Internalist Tendencies in Computational Journalism Research

When I say this paper adopts a critical perspective on computational journalism, I mean to indicate that it adopts a scholarly stance that not only makes itself useful to, but also maintains a skeptical distance from, current developments in our rapidly digitizing world. The idea of computational journalism has already affected the practices of news production. Leading educational institutions in the field of journalism education have taken notice of the computational journalism trend. In 2007 Georgia Tech University began teaching classes on "computational journalism" (Irfan and Diakopoulos 2007). Since 2008, Northwestern University's Medill School of Journalism has, with support from the Knight Foundation, funded a small cadre of "people with strong technology skills who are interested in pursuing a journalism master's degree" (Medill.northwestern.edu 2008) while Columbia University's Graduate School of Journalism is aggressively touting its' new five semester journalism-computer science joint degree (Journalism.columbia.edu 2011). Paralleling institutional developments at universities, the transnational grassroots journalism organization Hacks/Hackers was founded in 2009 "to create a network of journalists ('hacks') and technologists ('hackers') who rethink the future of news and information" (Hackshackers.com 2010). And in some elite newsrooms, so-called "web ninjas" (Taylor 2009) are already beginning to transform journalistic practice from the inside out. The New York Times Interactive Newsroom Technologies group—described by *New York Magazine* as "a group of developers-slash-journalists, or journalists-slash-developers, who would work on long-term, medium-term, and short-term journalism"—is one particularly prominent example of the application of computational and visualization practices to newswork (Nussbaum 2009), and similar teams are now working at organizations like the Huffington Post (Stoeffel 2011) and ProPublica (Ngyuen 2010).

Contemporary press accounts of these developments often adopt a particular, and particularly dramatic, narrative structure: while technology is destroying journalism, technology contains within it the seeds of journalism's rebirth. Coverage of the *Times'* Interactive Newsroom Technologies group by New York Magazine begins with a lament that 2008 was "catastrophic for the



Electronic copy available at: https://ssrn.com/abstract=2009292

New York Times," in part due to large-scale technological changes. But, it goes on:

> *"Even as the financial pages wrote the paper's obit, deep within that fancy Renzo Piano palace across from the Port Authority, something hopeful has been going on: a kind of evolution. Each day, peculiar wings and gills poke up on the Times' website—video, audio, "drillable" graphics. ... Throughout the election, glittering interactive maps let readers plumb voting results. There were 360-degree panoramas of the Democratic convention; audio "back story" with reporters like Adam Nagourney; searchable video of the debates* ... [these changes might] mean this terrible year was not for nothing: that someday, this hard era would prove the turning point for the paper, the year when it didn't go down, when it became something better." *(Nussbaum 2009,* np, emphasis mine*)*

What technology taketh away, technology giveth—only differently, and perhaps even better than before. Such, at least, is the hopeful story of computational journalism. And the few scholarly articles that have thus far been written about computational journalism generally echo this narrative arc. Turner and Hamilton state the matter plainly: "In recent years, ubiquitous computation has transformed the landscape of journalism. It has undermined business models, rebalanced the relative power of reporters and audiences, and accelerated the delivery of information worldwide." And while "computational journalism cannot transform the business situation of contemporary journalism, it can create new tools that may reduce the cost of watchdog reporting in certain circumstances, take better advantage of the new information environment, and ultimately help sustain watchdog work during the technological sea change now under way." Other scholars, such as Cohen et. al. (2011), write in similar terms about the perils and promise of new computational technologies.

I would argue that the presence of this "rise and fall" narrative within many scholarly accounts of computational journalism betrays the genre's *internalist* tendencies at this early stage of academic research. By "internalist" I mean a tendency to consider the problems of journalism scholarship from the point of view of the journalism profession. Articles by Cohen (2011) and Daniel



Electronic copy available at: https://ssrn.com/abstract=2009292

et. al (2010) are representative of this larger tendency. Daniel and her co-authors even title their piece, "The Promise of Computational Journalism," and spend the bulk of their paper enumerating the various ways in which computational practices might benefit the profession of journalism: they claim it will increase the depth of investigative reporting, differentiate digital brands, aid the bottom-line of mass market news providers, and even " minimize the risk of incorrect reporting [by] providing ... a factual basis for analysis" (Daniel et. al. 2010, 2). It is not surprising that, in this moment of occupational crisis, journalism scholars would be concerned with assisting journalism in addition to understanding it. But attempts at assistance can impede as well as assist. If Daniel et. al. are correct in their assessment of the financial impact of computational journalism, for instance, is it really an unalloyed positive development that, according to her argument, adopting computational practices will aid larger, legacy news outlets when compared to their smaller competitors? What does this say about the structure of the journalistic field as currently constituted? And if certain aspects of factual analysis might be aided by computational methods, is it possible that other journalistic methodologies will be impaired? In their rush to assist an occupation in crisis, these authors consider these sorts of trade-offs only briefly.

One way to re-orient this tendency towards internalism, without losing sight of the fact that scholarly research on computational journalism might in fact be useful for the journalism industry, is to leaven the professional focus of much journalism research with perspective from other academic fields. Here again, however, the early record of interdisciplinary research is rather spotty. The scholarship on computational journalism that has drawn the most from other academic fields has emerged (not surprisingly) out of the nexus between journalism with computer science. And while this intersection has produced some of the most interesting and rigorous scholarship on computational journalism to date, the research has, like journalism scholarship in general, been primarily concerned with "building things." This computational journalism research has primarily concerned itself with, first, understanding the dynamics of contemporary news practices and, second, designing digital tools that can supplement, routinize, or algorithmically expand the scope of these traditional practices. A series of genuinely path-breaking articles by Diakopoulos et. al. (Diakopoulos et. al. 2010; Diakopoulos et. al. 2011;



Electronic copy available at: https://ssrn.com/abstract=2009292

Diakopoulos and Naaman 2011a) represent well both the promise and the limitations of computational journalism research that draws primarily on the discipline of computer science; the primary focus with the primary focus of this research remains on the construction of a particular set of digital newsroom devices, while the impact of computation on larger social, political, organizational, and cultural currents in journalism goes unexamined.

The prominence of this particular scholarly habitus— this focus on the building, testing, and implementation of digital tools designed to solve particular institutional and occupational problems— speaks to some of the resonances between the cultures of journalism and computer science. In the current overlap between the two fields, it seems to me, there is both an admirable concern with results and a less admirable dismissal of approaches deemed to be overly discursive, abstract, or "scholarly." There is hostility, in short, towards too much talk. A pointed example of this practical attitude can be seen in a blog post by prominent "hack/hacker" Matthew Waite, who, after leading development work on the Pulitzer Prize winning website Politifact, summarized his guiding attitude as one of "demos not memos" (Waite 2009). The slogan has become a popular one in the journalism development community, often quoted in exasperation at moments when the discourse about the "future of journalism" appears overly abstract or intellectualized.[1] It is beyond the scope of this paper to analyze the roots and implications of this "just build it" attitude amongst the current crop of computational journalists, although research by Dossick and Neff on the differences between "clean tech and messy talk" points toward a useful example of this kind of analysis as applied other engineering-oriented domains (Dossick and Neff, forthcoming).

The disciplinary implications of this attitude, on the other hand, *are* one of the concerns of this paper. In the remainder of this article, I want to advocate on behalf of a more interdisciplinary and externalist perspective on computational journalism research. I argue that understanding the relationship between big data, algorithmic analysis, and journalism can be approached from a number of directions, all of which bracket the question of whether

[1] A thoroughly unscientific example: in the summer of 2011, the author of this paper initiated a conversation on a popular social network about hacker-journalists attitudes towards theory. In the course of three days, multiple internal "slogans" were offered up by a variety of workers in a variety of newsrooms, each of which expressed a negative attitude towards overly-abstract discourse.



Electronic copy available at: https://ssrn.com/abstract=2009292

computation is "useful" for journalism and at least temporarily forgo the pragmatic construction of digital tools. If computational practices are the salvation of journalism, this salvation is certainly a double-edged sword and depends on a number of social institutions and technological artifacts that lie outside of journalism itself; in a similar fashion, the creation and adoption of computational tools is as much a research problem for sociologists and anthropoligists as it is for engineers. The kind of analysis I advocate here would pay careful attention to the means by which public policies and struggles over the transparency of open-government data impinge on newswork. It would examine, in detail, the unequal distribution of computational resources in 21st century journalism and critically dissect how this inequality is either impeding or facilitating journalism's professional mission. It would look at the dynamics of the journalistic field and the manner in which struggles by entities outside that field are impacting the diffusion of computational practices. It would pay careful attention to the organizational dynamics and work processes that either facilitate, or hinder the adoption of computational techniques inside the newsroom. It would engage in a careful, historically grounded analysis of journalistic culture and the role that culture plays in our understanding of what counts as "news." It would, finally, bring a nuanced, socio-material perspective on technology to bear on questions regarding the relationship between computational technology and journalistic "sense-making."

Astute readers will have noted note that four of the lenses I advocate here— political, economic, organizational, and cultural frames— have long been considered to be the "classic" perspectives within the sociology of news (Schudson 2005) while two of them— technological and institutional, or "field," perspectives— though not classic, are increasingly utilized in 21st century journalism research. The remainder of this article, then, applies Schudson's "four approaches to the sociology of news" to a new object of analysis: computational journalism. Insofar as it adds two new frames to the previous four, this article serves as both an affirmation and gentle critique of the "four approaches" perspective; while affirming the value of these classic lenses, I argue that two new additional frames may also be helpful in understanding journalism in an increasingly trans-institutional, technologically mediated age. The purpose of this categorical analysis, in short, is to both examine the possible lenses through which to analyze a major new example of digital



Electronic copy available at: https://ssrn.com/abstract=2009292

knowledge production in the 21st century, and to highlight some of the early work that has already been done in this vein.

# 3. Six Approaches to a Sociology of Computational Journalism

Michael Schudson's "Sociology of News Production," first published in *Media, Culture, and Society* in 1989 and revised several times since then, remains one of the key citations in any sociological analysis of journalism and news. The article has actually undergone several important mutations over the course of its 20-plus year history, with the key changes being outlined rather frankly in the latest incarnation of the piece ("Four Approaches to the Sociology of News," published in the equally oft-revised *Mass Media and Society* [5th edition]). The different descriptions of each of the categories amount to more than simply revisions for the sake of revision, particularly in the latest piece, which signals its intentions by announcing the "approaches" to studying news have now expanded from three to four. The relevant nuances of these category changes will be explored below. An equally important change in Schudson's thinking—the argument that journalistic products are as real as they are "socially constructed"— lies outside the scope of this current paper, but would be valuable to analyze in its own right.

What, then, are the different lenses though which we can analyze the sociology of news production, and how can each of them be applied to the study of computational journalism?

## 3.1 Politics and Public Policy

In its simplest form, the "political context of news-making" can be used to study the production of news on a macro or state-level, particularly insofar as different political systems co-produce different systems and styles of journalism. Political approaches to the study of news usually emphasize one or two unique, meta-theoretical points. First, the approach emphasizes that journalism is not co-terminous with democracy; contra Carey (Carey 2000), undemocratic states may posess their own, unique journalistic forms. Second and more importantly, the political approach to the analysis of journalism has been especially useful in the comparative study of news practices, particularly



Electronic copy available at: https://ssrn.com/abstract=2009292

the different practices that exist within different democratic systems. Beginning with Hallin and Mancini's 2004 *Comparing Media Systems* (Hallin and Mancini 2004) a growing and persuasive body of research has pointed to variations within capitalist democracies (between France, Great Britain, and Germany, for instance) as being particularly important causal factors accounting for differences in news (Benson 2010; Hallin 2004; Hanitzsch 2008; Waisbord 2000) In part, this is a reaction to earlier generations of news research that tended to posit "democratic, capitalistic" journalistic systems as structurally homogenous when compared to the journalism of socialist or communist nation-states (Siebert et. al 1956).

At first glance, it is difficult to see how this political perspective might be at all useful in the analysis of computational journalism. Is computational journalism not a micro-level or (at best) an organization-level phenomenon? How can it be studied in its political context without losing sight of what makes it unique? There is some truth to this claim. On the other hand, it is precisely by trying to fit the development of computational journalism within a macro-political frame that our attention is drawn to aspects of it that we might otherwise neglect. When analyzing computational journalism through a political lens, scholarship should initially focus less on comparative differences in political structure and more on the bureaucratic, policy-level initiatives that either allow computational journalism to thrive, or retard its growth. Developments in the world of so-called "open government initiatives" the role of large-scale databases in the crafting of public policy, the transparency and accessibility of government data and its use in democratic decision-making, government surveillance of online data and political activities; all these state-level initiatives may ultimately influence the forms of computational, data-driven journalism that news-organizations adopt.

In general, work on internet-afforded public policy looks at one of two phenomena— either digitally enhanced citizen-participation (Noveck 2009) or the more passive activity of making government data easily accessible and searchable via digital archives. While "open" and "wiki" government initiatives have received much attention from policy-makers activists, and reporters, the scholarly community has yet to devote much attention to these phenomena. Noveck (2011) has been an exception, devoting the last several years to both open government policy-making in Washington D.C as a member of the



Electronic copy available at: https://ssrn.com/abstract=2009292

Obama administration as well as academic research. Her work, however, spends little time discussing the relationship between wiki-government and journalism, and has little to say about news. One of the few academic pieces to directly tie developments in digital transparency to journalistic practices comes (again) from Schudson, who argues that the growth of what he calls

> *"political observatories [...] and the new availability of databases for public-interest research [...] represent just one feature of the future for news. Yet it is a vital feature, and so far it has received little general notice. Political observatories do not replace journalists, nor do databases shove narratives aside. But the observatories are increasingly valuable partners for journalists, and databases lay new foundations for narrative. Both offer promise for developing the kind of public information that makes democracy possible." (Schudson 2010, 107)*

Of course, the flipside to digitally afforded public transparency is digitally-enabled surveillance, as Kreiss and Howard remind us (Kreiss and Howard 2010). But it is this complex overlap, embedded within a deep and nuanced understanding of that journalism's political context, which a critical analysis of computational journalism would attempt to understand. The public policy decisions that enable or retard the growth of computational journalism are surely a ripe area for research. Ultimately, given a lengthy enough amount of time, scholars might even be equipped to relate differences in computational practices to system-level differences between a variety of political contexts.

## 3.2 Economics

Inside in a large conference room on the upper-floor of a center city office building, several employees were grilling the Vice-President of Philly.com, the news website for Philadelphia's two daily newspapers, on her hiring priorities. The year was 2008, and the staff of Philly.com and the newspapers had gathered for an afternoon to plot their future. "You've been hiring all these web producers," one reporter complained, "these kids who just take newspaper content and move it around Philly.com. But what it really seems we need are



Electronic copy available at: https://ssrn.com/abstract=2009292

developers, computer programmers, silicon-valley types. Why aren't we hiring them?"[2]

There was a pause and some uncomfortable throat clearing. "Well," the Philly.com Vice-President finally ventured, "the problem, you see, is that the web producers I hire are cheap. But the tech team you and I both know we *need* to hire? They aren't cheap. I can hire six or seven web producers for what it costs to hire a single tech person. And we really just can't afford to do that."

Traditionally the analysis of the "economic organization of news," like the analysis of the political context of news making, has been considered on a macro-level. With its origins in both Cold War dynamics (the division of the world into socialist and capitalist economic systems) and the political-economy critique of concentrated media systems, the economic analysis of news production (Baker 2002; McChesney 1995; Siebert 1956) might seem a poor way of understanding computational journalism. But just as in the political lens discussed above, drawing on economic categories to understand the development "big data" journalism might point our analysis in directions we might not immediately expect. Rather than simply focusing on the manner in which large economic systems intersect with various forms of journalistic production, scholars might critically interrogate the way that different institutionally specific resources constrain the options available to various news outlets and industry segments. Such scholarship could attempt to correlate forms of computationally enhanced news production with levels of institutional economic capital. Researchers in this tradition would take seriously the lament of the local newspaper executive, above, the fact that certain technologically-focused innovations appeared out of reach for less wealthy news organizations, and they would interrogate that lament by examining the role played by those seeking to level the playing field-- open-source software makers and volunteers-- in the more evenly spread diffusion of technology.

To date, there has been next to no work done on the relationship between economic resources and computational journalism. Nevertheless, white-papers like the recently released FCC report on the "Information Needs of Communities," with its emphasis on capacity gaps between local and nationally focused news organizations, can serve as a starting point for a less

[2] The following anecdote is taken from the author's newsroom fieldwork, conducted between 2007 and 2011.



Electronic copy available at: https://ssrn.com/abstract=2009292

macro-level, more institutionally grounded analysis (Waldman 2011). Of course, a retreat from systemic economic perspectives does not obviate the feasibility of a systemic analytical lens over the long term; like the political context lens, drawing stronger and more generalizable correlations between economic system and journalistic form might be possible over time. In the interim, however, such correlations would require extensive, critically focused empirical work on the relationship between economics and newsroom innovation.

## 3.3 Institutions and Fields

The shifting focus in the political and economic perspectives discussed above, away from macro-level generalizations and towards institutionally-grounded analysis, points us toward a third lens that might be useful in understanding computational journalism: the so-called "field" or "institutional" perspective. Schudson does not discuss field analysis in his classic series of papers on the sociology of news; indeed, Rodney Benson, who has led the way in importing field perspectives into journalism research, advances mezzo-institutional field research as part of a critique of classic sociological lenses on the production of news, including Schudson's (Benson 1999; Benson 2004). I do not attempt to take sides in this dispute here, except insofar as I argue that *any one of the three scholarly lenses discussed so far — political, economic, and field — may be useful at different times and insofar as they attempt to answer different questions*. As a new object of study, scholarship on computational journalism would be best served by an openness to a variety of methods and theories. And understanding the development of computational journalism from the perspective of the journalistic field has its own advantages, as well as its own drawbacks.

Drawing on Bourdieu's highly influential work on the sociology of institutional power and applying it to the study of news, Benson argues that

> *"The starting point for understanding the media field paradigm is Pierre Bourdieu's general theory of fields (champs). Drawing on and modifying Weber's sociology of religion, Bourdieu sees society as differentiated into a number of semi-autonomous fields (e.g., fields of politics, economics, religion, cultural production, etc.) governed by their own "rules of the game" and offering their own particular economy of*



Electronic copy available at: https://ssrn.com/abstract=2009292

*exchange and reward, yet whose basic oppositions and general structures parallel each other." (Benson 1999, 466)*

"Journalism," according to this model, would best be understood as its own "field of power" whose general social dynamics are conceptualized as positional and relational. These dynamics are based on the possession of a particular "habitus" as well as a stock of capital. Institutions (and individuals) would struggle to possess certain levels of socially-defined capital, as well as to alter the definition of what counts as legitimate capital in the first place. Understanding computational journalism in this fashion has at least two benefits. First, it adds a vector of power dynamics to an area of socio-technical life (technological innovation) too often understood from within an "all boats will rise" mentality. Second, and perhaps more importantly, it understands journalism relationally, placing it in the context of other institutional clusters and fields.

This analysis of computational journalism from a Bourdieuean or new institutionalist (Ryfe 2006) perspective would begin from the premise that a new field, the field of data journalism, is emerging alongside traditional fields such as computer science and news reporting. It would study the social, economic, and cultural power struggles that occur inside this field, and the way these struggles shape practices of newswork and journalistic products. Importantly, however, the dynamics of this field would never be seen as operating in isolation from other nearby fields. Developments in traditional journalism, or in computer science, would obviously play key roles in the development of data journalism. And clusters of seemingly unrelated institutions, like non-for-profit foundations, would also play a key role in the shape of the computational journalistic field. Lewis' work analyzing the relationship between the Knight Foundation, news institutions, and cultural concepts of journalistic professionalism (Lewis 2011) is a particularly impressive example of this kind of research; similar methodologies and perspectives might be applied to the study of the role of the Knight Foundation, open source technology groups like the Mozilla Foundation, and other funders in the development of the computational journalistic field.



Electronic copy available at: https://ssrn.com/abstract=2009292

## 3.4 Organization-Level Dynamics

Despite the utility of political, economic, and field approaches to the study of computational journalism, each of these perspectives abstracts (to a greater or lesser degree) from the day-to-day organizational processes by which computational practices embed themselves in journalistic work. A fourth perspective, "the social organization of newswork," (Schudson 2005) takes us back down to earth, to ground-level newsroom dynamics. Research operating via this lens would examine how workflow routines, levels of technological adaptation, individual rivalries, bureaucratic divisions, and daily process imperatives affect the manner in which newsrooms integrate potentially far-reaching new technologies into their workflow. While earlier research of the impact of technology on journalistic practice often posited a dynamic in which new technologies generate distinct editorial effects, Boczkowski (2004) has provided researchers with a far more nuanced model by which to understand the relationship between technology and newswork. For Boczkowski, production factors (organizational factors, work routines, and representations of users) shape adoption processes, which are themselves afforded by technological changes. Only at the end of this far more complex process do distinct editorial products begin to emerge.

Most newsroom ethnographies either explicitly or implicitly adopt an organizational routines perspective, with detail-rich, on the ground observations often bringing to light just how slow an halting organizational change actually is. Alongside classic work in this vein (Epstein 2000; Gans 1979; Tuchman 1978), the last decade has witnessed a veritable outpouring of ethnographic newsroom research **(**AUTHOR 2010a; Boczkowski, 2004a; Cottle, 2007; Eliasoph, 1997; Klinenberg, 2005; Paterson and Domingo, 2011) research often prompted by macro-level technological change but whose conclusions have often been that ground-level journalistic changes are far less dramatic than might be assumed. And while much of this research has looked at technology in general— the use of crowdsourcing, blogs, Twitter, etc— little of it has explored computational journalism, if by computational journalism we mean "the combination of algorithms, data, and knowledge from the social sciences to supplement the accountability function of journalism" (Turner 2009). Royal's provisional research, which analyzed the organizational integration of the *New York Times* Interactive New Department into larger work



Electronic copy available at: https://ssrn.com/abstract=2009292

routines at the *Times,* is one of the few exceptions (Royal 2010). There is much room for additional scholarship in this area. Not only would it be easy to apply recently reinvigorated ethnographic methods to the study of computational practices, but it would be intriguing to see whether differences existed in the findings of earlier and later digital newsroom ethnographies, and what these differences might say about journalism in a digital age.

## 3.5 Cultural History

The intersection of technological change and newsroom practice is most often analyzed on the level of organizational routines; nevertheless, the growth of big datasets and computational practices can be expected to affect journalistic culture as much as journalistic work. Schudson draws on Sahlins to argue that cultural aspects of society, "while they may be uncovered by detailed historical analysis, cannot be extrapolated from features of social organization at the moment of study. They are part of culture— a given symbolic system within which and in relation to which reporters and officials go about their duties" (Schudson 2005, 187). There is much debate about the role played by culture in the operation of journalism. Benson, referring to forms of culture grounded in national differences, argues that the term should be abandoned. It must be admitted that there is a slight lack of clarity in Schduson's discussion of the journalistic culture; he does occasionally refer to national cultures and the role they play in the construction of news— national differences that, as Benson argues, may be more usefully be seen as the outcome of field dynamics rather than symbolic systems per se. Nevertheless, Schudson's arguments about culture and journalism are far subtler and far ranging than Benson allows. From the perspective of praxis, culture might be seen as the accrued, long-term detritus of daily newswork, a detritus that creates the symbolic background against which journalists guide their actions. Or we may see it as the background symbolic system that guides many of the most important and obvious but least understood aspects of journalistic practice. The culture of journalism is invoked by Schudson to guide researchers toward what I would call "interesting but tricky problems": "journalist's vague renderings of how they know news when they see it ... [why] news stories are so often personified... [why] reporters write of persons and not structures, of individuals not social forces ... and [journalistic] assumptions about narrative, storytelling,



Electronic copy available at: https://ssrn.com/abstract=2009292

human interest, and the conventions of photographic and linguistic presentation." (188-89). Obviously, all these questions can be reoriented to shed light on the nuances of data journalism. Does using big data create a difference sense of what news is? Does data encourage a greater emphasis on structures and social forces as opposed to personalities and incidents? The analysis of organizational routines might begin to answer these questions, but newsroom sociology cannot definitively answer them. For Schudson, Zelizer (Zelizer 1992) Carey (Carey 2008), and others, only a cultural analysis informed by nuanced historical research can do the heavy lifting required to answer these questions.

To date, most of the detailed historical research on the relationship between journalism's symbolic systems, computer programming, and big data remains to be done. While he might not call his research cultural history, Powers work on the evolution of journalistic "computational talk" from 1975 to the present day (Powers, forthcoming) is one possible exception; for Powers, discussions of computers and journalism usually frame these devices as providing continuity of work routines, threats to journalism, and opportunities for professional reform. Powers lengthy analytical time frame and his focus on value, technology, and work thus marks one model to be emulated by computational journalism scholarship. Research that embraces an even longer time-frame, such as Barnhurst and Nerone's analysis of the "forms of news" (Barnhurst and Nerone 2001) or my own work on "news objects" (AUTHOR, forthcoming) can also be seen as providing openings for the future analysis of the impact of big data on journalistic sense-making.

## 3.6 Technology and the News

In this paper, up until this point, "technology" has been discussed as a form of embedded materiality that intersects with journalism via the mediation of cultural, political economic, organizational, or institutional factors. To conclude, I want to advance the notion of technology as an independent lens of analysis without advocating, I hope, a pernicious form of technological determinism. For most sociology, an over-emphasis on the role played technology in the construction of news constitutes the primary sin, one to be assiduously avoided. But is there a way to talk about technology and the news on its own terms, without reducing said technology to either a political,



Electronic copy available at: https://ssrn.com/abstract=2009292

economic, cultural, or social construction? I would argue that, to a limited degree, there is.

Some key aspects of computational journalism are themselves partly technological in nature; they are in part determined by the vast increase in computer processing speed and a growing data ubiquity, as seen over the last thirty years. These technological developments are not, in and of themselves the primary determinants of the shape computational journalism has taken in the past or may take in the future. I hope that each of the five lenses discussed so far has properly inoculated readers against visions of technologically guided progress; the meaning and impact of computational journalism is as much affected by politics, economics, institutions, culture, and organizational routines as it is by microchips and Moore's Law. Nevertheless, while not determinative, a scholarly concern with actual role played by materiality and technology in the processes of journalism might emphasize at least three areas: the oft-hidden intersection between *imagined values* and *engineering design* during the construction of journalistic artifacts, the increasingly *hybrid* nature of newsroom sorting and filtering technologies, and the changing status of *journalistic evidence* fostered by the exponential increase in available evidentiary forms in the digital age. To date, little serious work has been done in any of these areas, though extensive examples from field of science and technology studies provide examples of what such a research agenda might look like if applied to journalism.

In a technologically oriented study of computational journalism, researchers might first wish to probe the manner in which journalistic values refract through the material lens of digital design. Here, Nissenbaum's work on "values in design" (Nissenbaum 2004) might serve as a potential guide. In designing semantic tagging systems to assist Google and other search engines in the proper sorting of news content, how do engineers filter newsroom notions such as authorship, the byline, and the fact though taxonomical systems originally designed for use in computer science? What aspects of the traditional story, in short, appear to be the most valued in the artifact building process, and what aspects are neglected? This focus on the manner by which human needs are embedded within and yet refracted through technological artifacts could draw researchers attention to a second aspect of computational journalism: the hybrid nature of newsroom sense-making technologies. Data-



Electronic copy available at: https://ssrn.com/abstract=2009292

crunching algorithms and other increasingly invisible information ordering devices are neither entirely material, nor are they entirely human— they are hybrid, composed of both human intentionality and material obduracy. In this way, the human becomes partially obdurate and the material partly intentional (Latour 1993). Third and finally, the analysis of computational journalism through a technological lens would focus on the manner in which the explosion of digital "objects" - Tweets, links, blog posts, databases, etc— changes journalists evidentiary calculations about what counts as proper story evidence. While none of these three angles amount to a methodological research prescription per se, a technological approach to computational journalism would be as much about asking particular types of research questions and emphasizing different areas of scholarly interest than the previously discussed lenses.

## 4. Conclusion

Each of the six lenses discussed in this article— political, economic, field, organizational, cultural, and technological— are grounded in decades of serious journalism scholarship and yet can also be adapted to serve the needs of an academic agenda that focuses on a new research object: computational journalism. Analyzing computational journalism through a political lens would focus on public policies that encourage or discourage the adoption of computational processes and practices. An economic perspective would critically examine the stratification of the computational journalism field, and the manner by which differences in institutional resources create patterns of asymmetry in journalism's ability to operate on behalf of the public. A field approach would expand this economic perspective, broadening the analytical lens to include other institutions in other fields— the field of computer programming and the foundation field, just to name two— and researching how particular field homologies create and recreate different distributions of financial, cultural, and symbolic power inside journalism. The organizational perspective would ground this research in real, daily work practices and routines, analyzing the manner in which potential technological advances succeed, failed, or were modified due to institutional reticence, bureaucratic red tape, and institutional priorities. A cultural lens on computational journalism



Electronic copy available at: https://ssrn.com/abstract=2009292

would not only examine the way human and material interactions affected the adoption of computational techniques, but also the way these interactions took place against the backdrop of larger symbolic systems and patterns of belief. Finally, a technological approach to studying the intersection of journalism with big data could discuss technology on its own terms, examining values in journalistic design, the hybrid nature of newsroom sense-making, and the changing status of journalistic evidence fostered by the exponential increase in types of digital evidence.

Most of this paper might be seen as the scholarly equivalent of the blind men and the elephant; each of the six analytical lenses might, in the aggregate, give us some useful insight on the shaggy, emerging beast I have called computational journalism. More realistically, however, these analytical techniques would be combined. There is tremendous overlap between political, institutional, and economic perspectives. Any one of the five other perspectives might help to add some three-dimensional context to often routine driven explanations advanced by the organizational approach. A focus on culture and history can counteract the reductionist, possibly determinist tendencies of the technological perspective. And so on. In the end, the research lens or lenses chosen should depend primarily on the questions a scholar hopes to ask and the problems he or she wishes to address.

I conclude on this methodologically catholic note because the purpose of this paper is not to advance one of the six analytical frameworks over others. Rather, my hope has been to advocate on behalf of what I have called a *critical approach to computational journalism*. By this I mean a research program that brackets (at least temporarily) questions of practical utility and the "future of journalism"—questions that currently dominate the agenda of much newsroom research. There is nothing wrong with being helpful, of course. The best scholarship would certainly keep one foot in current journalistic developments, if only to ask better questions and postulate better hypotheses. But research into emerging forms news production needs to keep in mind that all change comes with tradeoffs, that valuing certain technological advances might shortchange others, and that positive journalistic developments may ultimately find it difficult to escape the networks of power in which they are, in the last instance, embedded.



Electronic copy available at: https://ssrn.com/abstract=2009292

## 5. References


Baker, CE (2002) *Media, Markets, and Democracy*, Cambridge, UK: Cambridge University Press

Barnhurst, KG and Nerone JC (2001) *The Form of News: A History*. New York: The Guilford Press

Benson R (1999) Field theory in comparative context: A new paradigm for media studies. *Theory and Society* 28(3): 463-498

Benson R (2004) Bringing the sociology of media back in. *Political Communication* 21(3): 275-292.

Benson R (2010) What makes for a critical press? A case study of French and US immigration news coverage. *The International Journal of Press/Politics* 15(1): 3

Berry DM (2011) The computational turn: Thinking about the digital humanities. *Culture Machine* 12(0): np.

Boczkowski PJ (2004) The processes of adopting multimedia and interactivity in three online newsrooms. *Journal of Communication* 54(2): 197-213

Carey JW (2000) 'Journalism and Democracy Are Names for the Same Thing.' Available at: http://www.nieman.harvard.edu/reports/article/101943/Journalism-and-Democracy-Are-Names-for-the-Same-Thing.aspx.

Carey, JW (2008) *Communication As Culture: Essays on Media and Society*. New York: Taylor & Francis

Medill.northwestern.edu (200x) Coders wanted: Journalism scholarships available for programmers / developers. Available at http://www.medill.northwestern.edu/admissions/page.aspx?id=58645

Cohen P (2010) In 500 billion words, new window on culture. *New York Times*, December 17, 2010

Cohen S, C Li, J Yang, and C Yu (2011) Computational journalism: A call to arms to database researchers. In *Proceedings of the 5th Biennial Conference on Innovative Data Systems Research*. Sailorman, CA: ACM.

Dalbello M (2011) A genealogy of digital humanities. *Journal of Documentation* 67(3): 480-506

Daniel A, T Flew, and C Spurgeon (2010) The promise of computational journalism. In *Proceedings of the 2010 Australian and New Zealand Communication Association*. Canberra: ANZCA




Electronic copy available at: https://ssrn.com/abstract=2009292

Diakopoulos N, F Kivran-Swaine, and M Naaman (2011) Playable data: Characterizing the design space of game-y infographics. In *Conference on Human Factors in Computing Systems (CHI)*. May, 2011

Diakopoulos N, M Naaman, and F Kivran-Swaine (2010) Diamonds in the rough: Social media visual analytics for journalistic inquiry. In *Conference on Visual Analytics Science and Technology (VAST)*. October, 2010

Diakopoulos N and Naaman M (2011) Towards quality discourse in online news comments. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*.

Domingo D and Patterson C (2011) *Making Online News (Volume 2): Newsroom Ethnography in the Second Decade of Internet Journalism*. New York: Peter Lang

Dossick CS and Neff G (forthcoming) Messy talk and clean technology: Communication, problem-solving and collaboration using building information modeling *The Engineering Project Organization Journal*

Epstein EJ (2000) *News From Nowhere: Television and the News*. Lanham, MD: Ivan R. Dee

Gans HJ (1979) Deciding what's news: A study of CBS evening news, NBC nightly news. Newsweek, and Time. New York: Pantheon

Hallin D and Mancini P (2004) *Comparing Media Systems: Three Models of Media and Politics*. Cambridge, UK: Cambridge University Press

Hanitzsch T (2008) Comparing journalism across cultural boundaries.: State of the art, strategies, problems, and solutions. In: Weaver D and Loffelholz M (eds). *Global Journalism Research: Theories, Methods, Findings, Future*. London: Blackwell

Hayles NK (2011) How we think: The transforming power of digital technologies. In Berry, DM (ed) *Understanding the Digital Humanities*. London: Palgrave

Irfan E and Diakopoulos N (2007) Computational journalism: CS 4803CJ/8903CJ. Available at: http://www.cc.gatech.edu/classes/AY2007/cs4803cj_spring

Journalism.Columbia.com (2010) Journalism-computer science joint degree. Available at: http://www.journalism.columbia.edu/page/275-dual-degree-programs/278#computer.

Klinenberg E (2005) Convergence: news production in a digital age. *The ANNALS of the American Academy of Political and Social Science*, 597(1): 48

Kreiss D and Howard P (2010) New challenges to political privacy: Lessons from the first U.S. Presidential race in the web 2.0 era. *International Journal of Communication* 4 (2010)

Latour B (1993) *We Have Never Been Modern*. Harvard: Harvard University Press

Latour B (2004) Why has critique run out of steam? *Critical Inquiry* 30 (2): 225-248



Electronic copy available at: https://ssrn.com/abstract=2009292

Lewis SC (2011) Journalism innovation and the ethic of participation: A case study of the knight foundation and its news challenge. Unpublished doctoral dissertation, University of Texas-Austin

Manovich L (2011) Trending: The promises and the challenges of big social data. Available at http://manovich.net/2011/04/23/new-article-trending-the-promises-and-the-challenges-of-big-social-data

McChesney RW (1995) *Telecommunications, Mass Media, and Democracy: The Battle for the Control of US Broadcasting, 1928-1935* Oxford: Oxford University Press.

McPherson T (2009) Introduction: Media studies and the digital humanities. *Cinema Journal* 48(2): 119-123

Michel J , YK Shen, AP Aiden, A Veres, MK Gray, JP Pickett, and Hoiberg D (2011) Quantitative analysis of culture using millions of digitized books. *Science* 331(6014): 176-82

Ngyuen D (2010) Scraping for journalism: A guide for collecting data. Available at:

http://www.propublica.org/nerds/item/doc-dollars-guides-collecting-the-data.

Nissenbaum H (2004) Privacy as contextual integrity. *Washington Law Review*. 79: 119-158

Noveck B (2009) *Wiki Government: How Technology Can Make Government Better, Democracy Stronger, and Citizens More Powerful*. Washington DC: Brookings Institute Press

Noveck B (2011) The single point of failure. Innovating Government: *Information Technology and Law Series 20 (2): 77-99*

Nussbaum E (2009) The new journalism: goosing the gray lady. *New York Magazine,* January 11, 2009.

Powers M (forthcoming)

Rosenbloom PS (2010) Towards a conceptual framework for the digital humanities. Available at http://pollux.usc.edu/~rosenblo/Pubs/TCFDH.pdf

Rothberg M (2010) Quantifying culture? A response to Eric Slauter. *American Literary History* 22 (2): 341

Royal C (2010) The journalist as programmer: A case study of *The New York Times* interactive news technology department. Presented at *The International Symposium For Online Journalism*. Austin, TX

Ryfe DM (2006) Guest editor's introduction: New institutionalism and the news. *Political Communication* 23(2): 135-144

Schudson M (1997) The sociology of news production. In Berkowitz DA (ed) *Social Meanings of News: A Text Reader*. London: Sage, pg 7-20.

Schudson M (2010) Political observatories, databases & news in the emerging ecology of public information. *Dædalus* 139(2): 100-109.

Schudson, Michael. 2005. Four approaches to the sociology of news. In Curran J and Gurevitch M (eds.) *Mass Media and Society, (4th ed)* London: Hodder Arnold



Electronic copy available at: https://ssrn.com/abstract=2009292

Segel E and Heer J (2010) Narrative visualization: Telling stories with data. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)* 16 (6): 1139-48.

Siebert, FS, T Peterson, and Schramm W (1956) *Four Theories of the Press.* Champaign, IL: University of Illinois Press.

Slauter E (2010) Revolutions in the meaning and study of politics. *American Literary History* 22(2): **325-**

Stoeffel K (2011) Huffington Post hires Andrei Scheinkman to lead interactive news features team. Available at http://www.observer.com/2011/media/huffington-post-hires-andrei-scheinkman-lead-interactive-news-features-team

Taylor, Megan (2009). Washington Post's 'web ninjas' build map-timeline combo. Available at http://www.pbs.org/mediashift/2009/02/washington-posts-web-ninjas-build-map-timeline-combo047.html.

Tuchman, G. 1978. *Making News.* New York, NY: Free Press.

Turner F and Hamilton JT (2009) *Accountability Through Algorithm: Developing the Field of Computational Journalism.* Available at http://dewitt.sanford.duke.edu/images/uploads/About_3_Research_B_cj_1_finalreport.pdf

Waisbord SR (2000) *Watchdog Journalism in South America: News, Accountability, and Democracy.* New York, NY: Columbia University Press

Waite M (2009) The key lesson I learned building PolitiFact: Demos, not memos available at http://www.mattwaite.com/posts/2009/apr/27/key-lesson-i-learned-building-politifact-demos-not/.

Waldman S (2011) *The Information Needs of Communities: The Changing Media Landscape in a Broadband Age.* Washington DC: Federal Communications Commission

Yong E (2011) The Renaissance man: how to become a scientist over and over again. Available at: http://blogs.discovermagazine.com/notrocketscience/2011/06/08/the-renaissance-man-how-to-become-a-scientist-over-and-over-again/.

Zelizer B (1992) *Covering the Body: The Kennedy Assassination, the Media, and the Shaping of Collective Memory.* Chicago: University of Chicago Press.



Electronic copy available at: https://ssrn.com/abstract=2009292