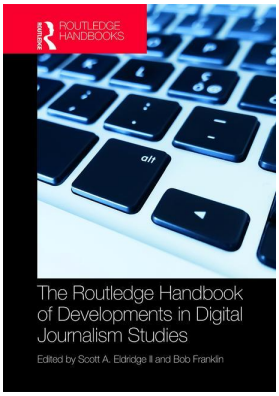

The Routledge Handbook of Developments in Digital Journalism Studies

Scott A. Eldridge, Bob Franklin

Defining and Mapping Data Journalism and Computational Journalism

Publication details
https://www.routledgehandbooks.com/doi/10.4324/9781315270449-18
Mark Coddington
Published online on: 30 Aug 2018

17

DEFINING AND MAPPING DATA JOURNALISM AND COMPUTATIONAL JOURNALISM

A review of typologies and themes

Mark Coddington

In every era, there is a subset of journalistic practice that draws substantial attention (perhaps inordinately so) from journalists and especially scholars of journalism, outpacing the prevalence of the practice itself or its interest among the public. During the 1990s, the discussion of public journalism ran throughout both journalism scholarship and the news industry itself, while its practice was largely limited to a few well-publicized examples. In the 2000s, citizen journalism became an industry buzzword and the subject of much scholarly rumination, though successful instances of citizen-led news operations and projects were difficult to come by. During this decade, data journalism and its many related forms has become an object of scholarly fixation and much excitement in the industry and in journalism education. Numerous news organizations developed specialized data teams, global data journalism awards were instituted, and attendance at the annual conference of the seminal U.S. group National Institute for Computer-Assisted Reporting (NICAR) tripled between 2010 and 2014 (Stiles, 2017). In academia, universities around the world developed data journalism courses to meet the industry's demand for those skills (Griffin and Dunwoody, 2016; Hewett, 2016), and scholarship on data journalism boomed.

Nevertheless, data journalism remains only a small subset of journalism, something that exists only in precarious forms, if at all, in many smaller newspapers and news organizations (Fink and Anderson, 2015). Likewise, scholarship on data journalism remains only a niche within journalism studies, published by a couple of dozen scholars worldwide, largely in a handful of journalism-centric journals. But that small constellation of scholars has produced a flurry of research over the past decade, often overlapping and occasionally contradicting each other in its rush to document and understand a rapidly emerging phenomenon.

I offered a typology for understanding data journalism, computational journalism, computer-assisted reporting, and other forms of quantitative journalism in a 2015 journal article I optimistically titled "Clarifying Journalism's Quantitative Turn" (Coddington, 2015). Little clarity has emerged since then, with academic studies continuing to churn out, and many of them either developing another novel way to define and organize forms of quantitative journalism or largely side-stepping the issue of definition and classification. But I hope to continue moving toward clarity in this chapter by examining some of those typologies (both complementary and competing) and outlining themes that have begun to congeal in scholars' characterizations of

data journalism, computational journalism, and other similar practices while acknowledging the breadth of scholarship and perspectives on this issue.

First, though, it is important to clarify why such an endeavor is important in the first place. Why devote so much attention to the classification and definition of various slightly differentiated specialized forms of journalism? Why not simply focus on the practices themselves, without fixating on what they should be called or how they should be grouped? This is a common sentiment among journalists themselves – including, famously, data journalism pioneer Adrian Holovaty (2009) – but one that ultimately should be guarded against, at least among scholars. I echo the arguments made by Nikki Usher (2016: 73–74; see also Usher, this volume, Chapter 26) in defense of such typological work by noting that as buzzwords within the industry, terms like data journalism and computational journalism are in danger of being discarded as meaninglessly trendy (and later, passé) without careful reflection that can attach substantial meaning to them. This type of categorization also reminds us that these terms and practices have institutional, cultural, and epistemological roots; they did not emerge out of nowhere, nor are they merely new names slapped onto an existing practice like computer-assisted reporting. As Usher argues, this work "helps us understand what is novel, old, and the same about these groups and as they compare to traditional journalism" (2016: 74). There may come a time when further work on the classification of these concepts becomes a pedantic and picayune exercise without real value, but the subfield is still too jumbled to have reached that point yet.

The state of data journalism scholarship

In the introduction to a special issue on data journalism edited for *Digital Journalism* that originally appeared online in 2014, Seth Lewis described the scholarship on quantitative journalism as a "rapidly growing body of work" that is "seemingly 'everywhere' based on the industry buzz and accelerating scholarly interest" (Lewis, 2015: 322–323). Lewis was right about acceleration; research into data journalism has only increased since then. Typical of early research into a phenomenon, much of that data journalism research has been exploratory and descriptive (Ausserhofer et al., 2017). Early on in the current wave of scholarship, C. W. Anderson voiced concern that research on computational journalism was largely internalist, with "a tendency to consider the problems of journalism scholarship from the point of view of the journalism profession" (2013: 1007). Such early internalist work has remained at the forefront of scholarship, picking up numerous citations, particularly in computational journalism (Ausserhofer et al., 2017; Young and Hermida, 2015). But research since then has eased off this more utilitarian perspective, with several studies taking a more conceptual or critical tack on journalism practice (Borges-Rey, 2017, and this volume, Chapter 21; Bucher, 2017; De Maeyer et al., 2015; Hammond, 2017). Some of the most useful work conceptually has centered on the epistemological approaches of various forms of quantitative journalism, using them to trace shifts in the philosophical frameworks and processes of knowledge production among journalism more broadly (Anderson, forthcoming; Borges-Rey, 2016, 2017; Parasie, 2015; Parasie and Dagiral, 2013).

Throughout the development of scholarship in this field, researchers (including myself) have complained of the proliferation of terms used to describe quantitative forms of journalism, yet the litany of labels continues to grow. In addition to the central terms of data journalism (and data-driven journalism), computational journalism, and the much older computer-assisted reporting (CAR), researchers have described these and related phenomena as "computational exploration in journalism" (Gynnild, 2014), "interactive journalism" (Usher, 2016), "big data journalism" (Tandoc and Oh, 2017), and others. In the last few years, automated forms of journalism have also arisen as a key set of journalistic practices adjacent to these, and those have been given various

names like 'automated journalism', 'algorithmic journalism', and 'robot journalism', as well (Dörr, 2016; Splendore, 2016).

As others have noted (Bucher, 2017; Hammond, 2017), the fact that scholars have hurried to come up with so many different terms for similar journalistic phenomena is notable in itself, an indication of the rapid change in the journalistic field and haziness in the scholarly thinking about it – and perhaps the need to distinguish oneself in a crowded academic marketplace as well. Though the maze of terms and concepts is indeed a sign of confusion and a bit of franticness in scholarship in this area, that's not necessarily a cause for deep concern (yet). This is a natural stage when scholars are encountering widespread new phenomena that are practiced differently in a variety of contexts and that change quickly. In part, the confusion stems from the fact that scholars are torn between borrowing terms from the profession and developing their own. The terms from the profession (such as data journalism) are far more widely used and therefore seem more 'natural', but since journalists themselves do not agree on what those terms mean, they bring a variety of sometimes conflicting meanings into the academic context. But when scholars try to counter this ambiguity by developing their own terms, they seem unnecessary and out of step with the way practices are being talked about on the ground. When the actors being studied have themselves developed widely used but imprecise terms about the work that is being studied, scholars are left with the unenviable task of either imposing clarity on those terms or supplanting them in the academic discourse.

And yet it is a task that must be undertaken. As data journalism scholarship transitions out of its initial exploratory phase into deeper and more conceptually rich research, it also needs to begin pinning down clear, stable definitions and conceptualizations of the practices it is studying. This is a necessary stage in the maturation of scholarship in this area, and it is one that researchers should be ready by now to enter.

Typologies of data journalism practices

A few scholars have attempted to move this clarifying work forward with typologies organizing and distinguishing between various types of quantitative journalism. Astrid Gynnild's (2014) was the first significant attempt to formally parse multiple forms of this type of work, as she gave the umbrella term "computational exploration in journalism" to work that encompassed use of algorithms and social science methods to retrieve, analyze, and visualize data. She distinguished between three approaches to computational exploration in journalism: a newsroom approach built on journalistic traditions that included computer-assisted reporting (CAR) and data journalism; an entrepreneurial approach that involves creating databases for web or mobile apps and included journalism as programming; and an academic approach that consisted predominantly of computational journalism. Gynnild's typology is notable for including data journalism and CAR within the same category, and it relies heavily on the journalistic setting in which the practice takes place. This places a valuable emphasis on professional values and influences, though it leaves aside important epistemological elements that others have fruitfully used to distinguish between practices. In some cases, the setting on which this typology relies can be fungible; a key aspect Gynnild uses to characterize the entrepreneurial approach is that it often takes place outside traditional newsrooms or on the margins of the profession. This can change quickly and lead to blurring: if a data journalist leaves the *New York Times* to do similar work at a news start-up, does she switch from the newsroom approach to the entrepreneurial approach by virtue of her new place of employment?

My own typology (Coddington, 2015) outlines three forms of quantitative journalism: data journalism, computational journalism, and CAR. It then introduces four dimensions with which

to evaluate them: a closed and professional vs. open and networked orientation, transparency vs. opacity, epistemological orientations toward targeted sampling vs. big data, and an active vs. passive vision of the public. This typology introduces both epistemological and professional elements, as well as audience-related factors, to distinguish types of quantitative journalism. (In contrast to Gynnild, I place CAR as a distinct practice from data journalism, which I classify as closer to computational journalism.) But though its four factors can be used on any journalistic case, it is not directly grounded in empirical research. Its characterizations of data journalism, computational journalism, and CAR are Weberian ideal types based on a review of literature and discourse rather than arising out of empirical data gathered from journalistic practice itself.

At least two other scholars have attempted to expand on this typology: Splendore (2016) adds a three-part comparative dimension to the four originally laid out, looking at epistemological implications in the access, process, and editing stages of news production. In the access stage, he differentiates between manual and automated access to data; in the process stage, between hypothesis-based investigation and data-driven analysis; and in the editing stage, between design for a broader public or for individual personalization. Splendore analyzes data journalism, computational journalism, and automated journalism through these new dimensions (he drops CAR from his analysis), finding data journalism to be largely distinct from the more similar computational and algorithmic journalism. Splendore's expansion to the typology is a useful process-based addition, though like my original typology, it consists of ideal types of practices and thus requires generalizations that may not match up with actual observed practice.

Michalski (2016) gives the typology a more empirical application with the development of a content analysis scheme based largely off the four dimensions in the original typology (Coddington, 2015) as well as elements drawn from numerous other studies. He uses this operationalization to compare projects by the *Guardian* and the *Washington Post*, placing the *Guardian* project closer to data journalism and the *Post* project near CAR. Michalski's 32-item coding scheme could use some refinement in several spots, but he does valuable work in bridging the conceptual framework with the kind of quantitative empirical analysis of journalistic output that has been largely absent from work in this area (Ausserhofer et al., 2017). Both Splendore and Michalski's studies are useful extensions of the typology that expand its conceptual range and empirical applicability, though both hold the same limitation as the original typology: it is a helpful way of seeing dimensions by which these forms of journalism might be evaluated, though it only does initial work in actually evaluating and distinguishing the journalistic forms themselves.

In her book on interactive journalism, Nikki Usher (2016) introduces a typology that diverges from mine and Gynnild's (2014) in several respects. First, Usher's is not at root a typology of journalistic practices but a typology of the people who engage in them. Second, Usher's area of concern only partly overlaps with the quantitative forms examined here; she is primarily focused on classifying the types of people who work to produce news interactives, some but not all of whom work with data or computation. Third and most helpfully, Usher's typology is built on field-based evidence, arising in a grounded manner from the way journalists themselves talk about who they are. She distinguishes between hacker journalists, programmer journalists, interactive journalists, and data journalists, finding that data journalists have the closest connection of those groups to traditional journalism, through CAR. Her categories are not mutually exclusive, though; she describes many hacker and programmer journalists as doing data journalism, but notes that some data journalists aren't programmers and don't code at all. Data journalists, she says, are defined by their orientation toward data: "they are primarily working with data in the service of stories, actively trying to tell stories with data, and spend most of their time working specifically with data" (2016: 90). In addition to her bottom-up generation from journalistic practice, Usher's study is also valuable for its classification of all of these forms as a *professional*

subspecialty, which absorbs many of the norms and values of professional journalism but offers a distinct knowledge claim to help expand the profession's jurisdiction and reassert its relevance. Though the professional subspecialty she describes is interactive journalism rather than data journalism, many of its characteristics transfer over smoothly and form an apt theoretical framework for thinking about data journalism as a set of practices.

Borges-Rey (2017) synthesizes the types laid out by myself and Gynnild (2014) as well as Parasie's research (Parasie, 2015; Parasie and Dagiral, 2013) into two basic epistemological paradigms: the "newshound", which is subordinate to journalism's professional norms and traditional ways of dealing with data, and the "techie", which switches between journalistic and computational mindsets in approaching data. Borges-Rey places CAR in the newshound paradigm and computational journalism in the techie paradigm, though data journalism is split: as Gynnild (2014) conceives of it, it is in the newshound approach, and as I conceive of it, it is in the techie approach. Borges-Rey's simple typology is useful as an overarching framework, and it addresses both the professional setting and epistemology together. It doesn't add to our ability to make fine-grained distinctions between forms (for example, the distinction between data journalism and computational journalism), but it does helpfully compress many of the themes Gynnild and I identify into a single unified spectrum.

Though it is perhaps a bit excessive to have at least four different typologies of a subfield emerge within three years, these frameworks are more complementary than competitive. Each serves a different purpose and best applies to different conceptual tasks. Borges-Rey's (2017) is most valuable as a general orienting device to gauge the overall position of a journalistic practice, while mine (Coddington, 2015) offers a toolkit to make more specific distinctions between various aspects of different practices. Usher's (2016) is most helpful for analyzing the roles, backgrounds, and orientations of the individual actors who are engaging in those practices, especially given the diverse roles and approaches of the people who work in this area, even within the same project. Scholars shouldn't feel obligated to use all of them – that would probably be conceptually counterproductive – but any of them might prove the most useful tool for the analytical job at hand.

Characterizing data journalism and computational journalism

Having examined the broader typologies by which these forms are organized, it's important also to look more closely at the forms themselves and the emerging scholarly consensus about their chief characteristics, alongside the disagreements about what values and practices they constitute. For both data journalism and computational journalism, some general themes have arisen that can help to broadly characterize both forms, even as they continue to shift and our knowledge of them expands.

Data journalism

The degree to which data journalism is tied to previous forms of journalism such as CAR is a key point of contention among scholars, but the story of how it rose to its newfound place of prominence has begun to solidify. Programmer-journalist Adrian Holovaty is often seen among both journalists and scholars as its most influential early innovator, with his 2006 blog post outlining a vision of journalism as structured data (Holovaty, 2006) and his late 2000s projects fleshing out that vision, ChicagoCrime and EveryBlock (Anderson, forthcoming; Ausserhofer et al., 2017; Gray et al., 2012). The *Guardian* also plays a crucial role in accounts of data journalism's rise, with its 2009 crowdsourced project analyzing UK MPs' expenses and 2010 analysis and presentation

of WikiLeaks' large-scale document leaks as key projects bringing data journalism into the public consciousness. Its Datablog, initially edited by Simon Rogers, extended that work in highlighting data-driven journalistic work on a regular basis (Gray et al., 2012; Howard, 2014; Knight, 2015). Soon afterward, others have identified 2013 and 2014 as a time when discourse about data journalism boomed, and the term began serving as a major form of self-identification for journalists (Howard, 2014; Royal and Blasingame, 2015).

Scholars have had a difficult time arriving at an agreed-upon definition for data journalism, which reflects the amorphous nature of the practice itself. Howard's (2014) is probably the most robust and straightforward, but even this is presented in a few parts: He defines data journalism as, "gathering, cleaning, organizing, analyzing, visualizing, and publishing data to support the creation of acts of journalism" and also as the application of data science, or the extraction of knowledge from data, to journalism, as well the combination of treating data as a source, applying statistics to interrogate it, and using visualizations to present it (ibid.: 4–5). As Ausserhofer and colleagues (2017) and Royal and Blasingame (2015) both note, scholars have differed on whether data journalism is fundamentally about a process or product, and I echo their assertion that it is both, though I think the process – an epistemological and professional approach to gathering, analyzing, and presenting data – is more fundamental to data journalism than any product.

The best way to characterize data journalism may not be through a precise definition but through a set of themes that emerge across numerous scholarly analyses of it. The most central theme in definitions of data journalism is *storytelling* – the idea that data journalism, as Howard (2014) puts it, "is telling stories with numbers, or finding stories in them" (5). In Royal and Blasingame's (2015) review of 63 definitions of data journalism by academics and professionals, storytelling emerges as the most common theme and in some cases is the sole or primary element of the definition (Mair and Keeble, 2013; Splendore, 2016). Storytelling is a particularly focal point in data journalists' understandings of their own work, where they see telling stories as the end and data simply as a means (Borges-Rey, 2016; Rogers, 2013). While this characterization of data journalism as storytelling serves to link data journalism with the traditional journalistic paradigm, other scholars have identified storytelling as a primary point of divergence between data journalism and CAR. Whereas in CAR data primarily served an investigation, in data journalism, data's main end is as a story, potentially on its own and often through interactivity or personalization (Felle, 2016; Usher, 2016).

Two other themes of data journalism are related – *transparency* and *visualization*. Rather than allowing data to recede into the background of a story, data journalists tend to advocate transparency that allows audiences to access the data on which the story is built, because of both the influence of open-source philosophy and the increased demand for unfiltered information online more generally (Coddington, 2015). This openness often involves publishing the full data online as part of the project, though this is not always an element of data journalism, and the published data has always been processed and formatted, in ways that are not always described to the audience (Lesage and Hackett, 2014; Usher, 2016). Transparency is also a key form of legitimation for data journalists, providing some accountability for their work, though most audience members do not have the skills to perform that kind of function (Borges-Rey, 2016; Lesage and Hackett, 2014). In many cases, the transparency work of the publication of data is done through a visualization that may be interactive. In this way, visualization – a product that is a key but not necessary element of data journalism – is tied not only to the idea of data as a story in itself but also to transparency, by allowing audience members to drill into datasets and find elements relevant to them (Gray et al., 2012; Tandoc and Oh, 2017). Not all data journalism involves visualization, nor does it all involve interactivity. But the emphasis on using visualization and interactives to aid transparency and allow audiences to act as coproducers of the truth claims of the journalistic

artifact is an important definitional characterization of data journalism as a whole (Splendore, 2016; Usher, 2016).

Crowdsourcing has also occasionally been a part of this transparency process of data journalism, particularly in high-profile examples such as the *Guardian*'s MPs' expenses project or ProPublica's "Free the Files", in which audiences analyzed massive troves of documents to help journalists discern trends. As journalists have characterized data journalism, they have sometimes included crowdsourcing as a key component (Gray et al., 2012; Rogers, 2013). But it's simply not widespread enough in practice to be classified as a central theme for data journalism. Especially in more recent studies, scholars have found that while data journalists are receptive to crowdsourcing and see it as an important part of data journalism work, few have actually successfully or consistently employed it (Borges-Rey, 2017; Felle, 2016). Crowdsourcing is an extension of data journalism's vision of an active public and transparent data analysis process, but only in rare cases is it a significant part of actual data journalism practice.

Another element emphasized by some researchers but not fundamental to data journalism is the size of datasets used. Ausserhofer and colleagues (2017) identify large datasets as a common part of definitions of data journalism, and Uskali and Kuutti (2015) define data journalism as exclusively working on large datasets. Indeed, data journalism tends to work with larger datasets than CAR has, which entails an epistemological shift from sampling to extrapolate from smaller datasets to mining and processing larger ones (Coddington, 2015). But while the number and size of datasets freely available served as a catalyst for data journalism's growth, the size of data is not an inherent part of data journalism's approach. Several studies have found that journalists see a sharp distinction between the type of time-intensive, team-based, investigative data journalism that has attracted much of the scholarly and professional attention and the more quick-turnaround, light, routine data journalism that happens on an everyday basis (Borges-Rey, 2016; De Maeyer et al., 2015; Rogers, 2013). Uskali and Kuutti (2015) have given the most thorough characterization of these two types (they call them "investigative data journalism" and "general data journalism"). Although Uskali and Kuutti center their definition of data journalism on large datasets, their conceptualization shows that a significant part of data journalism is smaller-scale work, a finding confirmed by other researchers (Borges-Rey, 2016; De Maeyer et al., 2015; Fink and Anderson, 2015).

The greatest disagreement among scholars regarding data journalism is over the degree to which it represents a pattern of continuity with previous forms of quantitative journalism (particularly CAR) vs. substantial change. As Borges-Rey (2017) notes, this gets to the heart of the question of what data journalism is: is it investigative journalism reborn with new tools at its disposal? Or is it a new combination of the logics of computer science and journalism that reshapes news production? Data journalists themselves tend to affirm the former, emphasizing their connections to the normative aims of traditional investigative journalism and foregrounding the importance of the journalistic mission (Felle, 2016; Rogers, 2013; Royal and Blasingame, 2015). And epistemologically, data journalism is indeed built on the largely positivist traits of objectivity that have governed journalism for decades, with its emphasis on ascertaining reality through verified observation in the form of data (Lesage and Hackett, 2014). Data journalism clearly has roots in CAR, whose key figures and institutions – notably the National Institute for Computer-Assisted Reporting, or NICAR – continue to play a pivotal role in structuring and developing data journalism as a subfield.

But some of data journalism's defining traits also represent departures from previous forms of quantitative journalism. Several of them have been outlined here: data journalism's emphasis on data itself as story, its embrace of transparency as a core norm, its focus on interactivity and personalization, and its orientation toward mining large datasets rather than sampling small ones are

all to varying degrees divergences from the paradigm of CAR. As Hammond (2017) notes, data journalists have professional legitimacy to gain by emphasizing their continuity with traditional journalism, which can lead us to overestimate that continuity based on their discourse. Anderson (2015) describes the continuity-based perspective as "deeply and fundamentally flawed" (351) and says quantitative journalism's story in the United States is one of rupture. But he argues that the fundamental changes that formed the foundation for data journalism took place not within the past decade or two but during the 1960s and 1970s, with the development of processing speed through computers, prestige for investigative journalism, a focus on patterns rather than incidents, and a move beyond he-said, she-said journalism (Anderson, forthcoming). The answer to the question of data journalism's continuity with previous quantitative forms, then, may depend largely on how far we zoom in or out. When we look closely at data journalism, we see important fissures between it and CAR. But when we zoom out, we see that both are built on the same substrate of positivistic values and methods fused with high-modern public service journalism.

Computational journalism

Though computational journalism has drawn a substantial amount of scholarly attention, it is a far smaller niche than data journalism. Early research on computational journalism overlapped heavily with data journalism, with some of the same projects such as the *Guardian*'s MPs' expenses and EveryBlock serving as examples of both (Flew et al., 2012; Hamilton and Turner, 2009). But in recent years it has been distinguished from data journalism as a primarily academic approach to quantitative journalism (Gynnild, 2014; Royal and Blasingame, 2015; Usher, 2016). Computational journalism's structural roots are in the American academy, with some early adoption in Europe, starting with courses at the Georgia Institute of Technology and programs at Northwestern University and Columbia University in the late 2000s (Anderson, 2013; Gynnild, 2014). Its common definitions as the application of social science or computer science methods to journalism (Flew et al., 2012; Hamilton and Turner, 2009; Usher, 2016) reinforce its deep academic ties.

But computational journalism has shifted into the professional sphere in recent years, largely through the rise of automated and algorithmic journalism. Consider the shift in definitions given to computational journalism by one of its pioneers, Nicholas Diakopoulos. In a brief but influential paper on computational journalism published in 2011, Diakopoulos defined computational journalism in terms of "the application of computing and computational thinking to the activities of journalism" (Diakopoulos, 2011). Five years later, he defined computational journalism instead as "finding, telling, and disseminating news stories with, by, or about algorithms" and referred to it interchangeably with algorithmic journalism (Diakopoulos and Koliska, 2017: 810; see also Koliska and Diakopoulos, this volume, Chapter 19). Computational journalism had moved from being more generally the application of all forms of computing and computational thinking to journalism to the more specific and practical application of algorithms to journalism. This change reflected a fusion of the two ideas in the practice of computational journalism itself, as the use of algorithms in journalism began to become, in many ways, *the* application of computing to journalism. Diakopoulos' practical merging of computational and algorithmic journalism is a perceptive move, sharpening its distinction from data journalism and grounding it more deeply in journalistic practice.

Beyond algorithms, the other core element of computational journalism is *computational thinking*. As articulated by Wing (2008) and others, computational thinking is the process of abstracting information or problems beyond their immediate material context and using fundamental concepts of computer science such as automation (this is where computational thinking is

especially closely tied to algorithms as an output) and recursivity. The thought process in computational thinking goes beyond mere abstraction, though – philosophers also specialize in abstracting problems, although most of them could hardly be called computational thinkers. Specifically, it involves approaching abstraction in computational terms – that is, binary code and the machine processes that code triggers – and understanding information in terms of what can be encoded and related between databases (Anderson, forthcoming). More broadly, it centers on the question of what can and cannot be calculated, and how – which, Bucher (2017) notes, "is not just a purely technical question; it is also a deeply social, cultural, political and economic one" (930).

These two elements – the cognitive process of computational thinking and the technical output of automation or algorithms – complement each other to form the core components of computational journalism, both as process and product. Where these elements had appeared in relation to journalism almost entirely in an academic context several years ago, they are beginning to make their way into professional practice through both processes of algorithmic accountability (Diakopoulos and Koliska, 2017) and journalism as structured data (Anderson, forthcoming).

Finally, it's important to note a couple of elements that are common among both data and computational journalism, as well as adjacent forms such as CAR and algorithmic journalism. As Splendore (2016: 345) notes, all of these forms are defined by their hybridity. They all represent a convergence of several fields: journalism, computer science, social science, and data analysis and visualization. This hybridity is not simply incidental; it is a constitutive and animating element of these forms. They are not defined by the norms of any one of these fields but instead by the generative tension created when these norms run up against each other and shape new articulations and practices. These forms of quantitative journalism are continually shifting between normative and epistemological poles drawn from these fields, and the tensions created by these shifts have led to both the explosion of new forms and the overlap between them (Borges-Rey, 2017). These varying forms are ultimately different combinations of very similar epistemological, social, and cultural elements.

More broadly, as Anderson (*forthcoming*) and Borges-Rey (2016) argue, all of these forms are means by which journalists seek to present their knowledge claims as more certain and to solidify their own professional and social legitimacy. The notions of 'objective' data, computational neutrality, social scientific rigor, and rituals of transparency are all at least in part strategies to garner public trust and authority for journalism in an era when both are flagging throughout much of the world and especially in the United States. All of these forms adhere to some type of objectivity and transparency but reinterpret them to incorporate norms from other converging fields and to make them more palatable for a more skeptical public (Splendore, 2016).

Conclusion

I do not conclude this chapter by offering yet another typology of quantitative journalism, or adding another dimension to an existing typology, or defining or redefining a form. As we have seen, there are plenty of typologies and definitions already; we have many of the conceptual, historical, and intellectual tools we need to think clearly and precisely about these journalistic forms. We would do well to use those tools to continue to prune excess classifications and allow a clearer and more focused conceptualization of quantitative journalism to bear fruit. But we must also be responsive to developments in the fields we are studying and leave conceptual space for forms to evolve and new practices to emerge. To that end, I offer a few final thoughts on the conceptualization of quantitative forms of journalism.

I have spent little time addressing CAR directly in this chapter, and scholars have been divided on whether CAR is still present in contemporary journalism in any meaningful way. I have

argued in the past that while CAR continues to be practiced, "it appears to be invoked more often as a historical mode of quantitative journalism than a contemporary practice" (Coddington, 2015: 334). CAR does indeed still exist: to the extent that journalists are still working with data in a way that is subordinated to the traditional professional, investigative journalistic paradigm and incorporates little interactivity, substantial transparency, or data as story, they are practicing CAR. This characterization generally tracks with statements about CAR made by Usher (2016) and Anderson (*forthcoming*). This is becoming a smaller part of journalism, though, and one whose significance is more in establishing the historical context for more widely practiced forms like data journalism than in describing widespread everyday journalistic practice. Because of that, dropping CAR from typologies of quantitative journalism (e.g., Splendore, 2016) is not necessarily a problem. But it remains important to continue to understand what CAR has stood for and how it relates to the other modes of quantitative journalism, as Anderson's (*forthcoming*) valuable history does. As we discuss the continual tension between continuity and change within data and computational journalism, CAR represents the continuity pole in that dialectic and thus remains a crucial area for continued study and conceptualization.

With CAR playing a more historical-contextual role and computational journalism more clearly splitting off to characterize largely algorithmically based work, it appears that data journalism is emerging as the central term for this area of journalistic activity, and appropriately so. It matches the dominant term used for this type of practice in many parts of the world, and in broad strokes, at least, its themes seem to accurately characterize the most prominent mode of quantitative journalism at this point – though more so at large or elite news organizations than smaller or local ones (Fink and Anderson, 2015). Other terms like computational journalism, interactive journalism, or hacker journalists have great value in clarifying different specializations within this area, and those should continue to find use and generate fruitful research. But if researchers begin to settle on a carefully explicated conceptualization of data journalism as their most common classification for this area of practice, that would likely be a healthy and welcome development.

I outlined this chapter with the goal of helping to clear up some of the haziness surrounding the classification of quantitative forms of journalism. But some haziness is inevitable when we are examining a constellation of practices and norms that is drawn from a complex and shifting set of disparate influences. Data journalism is both new and old, professional and marginal, participatory and exclusive, static and shifting. Those tensions make it very difficult to understand and convey fully, especially when we attempt to account for social, cultural, historical, epistemological, and professional factors in addition to the more conspicuous technological ones. But it is a worthwhile and necessary endeavor nonetheless, and even as we push toward a clearer picture of quantitative journalism, some of the most beneficial research will be the work that continues to trouble our neat distinctions with complex and nuanced perspectives arising from a fascinatingly unruly object of study.

Further reading

With so much research pouring out on this subject, it can be difficult to single out just a few pieces to commend for further reading. But there are a handful of recent works that have been especially insightful in clarifying quantitative forms of journalism. Nikki Usher's *Interactive Journalism* (2016) is especially valuable for its vivid and empirically grounded conceptualization of several emerging areas of journalistic work. Her concept of the professional subspecialty is also extremely useful in helping to clarify the reasons that data journalists adopt, transform, or reject professional journalistic values. C. W. Anderson's *Apostles of Certainty (forthcoming)*, on the history

of data journalism, is a crucial examination of the roots and social context of these phenomena, and a useful reminder of their shared epistemological lineage. And Eddy Borges-Rey's recent work (2016, 2017) is some of the most nuanced conceptualization of data journalism and other related forms that has been undertaken so far.

References

Anderson, C. W. (2013) "Towards a sociology of computational and algorithmic journalism." *New Media & Society*, 15(7), 1005–1021.

Anderson, C. W. (2015) "Between the unique and the pattern: Historical tensions in our understanding of quantitative journalism." *Digital Journalism*, 3(3), 349–363.

Anderson, C. W. (forthcoming) *Apostles of Certainty: Data Journalism and the Politics of Doubt*. Oxford: Oxford University Press.

Ausserhofer, J., Gutounig, R., Oppermann, M., Matiasek, S. and Goldgruber, E. (2017) "The datafication of data journalism scholarship: Focal points, methods, and research propositions for the investigation of data-intensive newswork." *Journalism* [online before print]. doi:10.1177/1464884917700667

Borges-Rey, E. (2016) "Unravelling data journalism: A study of data journalism practice in British newsrooms." *Journalism Practice*, 10(7), 833–843.

Borges-Rey, E. (2017) "Towards an epistemology of data journalism in the devolved nations of the United Kingdom: Changes and continuities in materiality, performativity and reflexivity." *Journalism* [online before print]. doi:10.1177/1464884917693864

Bucher, T. (2017) "'Machines don't have instincts': Articulating the computational in journalism." *New Media & Society*, 19(6), 918–933.

Coddington, M. (2015) "Clarifying journalism's quantitative turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting." *Digital Journalism*, 3(3), 331–348.

De Maeyer, J., Libert, M., Domingo, D., Heinderyckx, F. and Le Cam, F. (2015) "Waiting for data journalism: A qualitative assessment of the anecdotal take-up of data journalism in French-speaking Belgium." *Digital Journalism*, 3(3), 432–446.

Diakopoulos, N. (2011, April 22) *A Functional Roadmap for Innovation in Computational Journalism*. www.nickdiakopoulos.com/2011/04/22/a-functional-roadmap- for-innovation-in-computational-journalism/

Diakopoulos, N. and Koliska, M. (2017) "Algorithmic transparency in the news media." *Digital Journalism*, 5(7), 809–828.

Dörr, K. N. (2016) "Mapping the field of algorithmic journalism." *Digital Journalism*, 4(6), 700–722.

Felle, T. (2016) "Digital watchdogs? Data reporting and the news media's traditional 'fourth estate' function." *Journalism*, 17(1), 85–96.

Fink, K. and Anderson, C. W. (2015) "Data journalism in the United States: Beyond the 'usual suspects'." *Journalism Studies*, 16(4), 467–481.

Flew, T., Spurgeon, C., Daniel, A. and Swift, A. (2012) "The promise of computational journalism." *Journalism Practice*, 6(2), 157–171.

Gray, J., Bounegru, L. and Chambers, L. (2012) *The Data Journalism Handbook*. Sebastopol, CA: O'Reilly Media.

Griffin, R. J. and Dunwoody, S. (2016) "Chair support, faculty entrepreneurship, and the teaching of statistical reasoning to journalism undergraduates in the United States." *Journalism*, 17(1), 97–118.

Gynnild, A. (2014) "Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets." *Journalism*, 15(6), 713–730.

Hamilton, J. T. and Turner, F. (2009) *Accountability Through Algorithm: Developing the Field of Computational Journalism*. Behavioral Sciences Summer Workshop, Stanford, CA. Retrieved from www.stanford.edu/~fturner/HamiltonTurnerAccbyAlgFinal.pdf

Hammond, P. (2017) "From computer-assisted to data-driven: Journalism and big data." *Journalism*, 18(4), 408–424.

Hewett, J. (2016) "Learning to teach data journalism: Innovation, influence and constraints." *Journalism*, 17(1), 119–137.

Holovaty, A. (2006, September 6) *A Fundamental Way Newspaper Sites Need to Chang*. Retrieved from www.holovaty.com/writing/fundamental-change/

Holovaty, A. (2009, May 21) *The Definitive, Two-Part Answer to 'Is Data Journalism?'*. Retrieved from www.holovaty.com/writing/data-is-journalism/

Howard, A. B. (2014) *The Art and Science of Data-driven Journalism.* New York, NY: Tow Center for Digital Journalism. Retrieved from http://towcenter.org/wp-content/uploads/2014/05/Tow-Center-Data-Driven-Journalism.pdf

Knight, M. (2015) "Data journalism in the UK: A preliminary analysis of form and content." *Journal of Media Practice*, 16(1), 55–72.

Lesage, F. and Hackett, R. A. (2014) "Between objectivity and openness – The mediality of data for journalism." *Media and Communication*, 2(2), 42–54.

Lewis, S. C. (2015) "Journalism in an era of big data: Cases, concepts, and critiques." *Digital Journalism*, 3(3), 321–330.

Mair, J. and Keeble, R. L. (2013) *Data Journalism: Mapping the Future.* Suffolk: Abramis.

Michalski, D. (2016) *Reader Engagement with Data Journalism: Comparing the Guardian and Washington Post's Coverage of People Killed by Police.* Unpublished master's thesis, Hank Greenspun School of Journalism & Media Studies, University of Nevada, Las Vegas.

Parasie, S. (2015) "Data-driven revelation? Epistemological tensions in investigative journalism in the age of 'big data'." *Digital Journalism*, 3(3), 364–380.

Parasie, S. and Dagiral, E. (2013) "Data-driven journalism and the public good: 'Computer-assisted reporters' and 'programmer-journalists' in Chicago." *New Media & Society*, 15(6), 853–871.

Rogers, S. (2013) *Facts Are Sacred: The Power of Data.* London: Faber & Faber.

Royal, C. and Blasingame, D. (2015) "Data journalism: An explication." *#ISOJ*, 5(1). Retrieved from https://isojjournal.wordpress.com/2015/04/15/data-journalism-an-explication/

Splendore, S. (2016) "Quantitatively oriented forms of journalism and their epistemology." *Sociology Compass*, 10(5), 343–352.

Stiles, M. (2017, February 24) "Charting NICAR attendance, over the years." *The Daily Viz* [blog post]. Retrieved from http://thedailyviz.com/2017/02/24/charting-nicar-attendance-over-the-years/

Tandoc, E. C. and Oh, S.-K. (2017) "Small departures, big continuities? Norms, values, and routines in the Guardian's big data journalism." *Journalism Studies*, 18(8), 997–1015.

Usher, N. (2016) *Interactive Journalism: Hackers, Data, and Code.* Urbana, IL: University of Illinois Press.

Uskali, T. I. and Kuutti, H. (2015) "Models and streams of data journalism." *The Journal of Media Innovations*, 2(1), 77–88.

Wing, J. M. (2008) "Computational thinking and thinking about computing." *Philosophical Transactions: Series A, Mathematical, Physical, and Engineering Sciences*, 366(1881), 3717–3725.

Young, M. L. and Hermida, A. (2015) "From Mr. and Mrs. Outlier to central tendencies: Computational journalism and crime reporting at the Los Angeles Times." *Digital Journalism*, 3(3), 381–397.