

University of Nebraska - Lincoln  
**DigitalCommons@University of Nebraska - Lincoln**

---

Professional Projects from the College of Journalism and Mass Communications, College of  
and Mass Communications

---

11-2016

# The Ethics of Data Journalism

Rebekah E.D. McBride

*University of Nebraska-Lincoln*, [beahmcb@gmail.com](mailto:beahmcb@gmail.com)

Follow this and additional works at: <http://digitalcommons.unl.edu/journalismprojects>

 Part of the [Journalism Studies Commons](#)

---

McBride, Rebekah E.D., "The Ethics of Data Journalism" (2016). *Professional Projects from the College of Journalism and Mass Communications*. 9.

<http://digitalcommons.unl.edu/journalismprojects/9>

This Thesis is brought to you for free and open access by the Journalism and Mass Communications, College of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Professional Projects from the College of Journalism and Mass Communications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

## Professional Project Part 1

---

### THE ETHICS OF DATA JOURNALISM

---

© Rebekah McBride, 2016

#### Abstract

Data journalism is a growing trend that offers journalists new and exciting ways to tell stories. In a general sense it can be defined as “obtaining, cleaning and analyzing data for use in telling journalistic stories”<sup>1</sup> but its application is much more complicated than this simple definition. Data journalism has grown alongside the burgeoning computational technology industry, but these huge gains in technology have not been followed by similar gains in data-journalism scholarship and education. In particular, research and college-level courses related to the ethics of data journalism are lacking. The lack of ethics preparation will provide significant challenges for future journalists looking to obtain the interpretation and critical thinking skills they need to write compelling and ethical data-based stories. With that in mind, this paper investigates the ethics of data journalism in the context of education. Through interviews with industry experts and an analysis of the latest research, several best practices became clear. This paper highlights a few of those key ethical guidelines including the importance of placing all data in context and minimizing harm to those in the news. While data journalism warrants some special considerations, many of the ethical guidelines that are needed to support data journalism are already in place. Ultimately, journalists must simply make an effort to apply those guidelines to data and then share those guidelines through education and scholarship.

## Introduction

It has been said that “stories are just data with a soul,”<sup>2</sup> but contextualizing data in order to form a meaningful story provides many unique challenges. At its core, the mission to find meaning in data can be labeled as data journalism, but this overarching term is hard to define. In fact, scholars have long debated just what constitutes data journalism and whether it is different from computer-assisted reporting and computational journalism. In 2015, Mark Coddington, an assistant professor of journalism and mass communication at Washington and Lee University, attempted to offer an outline of the various data journalism practices in his paper, *Clarifying Journalism’s Quantitative Turn*. He argued that computer-assisted reporting, data journalism and computational journalism were all slightly different in their approach, but he ultimately admitted that no area is “mutually exclusive.”<sup>3</sup> According to Coddington, computer-assisted reporting is at the core of all data-driven journalism practice. Dating back to 1950 when Phillip Meyer, “the godfather of computer-assisted reporting,”<sup>3</sup> began outlining precision journalism, the method relies on the social science model or the idea of using “statistical analysis to achieve more definitive answers to journalistic questions.”<sup>3</sup> Therefore, Coddington believes that computer-assisted reporting is very much embedded in the ethics and practices of traditional journalism and is more about producing a “better story” than it is about technology. Computational journalism, on the other hand, is “centered on the application of computing and computational skills...rather than the journalistic use of data or social science methods more generally.”<sup>4</sup> In other words, computational journalism relies on algorithms to tell the story while computer-assisted reporting is much more about the story than it is about the data, algorithms or the technology. To clarify, examples of computational journalism

include *The Wall Street Journal's* "use of simulated user profiles to determine the algorithms governing price discrimination in online commerce."<sup>5</sup> Finally, Coddington defines data journalism as "essentially any activity that deals with data in conjunction with journalistic reporting and editing."<sup>6</sup> In his mind, data journalism includes *The Guardian's* reports on the expense claims made by Members of the United Kingdom's Parliament, because data journalism is "focused on opening data to the public... data analysis and data visualization."<sup>6</sup>

While an interesting analysis of the various methods and history, even Coddington admits that data journalism has become the overarching "term of choice" for the industry. As such, data journalism will be used here to describe any and all forms of journalism involving data or other computational methods including those that Coddington discussed in his research. In terms of a definition, data journalism will be defined as "a workflow, where data is the basis for analysis, visualization and –most importantly- storytelling."<sup>7</sup>

While the definition remains a topic of contention in the industry, the exact moment that data journalism was born is even harder to pinpoint. In fact, few scholars can agree on a story that started it all. For example, *The Manchester Guardian* (now known as *The Guardian*) claims to have published the very first data journalism story back in 1821, but some argue that Florence Nightingale's work may have been the very first attempt to visualize data in a story.<sup>8</sup> Meanwhile, other scholars argue that true data journalism requires computational technology, moving the birth to 1952 when CBS used a computer to predict the outcome of the presidential election.<sup>9</sup>

While its definition and exact date of conception may be highly debated, its impact on journalism is not. Over the past few decades, data journalism has grown rapidly, allowing

previously untold stories to emerge from the depths of databanks. Many influential stories including those involving the Afghanistan war logs and the recent Panama Papers would not have been possible without databases and journalists trained to interview such databases, but this new method of storytelling is not without its challenges. In particular, data journalism requires a journalist not only to gain the skills required to interview a database, but they must also be prepared to consider the ethical implications of hacking, dirty data and more. Journalists must also begin to think about harm in the news and other ethical issues in a new light. For example, journalists must become very aware of the fact that each data point represents a person's information, and therefore, datasets must be treated not as simple numbers, but as a living piece of the story that require careful consideration in terms of accuracy and ethics.

Unfortunately, acquiring these skills can be a challenge as a recent study showed. In fact, *Teaching Data and Computational Journalism*,<sup>10</sup> suggests that data journalism courses are a rarity on college campuses. The study showed that of the 113 institutions used in the study, only 59 offered one or more data journalism courses. The study further proves that as the increase in overall data collection goes up, education and scholarly research in this area lags. That's why education surrounding the methods and ethics involved in using data will be key to avoiding future missteps. As such this paper proposes a set of ethical guidelines that are supported by the latest scholarly articles as well as recommendations from those in the field. These guidelines will be positioned in terms of the four steps of data journalism outlined in *The Guardian's* article, *How to Be a Data Journalist* by Paul Bradshaw, a journalist and co-founder of helpmeinvestigate.com. These steps include: a) finding data b) interrogating data c) visualizing

data and d) mashing data.<sup>11</sup> In addition to the proposed ethical standards, this paper also seeks to serve as the jumping off point for a discussion about data journalism education and the need for courses that consider and discuss the ethical implications of this field.

### **Finding Data: The Challenges of Collecting Data**

All data journalism requires data, but finding or collecting data poses many logistical and ethical challenges. Despite the difficulty, it is important to remember that this step serves as the backbone to the story, so understanding the technical needs and ethical challenges is key to creating a quality end product. In terms of data collection, several decisions will need to be made along the way, but among the first decisions made will be whether the team will need to gather its own data or use a database that already exists. If a team of journalists decides to gather data, then step one will require “good contacts or the technical skills to gather data.”<sup>12</sup> In other words, journalists will need to have access to and preferably some expertise in programs that can gather and/or assist with the visualization of data. Some examples of these programs are Google Fusion Tables and Python, but many others exist.

Given the many programs available, journalists are not at a loss for technological resources, but a clear understanding of the ethics and capabilities of each program is lacking. In fact, it’s quite difficult to find scholarly research on data collection for journalistic purposes, despite the fact that databases potentially open the door for ethical infractions. For example, Google Fusion allows users to access publically available data. While convenient, Google Fusion Tables’ public nature also brings into question the quality of that data as well as the ownership of the data. In other words, who collected the data and why? This, in turn, brings up the ethical question, should journalists be accountable for the collection, analysis and presentation of the

data? Accountability is a part of many ethical codes. In fact, the Society of Professional Journalists states that accountability is at the “heart of journalism.”<sup>13</sup> The society’s ethical code also states that a “willingness to be accountable is what cements a journalist’s credibility”<sup>14</sup> making accountability one of the core ethical concerns that data journalists need to consider as they collect and interpret data.

Accountability, is not always easily defined, however, as data journalist Eric Litke<sup>15</sup> explained, “A journalist is responsible for understanding the data. Period. This means knowing how it was gathered and entered, whether there were changes in that entry over the time period being studied, what limitations the data has, and where errors could enter the system. If these questions have been asked and the reporter was lied to, there’s nothing to do about that, but these are critical questions that must be asked.”

In other words, Litke believes that a journalist is accountable for asking the right questions but not for falsified data. On the other hand, Litke continued, “Data cannot be taken at face value, as doing so requires a litany of assumptions. I’m reminder of a story where I wrote about data covering a 10-year timeframe. I made what I thought was a reasonable assumption, that the data in the database was internally consistent, but that was wrong and led to reporting wrong numbers. I should have asked the questions necessary to find that. When such data inconsistencies surface, the journalist at a minimum has to reveal this prominently to the reader, but more likely the data should be limited to a group that is internally consistent or new data should be gathered. We must not just investigate with data, but investigate the data we are using.”

Despite some of the concerns surrounding accountability and data input programs, Google Fusion Tables and programs like it have led the way for data journalism. In fact, Google Fusion Tables and Google Maps were used in *The Guardian's* coverage of The War Logs.<sup>16</sup> Tableau, a fee-for-service, analytics program has also been used by various news outlets, including *The Post-Crescent* (a part of the *USA Today* Network), of Appleton, Wisconsin. According to Kathleen Culver, the director of the Center for Journalism Ethics at the University of Wisconsin-Madison, *The Post-Crescent* has done a great deal of ethical work in the data journalism field, especially in terms of collecting its own data. For example, in 2016, *The Post-Crescent* used data from the Wisconsin online court records (CCAP) to “examine the 208 sealed felony cases from 2005 to 2014.”<sup>17</sup> The results were then turned into a data map that showed “which judges and counties most frequently sealed court records over the last decade in Wisconsin.”<sup>18</sup>

Litke was the reporter on that project and utilized Tableau to create his story. He said, “Tableau is the most nuanced and agile data visualization software I’ve used, and has been my go-to for years. The problem is that it is rarely mobile-optimized and doesn’t play well with our content management system on mobile. With a quickly growing portion of the audience on mobile, in some cases the majority, this is not acceptable, so we, for now, have had to abandon embeds in favor of static images. We’ll then link to a standalone page with a Tableau or other graphic,” Litke said. “For mapping, we use Google Maps, Fusion Tables and increasingly CartoDB. All have strengths and weaknesses, but again, mobile implementation is not ideal so it limits options. I have no ethical concerns unique to these tools. The data presentation just



requires the same transparency and context as any other data presentation. I typically include a tab among the various options that includes an about/how we did it explainer.”

Litke also worked on *The Post-Crescent's* collection of data for a story on heroin overdoses in Wisconsin.<sup>19</sup> In 2015, after several months spent interviewing coroners in every county to develop data on the number of heroin overdoses in Wisconsin, *The Post-Crescent* published an interactive story that alerted people to the serious heroin problem in Wisconsin. It also showed which counties needed the most aid for addiction and recovery.

In this case, Litke started with a basic dataset. He said, “I followed that up with a lot of traditional reporting, requesting records and reaching out to judges and clerks to find out what the process was to seal cases. That provided examples to show and along the way it showed me what kinds of cases to omit from the tally, aiding in the data cleaning. The key to me was talking to the people that made the decisions to seal and entered the record of that online, so I can get firsthand accounts of the process, and how it varied between counties. Of course, the very nature of this case, sealed records, brought up a unique ethical consideration that the attorneys in some cases pointed out — we are reporting on cases that judges had ordered sealed, and in some cases identifying the people who were sealed. We wouldn't have done this in cases where it was sealed to protect someone from physical harm (like a domestic violence victim), but these are otherwise records that are or were previously public. Digging into this was critical to allowing the public to gauge whether judges are properly using the ability to seal, and whether that's being done consistently around the state. We found it's certainly not consistent, and there are certainly arguments to be made that many of the cases we revealed had no business being sealed.”

The point is that there are many examples of useful data journalism methods and tools used in stories from the Warlogs to local issues, but these powerful stories must be backed by powerful ethical standards in order to maintain journalistic integrity. That means not only being accountable for understanding the data but also understanding who is represented in the data. Scholars have noted that there are several issues with sampling methods that can lead to inaccurate data. As Culver said, survey methodology is one of the main concerns when it comes to the ethical collection of data for a story.

“When we talk about surveys there’s a lot of concern about the methods for sampling. For example, surveys are often set up for landline phone but we’ve discovered that ethnic minorities are more likely to have cell phones, which leaves them out of these surveys,” Culver said.

Michelle Robinson, a data analyst for Race to Equity in Madison, Wisconsin, and a graduate student in the Sociology department at the University of Wisconsin-Madison, agreed with Culver, but also stated that there doesn’t seem to be any one methodology that is truly capturing the impact behind the numbers.

“There seems to be a mismatch between racial inequality and the types of methodologies I’ve been trained in,” said Michelle Robinson. “I’m very skeptical of the by-design reductionist methods that tend to flatten out an interactive environment.”

Given these inherent issues with data collection, journalists must take the time to investigate the collection methodology before moving forward with a story. In 2014, the Canadian government discovered the need for such an ethical standard the hard way. At the time, the government was looking to gather data on unemployment rates to get a better

handle on where temporary foreign workers were needed and where their presence might hinder companies from hiring citizens. Statistics Canada, a public agency within the Government of Canada, set out to measure the unemployment rate in various areas and the results showed that the official unemployment rate in Manitoba, Alberta and Saskatchewan were under the high unemployment rate of six percent, meaning that the temporary foreign workers program could be used in those areas. It has since been discovered, however, that the data did not include any of the Aboriginal groups living in those areas, groups, which *The Globe and Mail's* story on the topic described as "highly likely to be without work."<sup>20</sup> This is a serious oversight considering the fact that Saskatchewan alone hosts more than 30 First Nation reserves. The ethical misstep is clear; by eliminating such a large amount of people from the survey, the results were severely slanted resulting in an inaccurate and potentially damaging account of the employment situation in Canada. This lack of information alone is a major ethical issue in terms of accountability, but this also brings up the even greater issue of data's potential effect on minority groups.

In regards to this topic, Culver said, "Data sets generally focus on publicly available information such as incarceration rates or who is receives aid. Unfortunately, these are issues that often highlight minority groups. It's important to keep in mind, however, that this is an issue in society and not a journalism issue. For example, minority groups are facing high re-incarceration rates in many areas, but that is a social issue. The data is simply bringing it to people's attention."<sup>21</sup> Culver cautioned, however, that this topic could become a journalism issue if the reporter draws false conclusions about the data or places the data in a context that perpetuates stereotypes. Overall, she warned that context is key and that helping people to

understand the story behind the statistics is a huge part of responsible data-based journalism. She also recommended gathering or locating multiple data sets that represent the same topic, to better identify outliers or mistakes in the data.

*Big Data, Big Problems: Emerging Issues in the Ethics of Data Science and Journalism* by Joshua Fairfield and Hannah Shtein<sup>22</sup> also echoed this concern stating that “out of any dataset, some people will be disproportionately affected by data,” but that disadvantaged subgroups are likely to be more vulnerable.”

In regards to this Michelle Robinson<sup>23</sup> said, “People often come to data with their own preconceived notions, so we need to think ethically about how to approach data.” She noted that real structural constraints on time and resources exist for journalists that seem to justify the fact that releasing the numbers is good enough, but in her opinion it is not. “Even when you release numbers with context, it doesn’t mean it will be received, no matter how argued the contextualized data is.”

In fact, Michelle Robinson cited a recent study by Stanford assistant professor of psychology Jennifer Eberhardt and postdoctoral fellow Rebecca Hetey titled, “Racial Disparities in Incarceration Increase Acceptance of Punitive Policies,”<sup>24</sup> which showcased some of the preconceived notions that data journalists must work against when presenting data. In particular, the study showed “that exposing people to extreme racial disparities in the prison population heightened their fear of crime and increased acceptance of the very policies that lead to those disparities.”

Michelle Robinson believes in the power of context in fighting these stereotypes but, in her opinion, context must go beyond an interview with just one person from a particular

community. In fact, an interview with only one member of the community can be just as harmful as no context at all.

“Locally, I see either you have either a stereotype or an exception interviewed” Robinson said. “Journalists just tend to draw from the tail end of the data. The more responsible thing to do, however, is to become embed in a community, build relationships with folks and really diffuse an understanding of what it means to be a member of that community.”

Michelle Robinson also encourages news organizations to continue to push for more diversity in the newsroom. In Madison, Wisconsin, for example, she is seeing non-profit community publications such as Madison 365, which “fosters dialogue between members of diverse communities,”<sup>25</sup> working with larger local publications to incorporate a broader sense of the community into journalism.

Sue Robinson, an assistant professor of journalism at the University of Wisconsin-Madison who recently completed a book on race and the media<sup>26</sup>, also believes in the need for these partnerships as she has spent a great deal of time studying what voices are not being heard and how power affects who gets to say what about communities.

“We’re seeing and hearing more voices than ever,” Sue Robinson said. “But if someone is in a marginalized community blogging, they’re still in a silo and not networked into the power structure.”

Given this issue, she suggests that regardless of whether a reporter is a data journalist or just someone covering data, each reporter needs to “adopt practices that are more inclusive.”

This need for inclusivity and appropriate context that fights against reader's preconceived notions is something that experts promote, but the continuing cuts occurring at newspapers around the country have left fewer resources available to journalists who are tasked with analyzing huge amounts of data.

"My research shows that newsroom resources have been cut, in many cases, by about 60 percent over the last ten years," Sue Robinson said. "This affects data coverage because we can have all the tools that allow us to create databases that look at incarceration rates but unless you have someone serving as a synthesizer and analyzer that data is essentially lost."

Despite the lack of resources facing many data journalists, these issues are no excuse for causing harm to those in the news. A part of many ethical codes, including the popular Society of Professional Journalists (SPJ) Code of Ethics, minimizing harm to those in the news is essential for ethical journalism. As such, any ethical journalist should be considering potential harm whenever a story is published but thinking about harm in terms of data is a new concept for many editors and writers. As Matt Waite, a professor of practice in the College of Journalism and Mass Communications at the University of Nebraska-Lincoln and principal developer of the Pulitzer Prize-winning Politifact stated in an interview, "journalists don't necessarily need a new set of ethics or ethics codes; they need to think about harm through a new lens."<sup>27</sup>

He suggests that editors consider the difference between a journalist having public access to a document and a journalist's actually publishing that document to a mass audience. Like Culver, he cautions that the success of a data-based story relies heavily on a journalist's ability to put the data in context. Without such context, harm to those in the news can occur in such a way that the government steps in and shuts down records. As such, Waite believes all

journalists need to realize that they are accountable and should consider the possible outcomes of releasing data.

One example of a negative outcome as it related to databases, occurred in 2012 when *The Journal News* of Westchester, New York, published an interactive map with the names and addresses of citizens in two local counties who held gun permits. The article was published in response to the Newton shootings<sup>28</sup> but instead of opening up a dialogue about a public issue such as flaws in gun permits or the relationship between permits and crimes<sup>29</sup> the publication opened up a discussion about privacy and safety concerns in relation to data. The public outcry surrounding the release of this data was so intense that a local state Senator began pushing for legislation that would close the records. While there are many issues to discuss in this case, newsworthiness and harm to those in the news are two of the greatest issues at play here. Publishing data for the sheer sake of publishing, does not take into account harm to those in the news or the overall newsworthiness of the data, and dismissing these factors can lead to unfortunate changes in public record laws. So, just as the Society of Professional Journalists states in its Code of Ethics, journalists must “recognize that legal access to information differs from an ethical justification to publish or broadcast.”<sup>30</sup>

While there are many ethical and technical issues to take into consideration when collecting data, if done correctly, data-based journalism can lead to groundbreaking stories that open the door for an innovative era in journalism. As Waite says, “To not involve data, when the amount of data available is exploding is insane. Data helps to tell the story of a community and to ignore such data is to make yourself willfully ignorant. The biggest challenge with data

journalism, however, is that many journalists do not have the skills to analyze or interview the data.”<sup>19</sup>

### **Interrogating Data: The Challenges of Interrogating Data**

Information accessibility and diversity have continued to grow, but “knowing how to find what one is looking for is a threshold many people have yet to cross.”<sup>31</sup> The “data rush”<sup>32</sup> has put huge amounts of data at journalists’ fingertips but, “the material is [generally] not indexed in any meaningful way.”<sup>33</sup> Therefore, accessing data and then interpreting it require technical skills that previously were not associated with journalism. Once a field dominated by professionals versed in writing and editing, journalism has shifted to include “journalists, software developers and computer scientists inside and outside of established news organizations.”<sup>34</sup> This technological shift has been brought about many positive opportunities, but it is not without its challenges. As Culver, said, “Technology is across the board giving us new ways to tell stories so data journalism is an important development, but if it’s not done ethically it’s a huge problem. We need to be teaching common statistical reporting to all journalism students so that they are not dismissing this as an ‘I can’t do this’ scenario.”<sup>35</sup>

To avoid such a scenario, it is exceedingly important that journalism students and professionals alike understand the process of interrogating data. In fact, literature is now suggesting that “the press should be responsible for knowing how to work with large data sets in order to hold governments, or anyone else, accountable.”<sup>36</sup> Some publications have taken this to heart, developing their own databases and data-based journalism teams. For example, *The Guardian* has *Content API*, *Datastore* and a *Data Blog*, which are a part of their open-



platform initiative.<sup>37</sup> It's also worth noting that the Data Blog allows readers to work with data sets and submit their own visualizations and feedback. *The New York Times* has also been involved in several data-based journalism stories including the Investigative Reporters and Editors (IRE) Medal-winning *Toxic Water* series, which required a team of reporters, programmers and videographers to create a story from data that was collected through the Freedom of Information Act.<sup>38</sup>

As publications have experimented with data collection and interrogation, a few ethical issues have come to the surface, namely “comparatively shallow measures, lack of context awareness and a dominance of automated methods of analysis.”<sup>39</sup> In terms of interrogating data, data verification has been a particular technical and ethical challenge. When a large dataset is located, it is often difficult for journalists to verify the accuracy of the data or immediately understand the motivation behind those collecting the data. As such, interrogating data requires a great deal of time and resources. Since the time constraints are so large, some scholars have pushed for the “non-interpretive use of raw data.”<sup>40</sup> This idea was perpetuated by Adrian Holvaty, a journalist and co-creator of the Django web framework, who stated that “newspapers need to stop the story-centric view.” He claims that if given access to databases, people can “create their own news.”<sup>41</sup> While perhaps more convenient, this method of data-based journalism revisits the issue of newsworthiness and whether data should be published just because it can be published. Additionally, anecdotal evidence suggests that despite the cost of interrogating data, publications see value in providing a context for data and analyzing it before publication.

*The New York Times* coverage of the Panama Papers is an excellent example of data interrogation ethics at work. On April 3, 2016, the International Consortium of Investigative Journalists began releasing stories that detailed the leaked data. When the original stories were released by the consortium, several papers printed those stories, but *The New York Times* decided to hold off on a major Panama Papers exposé. Perplexed by this decision, readers began writing to then-Public Editor Margaret Sullivan asking why *The New York Times* had not made a “big splash” in regards to the leaked data. Matt Purdy, a deputy executive editor, spoke with Sullivan about the decision saying, “The Times was not a part of the global consortium and was not aware that the story was coming, it needed some time to get its own story going. We didn’t know these documents were out there and being worked on,” he said. So, in the end, *The New York Times* decided to take a few days to interrogate the data and the stories before publishing a piece of their own. This is in line with the Society of Professional Journalists guideline to seek truth and report it, which includes verifying information before publishing it<sup>42</sup>. This is an example of an established ethical guideline that just simply needs to be looked at through the data-based journalism lens.

In terms of data interrogation, Sullivan also wrote a piece for *The New York Times* on the importance of context in data journalism titled, “Awash in Data, Thirsting for Truth.”<sup>43</sup> The article looked at two case studies, an exposé on working conditions at Amazon and an article that used data to prove that the creative apocalypse, or the idea that in the digital world it would be difficult to make money creating art, never happened. In the case of the Amazon article, the authors told the story through anecdotes instead of data and this angered the audience as they felt the authors had little data to prove that Amazon had made mistakes. On

the other hand, the magazine cover on the creative apocalypse used a great deal of data but, ultimately, the data was used out of context and readers felt the story was inaccurate. Sullivan used these examples to demonstrate the challenges editors and writers face when using data as a part of a story, but the stories also drive home the point that context plays a huge role in writing an ethical and well-balanced data journalism piece. “Death to Data Journalism”<sup>44</sup> by David Leonhardt made a similar point, stating that data can be deceiving. He argued that data journalism is not any different from any other form of journalism since it is simply about seeking the truth—just through a different means. Leonhardt also brought up a point that Waite and many other sources mentioned, which is the importance of complementing all data with good reporting. These examples show that without the human aspect, additional interviews or anecdotes, data is fairly meaningless, essentially just numbers floating on the page.

“Can Big Data Algorithms Tell Better Human Stories”<sup>45</sup> by Bernard Marr, challenges this idea, however, as the article discusses the development of computer programs that interrogate data without human intervention. As outlined above, accountability and context rank among the most important elements of a credible and ethical story, but the invention of *Quill*, *Automated Insights* and *SlamTracker*, technologies that review data and turn it into stories for publications, bring about a new set of ethical challenges. While the technology is currently only able to look at number-related patterns and turn those into basic stories, the inventor says that by 2017 Quill will be writing Pulitzer-Prize quality journalism without human help. This raises a huge ethical issue that needs to be considered. As many experts have stated in other articles, one of the best ways to avoid the traps of data journalism is to consider the harm to

the individuals and to make sure that the data is put in context. As such, automated data stories bring up the question of accountability and data validation. In other words, does “an increase in algorithmic judgment lead to a decline in the authority of human judgment?”<sup>46</sup>

Rather than allowing a computer to interrogate data, it would be best to educate the next generation of data-based journalists on the importance of a computational or algorithmic mindset. This means thinking about prioritization, classification, association and filtering.<sup>47</sup> Instead of simply focusing on the invention and use of technology, the future of journalism should be focused on developing news professionals who are able to “think abstractly and work collaboratively to solve problems, with the assistance of high tech tools.”<sup>48</sup>

### **Visualizing and Mashing Data: How the Final Product Will Affect the Future of Journalism**

In addition to the challenges of collecting and interpreting data, academics and professionals are also considering the role the final product will play in the future of journalism. While some believe data will “save” journalism, others believe it will be the downfall. At this point, the research is lacking. A clear understanding of data journalism or a consensus regarding its role in the future and its ethical effects does not exist. But, addressing the lack of education as well as formulating clear ethical codes will be paramount if data journalism is to become a successful part of the news industry.

*Teaching Data and Computational Journalism*<sup>49</sup> presents an in-depth look at the gap in data-based journalism education while also offering recommendations for industry educators and leaders who are interested in preserving and promoting this aspect of journalism. The study essentially looks at four areas of data journalism and how those areas are currently being

addressed as well as how they could be better addressed in the future. The areas include data reporting, data visualization and interactives, emerging journalistic technologies and computational journalism. Additional areas of concern noted in the research included the absence of foundational courses and textbooks available to students as well as a lack of coding and programming education opportunities. Essentially, the research showed that “graduates with data journalism skills are better equipped to succeed, [but that] many journalism programs offer few courses, and nearly half offer no classes at all.”<sup>50</sup>

The authors theorize that this gap is in part due to a lack of resources on campus, so they recommend pooling resources to provide experience and instruction in the area of data-journalism. For example, the Columbia School of Journalism has collaborated with the Earth Institute to create a database of climate-related information. The students then take courses on Python programming, which gives students the opportunity to use the data in their stories.<sup>51</sup> The report also discussed a collaboration between a watchdog reporting class and an engineering professor whose computer science students worked together with journalism students to learn and share skills. In addition to these collaborations the authors advocated for more education on actual technical skills such as coding, in hopes of providing students with a data mindset. In other words, students need to learn to “think beyond rows and columns as they search for answers in data of all forms, structured and unstructured.”<sup>52</sup> Developing this mindset could include a statistics-related course or collaborations such as those listed above, but it will also require practice and support from educators who are versed in the topic, something that many schools lack.

While a useful look at the overall gaps in education, the report does not spend a great deal of time elaborating on the need for ethics-based courses. The reason may be that “data journalism emerged from the professional practice, not academia.”<sup>53</sup> It’s still a fairly new practice that is in the experimentation phase and has, therefore, not been heavily studied by scholars in terms of its ethical implications. Whatever the reason, it is time to start a discussion about data-journalism ethics and the role this area will play in the future of journalism.

C.W. Anderson discusses this issue in his scholarly article, “Towards a Sociology of Computational and Algorithmic Journalism.”<sup>54</sup> After a thorough investigation of data journalism or as he calls it, “computational journalism,” and its role in the fourth estate, Anderson argues that to date, the scholarly and sociological approach to data journalism has not been properly executed. Essentially, he feels that a more academic approach to the topic is needed. Anderson points out that data journalism has generated a great deal of interest thanks to the narrative being circulated, which essentially says, “while technology is destroying journalism, technology contains within it the seeds of journalism’s rebirth,” but he argues that this statement has not been backed by research. To investigate this statement and form a framework, Anderson used Michael Schudson’s *Sociology of News Production* written in 1989 as the basis for his research, adopting Schudson’s four sociology “lenses” including political, economic, organizational and cultural. Anderson also adds technological and institutional as additional lenses that aid in understanding data journalism and its effects on society.

In regards to the political lens, Anderson says it is important to consider the “policy-level” initiatives that help or hurt data journalism. Anderson points out that so far there is little scholarly research on how public-policy or political sites such as *open* or *wiki* have aided in

journalism, and he is correct. Much of the research and publications on data fail to provide a clear look at how data can be properly used to aid journalism in its role as the fourth estate.

Anderson simply gives suggestions for future studies. For example, he would like to see someone investigate the data gap, in other words “critically interrogate the way that different institutionally specific resources constrain the options available to various news outlets and industry segments and the potential for software makers and those with access to data to spread it more evenly among small and large news outlets.”<sup>55</sup> As for the organizational lens, Anderson suggests that researchers focus on the ethnographic elements of the traditional newsroom versus the computational journalism newsroom. Specifically Anderson says “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.”<sup>56</sup>

In regards to cultural history, the final Schudson lens, Anderson is interested in the “impact of big data on journalistic sense-making.”<sup>57</sup> For example he would like researchers to investigate whether “data create a difference sense of what news is?” and whether “data encourages a greater emphasis on structures and social forces as opposed to personalities and incidents?”

Finally, Anderson addresses his new lens, technology and the news, which calls for more research in terms of news values and their role in digital design. Specifically, Andersen says there should be, “a scholarly concern with the actual role played by materiality and technology

in the processes of journalism” he believes that research might emphasize, “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.”<sup>58</sup>

Overall, Anderson’s article provides a summary of research and a number of excellent questions, but it fails to answer any. Anderson simply poses research questions and ideas but does not offer any clear thoughts on how data journalism has affected the industry. Instead, Anderson spends a great deal of the article pointing out that, in his mind, the industry is wrong in its approach toward data journalism. He says that thanks to the influence of bloggers and computer scientists, a proper scholarly discussion regarding the effects of data is off the table.

When discussing the ethics of data-based journalism, it’s also important to discuss the role Big Data, “an umbrella term for a variety of strategies and tactics that involve massive data sets”<sup>59</sup> will have on the future of data journalism. “Big Data for Media”<sup>60</sup> by Martha L. Stone offers a different look at the issues in the field. This article, which is the result of a two-year study by Stone for the Reuters Institute for the Study of Journalism, offers definitions and parameters for understanding Big Data. While the article focuses on the usage of Big Data by media companies in order to create more targeted news and understand content preferences, the article does briefly address data journalism on page 13- 16. Stone utilized case studies from the British Broadcasting Corporation (BBC) and Cable News Network (CNN) to illustrate the ways in which major media companies are approaching data journalism. In the case of the BBC, the editor for visual journalism explained that the BBC is utilizing data journalism as a way to



tell visual stories. The editor mentioned that creating reader surveys and then visualizing that data has been a big draw for readers. In terms of CNN, the company claims to use Dataminr, a program that uses “an algorithm that identifies, qualifies and alerts clients to key information in real-time”<sup>61</sup> to track important stories. The company also does some data journalism of their own, which again, is a new method for sharing visuals with the online audience. While an intriguing piece, the article focuses more on the use of data rather than the reporting of data.

*Data and Goliath* by Bruce Schneier<sup>62</sup> helps to balance Stone’s work by providing a more critical look at data collection and data journalism. Filled with a wealth of information on the capturing of data, this book offers a complete history of big data. Chapter Three on analyzing data is particularly helpful in describing the results of data mining and how the data captured is used to tell the government or private companies more about individuals. While the author clearly did a great deal of research to complete this book, too often Schneier includes his own opinions and ideas, which means anyone reading the book needs to take his revelations with a grain of salt.

*Big Data* by Viktor Mayer-Schonberger and Kenneth Cukier<sup>63</sup> is similar to *Data and Goliath* in its ability to clearly outline the history of big data, but it takes a more balanced look at data, describing the benefits of data and the negatives, including the loss of privacy that inevitably comes with data mining. It takes a broader look at data, but the background information it provides is quite useful. The authors are certainly well-versed in the topic as Cukier is the data editor at *The Economist* and Mayer-Schonberger is a professor of Internet governance and regulation at Oxford. Nieman Lab also wrote an interesting article on this book and the topic in general when it published an interview with Cukier called “We’re Going to Tell

People How to Interview Databases: The Rise of Data (big and small) in Journalism.”<sup>64</sup> This article provided an interesting look at Cukier’s reasons for writing the book and why he believes that data will change journalism and many other industries forever. In this article, Cukier specifically addressed data journalism, stating that data can help journalists be more accurate and accountable in their stories. That’s a sentiment Waite also echoed. But, while Cukier sees the benefits and challenges of data journalism, he also believes that most journalists are likely opposed to data usage because they don’t like the idea that algorithms can replace their work or that the “quality of their work can be quantifiable.”

In the end, most scholars agree that to secure data-based journalism’s future in the industry, journalists must become educated on the topic and begin to think about stories and ethics through a new lens. More scholarly research is required to provide this understanding, but the desire to learn and the need for data journalism cannot be denied.

## **Conclusion**

Overall, there is an agreement within the industry that data journalism can be a positive part of the news industry if proper standards are applied. After reviewing the material, it’s clear that there are several principles that should be a part of every data journalist’s thought process.

- First and foremost, journalists must take responsibility for the data they are planning to publish. That means going through the data with a fine-toothed comb and giving it, as Waite called it, the “data smell test” to be sure that everything seems legitimate.

- Journalists should also consider where the data is coming from, who put the data together, what were the person's motives, etc.
- During this process, a journalist should also look for missing data and values and try to determine how that affects the end result.
- Journalists should not draw from the tail end of the data by interviewing or focusing on a stereotype or an exception. Instead, they must become embedded in the community and speak with sources that range in age and experiences.
- Take part in systemic reporting that addresses the larger issues and not just the numbers or one individual behind the numbers. In other words, address the societal issues surrounding the data and explore solutions.
- "Unpack" the concepts in story form (e.g. give many examples and explain what it really means to be a member of the data). In other words, put the data into context. Nearly every scholar touted the importance of context in regards to data. Putting numbers into context allows readers to better understand data's meaning; journalists are less likely to insinuate false correlations or lead readers to incorrect conclusions.

In addition to these data specific guidelines more general ethical guidelines relating to harm and transparency should also be applied. A data journalist should always be forthcoming about the source of the data and how it was compiled for the article. It is also worth noting again the importance of a proper education that provides future professionals with the skills needed to analyze data.

Data journalism has shown that it has the ability to provide a wealth of information to the masses on many topics that affect society. In fact, the experts agree that data journalism has the power to bring more voices to the forefront if only proper time and resources are spent on these stories. But, additional studies are needed to truly evaluate the benefits of data journalism and understand the ethical implications of this new reporting methodology. While so much more is needed, it is exciting to think that most of this research has emerged in the last few years and there is certainly more to come. It is also comforting to know that many of the ethical standards against harm and inaccurate reporting are already in place; journalists must simply view them in the context of data journalism and allow for a broader discussion of the issues.

## Bibliography

- 
- <sup>1</sup> Berret, Charles, and Cheryl Phillips. "Teaching Data and Computational Journalism." *Teaching Data and Computational Journalism*. GitBook, Mar. 2016. Web. 15 Mar. 2016: 6
- <sup>2</sup> Quote from author, Brené Brown, 2015
- <sup>3</sup> Mark Coddington (2015) Clarifying Journalism's Quantitative Turn, *Digital Journalism*, 3:3, 331-348, p. 333
- <sup>4</sup> Mark Coddington (2015) Clarifying Journalism's Quantitative Turn, *Digital Journalism*, 3:3, 331-348, p. 335
- <sup>5</sup> Mark Coddington (2015) Clarifying Journalism's Quantitative Turn, *Digital Journalism*, 3:3, 331-348, p. 336
- <sup>6</sup> Mark Coddington (2015) Clarifying Journalism's Quantitative Turn, *Digital Journalism*, 3:3, 331-348, p. 334
- <sup>7</sup> Baack, Stefan: A new style of news reporting: Wikileaks and data driven-journalism. In: *Cyborg Subjects* (2011). p.2
- <sup>8</sup> Rogers, Simon. "The First Guardian Data Journalism: May 5, 1821." *The Guardian*. September 26, 2011. <http://www.theguardian.com/news/datablog/2011/sep/26/data-journalism-guardian>.
- <sup>9</sup> Berret, Charles, and Cheryl Phillips. "Teaching Data and Computational Journalism." *Teaching Data and Computational Journalism*. GitBook, Mar. 2016. Web. 15 Mar. 2016: p. 14
- <sup>10</sup> Berret, Charles, and Cheryl Phillips. "Teaching Data and Computational Journalism." *Teaching Data and Computational Journalism*. GitBook, Mar. 2016. Web. 15 Mar. 2016: p. 10
- <sup>11</sup> Bradshaw, Paul. "How to Be a Data Journalist." *The Guardian*. Guardian News and Media, 01 Oct. 2010. Web. 1 Apr. 2016.
- <sup>12</sup> Baack, Stefan: A new style of news reporting: Wikileaks and data driven-journalism. In: *Cyborg Subjects* (2011). p.3
- <sup>13</sup> "SPJ Code of Ethics | Society of Professional Journalists | Improving and Protecting Journalism since 1909." *SPJ Code of Ethics | Society of Professional Journalists | Improving and Protecting Journalism since 1909*. SPJ, 6 Sept. 2014. Web. 19 Apr. 2016.
- <sup>14</sup> "SPJ Code of Ethics | Society of Professional Journalists | Improving and Protecting Journalism since 1909." *SPJ Code of Ethics | Society of Professional Journalists | Improving and Protecting Journalism since 1909*. SPJ, 6 Sept. 2014. Web. 19 Apr. 2016.
- <sup>15</sup> Interview with Eric Litke, a reporter with the USA Today Network in Wisconsin, July 5, 2016.
- <sup>16</sup> Baack, Stefan: A new style of news reporting: Wikileaks and data driven-journalism. In: *Cyborg Subjects* (2011). p.2
- <sup>17</sup> Litke, Eric. "How Many Sealed Records in Your County?" *Post-Crescent Media*. USA Today Network, 10 Mar. 2016. Web. 19 Apr. 2016. <http://www.postcrescent.com/story/news/investigations/2016/03/10/how-many-sealed-records-your-county/81536180/>
- <sup>18</sup> Litke, Eric. "How Many Sealed Records in Your County?" *Post-Crescent Media*. USA Today Network, 10 Mar. 2016. Web. 19 Apr. 2016. <http://www.postcrescent.com/story/news/investigations/2016/03/10/how-many-sealed-records-your-county/81536180/>
- <sup>19</sup> Litke, Eric. "Heroin overdoses in Wisconsin (2000-2013)?" *Post-Crescent Media*. USA Today Network, Web. 1 Apr. 2016. <http://www.postcrescent.com/story/news/investigations/2015/08/03/heroin-overdoses-wisconsin-2000-2013/31049227/>
- <sup>20</sup> "Why Do We Know so Little about Native Joblessness?" *The Globe and Mail*. The Globe and Mail, 20 Oct. 2014. Web. 10 Apr. 2016. <http://www.theglobeandmail.com/opinion/editorials/why-do-we-know-so-little-about-native-joblessness/article21180812/>
- <sup>21</sup> Interview with Kathleen Culver, the assistant director of the Center for Journalism Ethics at the University of Wisconsin-Madison

---

<sup>22</sup> Fairfield, Joshua, and Hannah Shtein. "Big Data, Big Problems: Emerging Issues in the Ethics of Data Science and Journalism." *Journal of Mass Media Ethics* 29.1 (2014): 38-51. Web.

<sup>23</sup> Interview with Michelle Robinson, a data analyst for Race to Equity in Madison, Wisconsin and a dissertationist in the Sociology department at the University of Wisconsin-Madison, August 1, 2016.

<sup>24</sup> Eberhardt, J. L., & Hetey, R. C. (2014, August 5). Racial Disparities in Incarceration Increase Acceptance of Punitive Policies. Retrieved August 08, 2016, from <http://pss.sagepub.com/content/early/2014/08/05/0956797614540307.abstract>

<sup>25</sup> <http://madison365.com/index.php/about-us/>

<sup>26</sup> Interview with Susan Robinson, Ph.D., associate professor of journalism at the University of Wisconsin-Madison, July 13, 2016

<sup>27</sup> Interview with Matt Waite, a professor in the College of Journalism and Mass Communication at the University of Nebraska Lincoln and developer of Politifact, Fall 2015

<sup>28</sup> Tompkins, Al. "Where The Journal News Went Wrong in Publishing Names, Addresses of Gun Owners." *Poynter*. Poynter, 27 Dec. 2012. Web. 19 Apr. 2016. <http://www.poynter.org/2012/where-the-journal-news-went-wrong-in-publishing-names-addresses-of-gun-owners/199218/>

<sup>29</sup> Tompkins, Al. "Where The Journal News Went Wrong in Publishing Names, Addresses of Gun Owners." *Poynter*. Poynter, 27 Dec. 2012. Web. 19 Apr. 2016. <http://www.poynter.org/2012/where-the-journal-news-went-wrong-in-publishing-names-addresses-of-gun-owners/199218/>

<sup>30</sup> "SPJ Code of Ethics | Society of Professional Journalists | Improving and Protecting Journalism since 1909." *SPJ Code of Ethics | Society of Professional Journalists | Improving and Protecting Journalism since 1909*. SPJ, 6 Sept. 2014. Web. 19 Apr. 2016.

<sup>31</sup> (Astrid Gynnild, 715).

<sup>32</sup> Merja Mahrt & Michael Scharrow (2013) The Value of Big Data in Digital Media Research, *Journal of Broadcasting & Electronic Media*, 57:1, 20-33.p. 21

<sup>33</sup> Merja Mahrt & Michael Scharrow (2013) The Value of Big Data in Digital Media Research, *Journal of Broadcasting & Electronic Media*, 57:1, 20-33. p. 21

<sup>34</sup> Gynnild, A. "Journalism Innovation Leads to Innovation Journalism: The Impact of Computational Exploration on Changing Mindsets." *Journalism* 15.6 (2013): 713-30. Web. p. 713

<sup>35</sup> Interview with Kathleen Culver, the assistant director of the Center for Journalism Ethics at the University of Wisconsin-Madison

<sup>36</sup> Gynnild, A. "Journalism Innovation Leads to Innovation Journalism: The Impact of Computational Exploration on Changing Mindsets." *Journalism* 15.6 (2013): 713-30. Web. p.716

<sup>37</sup> Gynnild, A. "Journalism Innovation Leads to Innovation Journalism: The Impact of Computational Exploration on Changing Mindsets." *Journalism* 15.6 (2013): 713-30. Web. p. 719

<sup>38</sup> Gynnild, A. "Journalism Innovation Leads to Innovation Journalism: The Impact of Computational Exploration on Changing Mindsets." *Journalism* 15.6 (2013): 713-30. Web. p.719

<sup>39</sup> Merja Mahrt & Michael Scharrow (2013) The Value of Big Data in Digital Media Research, *Journal of Broadcasting & Electronic Media*, 57:1, 20-33. p. 26

<sup>40</sup> Gynnild, A. "Journalism Innovation Leads to Innovation Journalism: The Impact of Computational Exploration on Changing Mindsets." *Journalism* 15.6 (2013): 713-30. Web. p. 721

<sup>41</sup> Gynnild, A. "Journalism Innovation Leads to Innovation Journalism: The Impact of Computational Exploration on Changing Mindsets." *Journalism* 15.6 (2013): 713-30. Web. p. 721

<sup>42</sup> "SPJ Code of Ethics | Society of Professional Journalists | Improving and Protecting Journalism since 1909." *SPJ Code of Ethics | Society of Professional Journalists | Improving and Protecting Journalism since 1909*. SPJ, 6 Sept. 2014. Web. 19 Apr. 2016.

<sup>43</sup> Sullivan, Margaret. "Awash in Data, Thirsting for Truth." *The New York Times*. The New York Times, 05 Sept. 2015. Web.

<sup>44</sup> Leonhardt, David, "Death to Journalism." *The New York Times*. Web.

<sup>45</sup> Marr, Bernard. "Can Big Data Algorithms Tell Better Human Stories." *Forbes*. Forbes Magazine, Web.

---

<sup>46</sup> Seth C. Lewis (2015) Journalism In An Era Of Big Data, *Digital Journalism* 3:3, 321-330. p. 326

<sup>47</sup> Seth C. Lewis (2015) Journalism In An Era Of Big Data, *Digital Journalism* 3:3, 321-330. p. 326

<sup>48</sup> Gynnild, A. "Journalism Innovation Leads to Innovation Journalism: The Impact of Computational Exploration on Changing Mindsets." *Journalism* 15.6 (2013): 713-30. Web. p. 727

<sup>49</sup> Berret, Charles, and Cheryl Phillips. "Teaching Data and Computational Journalism." *Teaching Data and Computational Journalism*. GitBook, Mar. 2016. Web. 15 Mar. 201.

<sup>50</sup> Berret, Charles, and Cheryl Phillips. "Teaching Data and Computational Journalism." *Teaching Data and Computational Journalism*. GitBook, Mar. 2016. Web. 15 Mar. 201; p. 7

<sup>51</sup> Berret, Charles, and Cheryl Phillips. "Teaching Data and Computational Journalism." *Teaching Data and Computational Journalism*. GitBook, Mar. 2016. Web. 15 Mar. 201; p. 28

<sup>52</sup> Berret, Charles, and Cheryl Phillips. "Teaching Data and Computational Journalism." *Teaching Data and Computational Journalism*. GitBook, Mar. 2016. Web. 15 Mar. 201; p. 47

<sup>53</sup> Berret, Charles, and Cheryl Phillips. "Teaching Data and Computational Journalism." *Teaching Data and Computational Journalism*. GitBook, Mar. 2016. Web. 15 Mar. 201; p. 47

<sup>54</sup> Anderson, C. W. "Towards a Sociology of Computational and Algorithmic Journalism." *New Media & Society* (2012): Web.

<sup>55</sup> Anderson, C. W. "Towards a Sociology of Computational and Algorithmic Journalism." *New Media & Society* (2012): Web. p. 1012

<sup>56</sup> Anderson, C. W. "Towards a Sociology of Computational and Algorithmic Journalism." *New Media & Society* (2012): Web. p. 1014

<sup>57</sup> Anderson, C. W. "Towards a Sociology of Computational and Algorithmic Journalism." *New Media & Society* (2012): Web. p. 1015

<sup>58</sup> Anderson, C. W. "Towards a Sociology of Computational and Algorithmic Journalism." *New Media & Society* (2012): Web. p. 1016

<sup>59</sup> Stone, M. L. (2014, November). Big Data for Media. Retrieved from <http://reutersinstitute.politics.ox.ac.uk/sites/default/files/Big Data For Media 0.pdf> p. 1

<sup>60</sup> Stone, M. L. (2014, November). Big Data for Media. Retrieved from <http://reutersinstitute.politics.ox.ac.uk/sites/default/files/Big Data For Media 0.pdf>

<sup>61</sup> Stone, M. L. (2014, November). Big Data for Media. Retrieved from <http://reutersinstitute.politics.ox.ac.uk/sites/default/files/Big Data For Media 0.pdf> p. 13

<sup>62</sup> Schneier, Bruce. *Data and Goliath: The Hidden Battles to Capture Your Data and Control Your World*. New York: W. W. Norton, 2015. Print.

<sup>63</sup> Mayer-Schönberger, Viktor, and Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston (MA),: Mariner, 2014. Print.

<sup>64</sup> O'Donovan, Caroline, "We're Going to Tell People How to Interview Databases:" *The Rise of Data (big and small) in Journalism*. Nieman Lab. Web.

## Professional Project Part 2:

©Rebekah McBride, 2016

### **Specialized Journalism: Investigating the Ethics of Data Journalism and Science Writing**

This three credit course for technical school undergraduates will explore the ethical principles of journalism, particularly as they apply to data journalism and science writing. The course is to be held for a total of three hours a week (50 minute course per day).

\*I envision this course being taught at the local technical school where the class sizes are around 20 students.

#### **Students will pick one of these books for their mid-term report:**

- *Geeks Bearing Gifts*, Jeff Jarvis
- *News For All the People*, Juan Gonzalez and Joseph Torres
- *Silent Spring*, Rachel Carson

**Headline of the Week:** Each week students will be asked to write a news story, press release, op-ed or other journalism-related story based on a writing prompt given in class. This will allow students to hone their journalism skills while also compiling a portfolio of material. There will be 10 writing prompts worth a total of 30 points each. The grade will be based on the writing rubric attached here.

#### **WRITING RUBRIC:**

A	B	C	D	F
<ul style="list-style-type: none"><li>• The content is written in the appropriate format for the assignment</li><li>• The content is in line with the prompt</li><li>• The writer uses the correct spelling and punctuation</li><li>• The story flows well</li></ul>	One of these elements is missing	Two of these elements are missing	Three of these elements are missing	The writer did not follow the assignment at all or the writer did not turn in an assignment

**Ethics Powerpoint Project:** During Week 3 the class will be learning about ethical principles. To assist students in applying these concepts, students will be asked to get into groups of 3-4 students and then select an article that discusses an ethical issues (for example: misuse of Photoshop, bias, anonymous sources, etc). Each student will then be asked to write a Times New Roman, 12 pt font, double spaced, 1,000 word paper summarizing the article and discussing how the various philosophical theories could be applied to the situation. For example, if the article discussed the use of anonymous sources, a student might discuss Kant's Categorical Imperative and whether



the decisions made in the article regarding sourcing can be applied as a universal law. As a group students will then be asked to develop a PowerPoint presentation that summarizes the article and their application of the various ethical principles. Students will present their finding in front of the class to facilitate further discussion and learning. The grading for the paper will be based on the writing rubric attached here. The presentation will be based on the presentation rubric below as well as the group feedback form handed out in class.

**PRESENTATION RUBRIC:**

A	B	C	D	F
<ul style="list-style-type: none"> <li>• The presentation covers the areas/topics assigned</li> <li>• The presentation was of appropriate length</li> <li>• The presentation was well organized</li> <li>• The information in the presentation was well researched and correct</li> <li>• The presentation was turned in on time</li> </ul>	One of these elements is missing	Two of these elements are missing	Three of these elements are missing	The writer did not follow the assignment at all or the writer did not turn in an assignment

**Law Project:** After learning about the basic laws that affect journalism, students will be placed into small groups and then assigned an area of law to research. The choices will be:

- ❖ **Basic Look at the First Amendment**
  - Should include: levels of speech (strict scrutiny, intermediate scrutiny), types of public forums and prior restraint
- ❖ **Libel**
  - Should include: definition of the elements of a libel suit (defamation, identification, publication, injury, falsity and fault), also discuss public vs. private
- ❖ **Privacy**
  - Should include: What constitutes an invasion of privacy? Is newsworthiness a defense? False light
- ❖ **Open Record Laws and Freedom of Information Act**
  - History of the law, who can access records? How? Discuss classified records

The expectation is that each group will develop a deep understanding of a particular area of law. After learning about the various court cases and examples of the law in action each group will develop a presentation that they will give in front of the class. The presentation must be approved by me prior to the presentation so I will be meeting with all groups on Week 5, Days 1 & 2 while students are given work time. The PowerPoint will be due at the end of the day (midnight) on Week 5, Day 3. Grading for the PowerPoint can be seen in the presentation rubric. Your grade will also be based on the group feedback form handed out in class.

**Mid-Term Project:** This class is all about unique forms of journalism so students will be encouraged to explore these areas. As such, the mid-term assignment asks each student

to select one of the books from the reading list which includes: *Geeks Bearing Gifts* by Jeff Jarvis, a book for those interested in data journalism, *Silent Spring* by Rachel Carson, a book for those interested in science writing and *News For All the People*, Juan Gonzalez and Joseph Torres, a book for those who wish to learn more about the ethical issues surrounding race and the media. Each student will need to read the book they select and then write a 2,000 word report. The report should include:

- A brief summary of the book
- How 4 of the ethical principles could be applied to certain situations in the book
- What the student learned and how he/she can use the information in his/her future pursuits

In addition to the paper, students who read the same book will be put into small groups and asked to create a 25 minute PowerPoint presentation on their book followed by a class discussion.

The grade for the paper will be based on the writing rubric while the presentation will be based on the presentation rubric and the peer feedback form handed out in class.

**Final Project:** Students will write a 1,500-2,000 word feature story on a topic that relates to ethics and either data journalism or science writing. Students must interview at least three knowledgeable members of the community on their topic of choice and include those interviews in the story. Students must also use information from at least two scholarly articles in the story. Topic ideas might include: the ethics of reporting on minority data, finding unbiased sources for science journalism, struggles local reporters have faced when utilizing the Freedom of Information Act to secure information, etc. After completing the paper students will be asked to share what they learned in a short 5 minute speech/presentation.

The grade will be based on the final project rubric.

**FINAL PROJECT RUBRIC**

A	B	C	D	F
<ul style="list-style-type: none"> <li>• The content is written in the appropriate format for the assignment</li> <li>• The content is in line with the assignment and includes well researched material with proper sourcing</li> <li>• The writer uses the correct spelling and punctuation</li> <li>• The story flows well</li> <li>• The story is of the appropriate length</li> <li>• The student presented the material in an appropriate manner</li> </ul>	One of these elements is missing	Two of these elements are missing	Three of these elements are missing	The writer did not follow the assignment at all or the writer did not turn in an assignment

**Points:**

Participation and Attendance: 10 points per class, 15 weeks, 150 points

Headline of the Week: 30 points per assignment, 10 weeks, 300 points

Ethics Powerpoint: 100 points

Law Project: 100 points

Mid-Term: 150 points

Final Project: 200 points

**Total: 1,000 points**

**Grades:**

**900-1000 points A**

**800-700 points B**

**700-600 points C**

**600-500 points D**

**500 or less F**

---

**Week 1: Journalism 101 (refresher)****Day 1: Introductions**

The learner will understand the syllabus and share a bit about his or her experience and goals. In particular students will be asked to share what they wish to gain from this course and how they plan to apply this knowledge to their future ventures. We will also discuss the “Headline of the Week” assignments, which will allow students to have pieces to take with them when they leave the course.

*HOMEWORK: Headline of the Week #1, Students should write a 300 word press release utilizing information from the writing prompt given in class. This will be due on Day 3.*

**Day 2: Professional Expectations and Protocol**

The learner will gain a clear understanding of the professional expectations and protocol.

The lesson will include a review of interviewing skills, AP Style, contacting sources, how to make a good impression, etc... This information will be presented in a PowerPoint followed by a class discussion.

**Day 3: Professional Expectations and Protocol Continued...**

Students will review how to write a press release and the basics of a story. This information will be presented in a PowerPoint followed by a class discussion.

*Homework: Over the weekend students should read the two articles due for Week 2, Day 1. The final press release draft will also be due on Week 2, Day 1.*

## **Week 2: So, what is specialized journalism?**

### **Day 1: Introduction to Science Writing and Data Journalism**

Prior to the start of class students should read the following articles:

“The Ethics of Journalism Don’t Work for Science” by Jonathan Wolff:

<https://www.theguardian.com/education/2007/jul/03/highereducation.news>

“The Problem with Data Journalism” by Allison Schrager

<http://qz.com/189703/the-problem-with-data-journalism/>

The class period will be spent discussing the two articles and questions that students have after reading the material.

*Students will also be introduced to the books they can select for the mid-term project. By the end of the week each student will need to send me an email stating which book they will be using for their report due in Week 8.*

*HOMEWORK: Students should write The Headline of the Week #2 (News Release) based on the writing prompt outlined in class, due Week 2, Day 3*

### **Day 2: History of Science Writing**

I will lecture on what science writing means. I will also highlight some of the most famous writers and articles throughout science writing.

### **Day 3: History of Data Journalism**

*Headline of the Week 2 Due*

I will lecture on what data journalism means. I will also highlight some of the most famous writers and articles throughout data journalism’s history.

## **Week 3: Ethics, The Basics**

### **Day 1: Philosophers**

The learner will gain a basic understanding of ethical principles and philosophers. During class I will present a PowerPoint on the philosophers and the correlating ethical principals.

Discuss Ethical PowerPoint project, due Week 4, Day 1

*Homework: Groups must select their article for the Ethical PowerPoint presentation on Week 4, Day 1 and submit the choice and group list to me via email by the end of Week 3, Day 2. Also read the articles for Day 2.*

### **Day 2: Ethical Codes**

Students should read the following codes and come to class prepared to discuss.

“The Society of Professional Journalists Code of Ethics”

<http://www.spj.org/ethicscode.asp>

“The New York Times Code of Ethics” [http://www.nytco.com/wp-content/uploads/NYT\\_Ethical\\_Journalism\\_0904-1.pdf](http://www.nytco.com/wp-content/uploads/NYT_Ethical_Journalism_0904-1.pdf)

The National Association of Science Writers Code of Ethics”

<https://www.nasw.org/code-ethics-science-writers>

### **Day 3: Applying the Ethical Principles**

After reviewing the PowerPoints and information from day 1 and 2 students will come to class ready to apply that knowledge. During the class students will be given several journalism-related ethical scenarios. Students will be put into groups and then asked to apply the ethical principals to the scenario. Using “numbered heads together” (Kaegan 1994), one individual from each group will report out on their response and then the class will have a discussion.

*Homework: Work on your powerpoint, due Week 4, Day 1*

### **Week 4: Ethics, the Basics Continued and Intro to Law**

*Headline of the Week #3 : Choose a philosopher and pretend that he/she just announced his finding. Write a short news story about his philosophy and its impact. Use a quote off of the sheet I've provided. Be sure to use the correct quotation punctuation. DUE Week 4, Day 3*

#### **Day 1: Work Time**

Work time on the ethics presentation

#### **Day 2: Presentations**

Come to class with your example of data journalism or science journalism that you selected on Monday as well as a brief PowerPoint explaining what you learned in the article and what ethical principles apply. In class you will give a 10-minute PowerPoint to share with the class so that others can benefit from what you have learned.

#### **Day 3: Presentations**

Presentations will continue

*Headline of the Week# 3 Due*

### **Week 5: Media Laws and FOIA**

#### **Day 1: Introduction to Law**

The learner will gain a basic understanding of the laws surrounding copyright, database rights, data protection, discrimination and hate speech and freedom of information.

At the start of class I will give a brief overview of these areas of law as well as an example court case from each area.

After the presentation students will be placed into small groups and then assigned an area of law to research as a part of the law assignment. The expectation is that each group will develop a deep understanding of a particular area of law. After learning about the various court cases and examples of the law in action each group will develop a presentation that they will give in front of the class. The presentation must be approved by me prior to the presentation so I will be meeting with all groups on Week 5 Day 1 & 2 while students are given work time. The PowerPoint will be due at the end of the day (midnight) on Week 5, Day 3.

*Headline of the Week 4 assigned*

**Day 2: Work Time**

Work Time on the Project and meetings with the instructors

**Day 3: Work Time**

Work time on the project and meetings with the instructor

*Headline of the Week 4 due*

*Homework: all presentation due to me the night before Week 6, Day 1*

**Week 6: Media Laws and FOIA Continued**

**Day 1: Presentations**

Presentations and discussions of the law projects will take place all week.

*Headline of the Week #4 (Op-Ed) assigned*

**Day 2: Presentations**

**Day 3: Presentations**

*Headline of the Week #4 due*

*Homework: Write down one question for the mid-semester clarification session*

**Week 7: Mid-Semester Projects**

**Day 1: Mid-Semester Clarification**

Final wrap up of the law discussion

Students should also come to class with one question regarding the course written on a piece of paper. I will read through the questions and go through the answers with the class. Anything I cannot answer in the moment, I will work on and get back to the class.

### **Day 2: Review of editing marks**

Review via a PowerPoint and work time on mid-term and final assignment

### **Day 3: Editing Review #1**

Students should come to class with a hard copy of one previous Headline of the Week article. Students will then exchange their paper with two other students who will edit the paper using proper editing marks. Students who edit the paper must sign the back so that I know each student has read at least two other student's papers. Students will then be given time to make the edits. A final version of the Headline of the Week will be due Week 8, Day 1 at the start of class. This will count as the Headline of the Week, #5.

*Homework: Finish headline of the Week #5 due Week 8, Day 1 and complete mid-term assignment*

## **Week 8: Presentations and Diversity**

### **Day 1: Presentation of Geeks Bearing Gifts**

All Mid-term PowerPoints and papers due the night before (midnight) Week 8, Day 1. Presentation and discussion of *Geeks Bearing Gifts*

*Headline of the Week #5 due*

### **Day 2: Presentation and discussion of *Silent Spring***

### **Day 3: Presentation and discussion of *News for All the People***

*Headline of the Week #5 due*

## **Week 9: Diversity**

### **Day 1: Diversity introduction**

Does data journalism disproportionately affect minority groups? How to avoid missteps in reporting on minority data including information on sampling methods, the importance of presenting information in the proper context and a look at a few case studies.

Students should read this blog before coming to class:

“The Legacy of the Central Park Jogger “Wilding” Case by Natalie Byfield  
<http://www.huffingtonpost.com/natalie-byfield/the-legacy-of-the-central-park-jogger-wilding-case-by-natalie-byfield.html>

*Headline of the Week #6 (per prompt) assigned*

### **Day 2: Diversity Panel Discussion**

Invite a few speakers to a panel discussion of the issues. This will include individuals from a variety of local papers, including independent and diverse publications such as Madison 365 as well as individuals from organizations such as Race to Equity.

### **Day 3: Panel Wrap-up and Discussion**

Discuss the panel and what journalists can do to avoid stereotypes within a story  
*Headline of the Week #6 due*

## **Week 10: Applying the Ethical Principles to Data Journalism**

### **Day 1: Introduction to Data Journalism Ethics**

The learner will begin to apply the ethical principles to basic research, statistic and data collection methods used in data journalism.

Prior to class students will be asked to read the following stories to gain an understanding of how the data was collected, how it was utilized and what ethical concerns stemmed from those major stories.

STORIES:

1. “The Panama Papers: Here’s What We Know”

[http://www.nytimes.com/2016/04/05/world/panama-papers-explainer.html?\\_r=0](http://www.nytimes.com/2016/04/05/world/panama-papers-explainer.html?_r=0)

2. “Afghan War Logs: what did we learn?”

<https://www.theguardian.com/media/2010/aug/02/afghan-war-logs-wikileaks>

3. “Scales of justice or roulette wheel?”

<http://www.postcrescent.com/story/news/investigations/2015/11/23/judicial-sentencing-varies-wisconsin/76278810/>

Students will create a dyad (two students partner up, student A shares their thoughts for one minute, then student B shares their thoughts for one minute). Following that dyad, the partners will discuss and come to a conclusion about the thoughts they would like to share with the class. Then each group will report out, answering the following questions:

1. What ethical principles do you think the journalists considered when writing about these stories?
2. What should they have considered?



*HOMEWORK: Headline of the Week #7 (per prompt) assigned*

**Day 2: Tools of the Trade**

Begin to learn the tools of the trade, watch Lynda videos on Python and Tableau and then I will present a few example stories that have utilized those tools

**Day 3: Ethics of the Tools**

Discuss the ethical issues that might result from using those tools (Web Scraping, automated story generators, etc)

Students should come to class having read:

“AP's 'robot journalists' are writing their own stories now”

<http://www.theverge.com/2015/1/29/7939067/ap-journalism-automation-robots-financial-reporting>

“Guide to Automated Journalism”

<http://towcenter.org/research/guide-to-automated-journalism/>

*Headline of the Week #7 due*

**Week 11: Ethics of Data Journalism Continued**

**Day 1: Dirty Data**

How to evaluate the strengths and weaknesses of a data base, look at case studies of “dirty data” and discuss

Students should come to class having read:

“Big Data’s Dirty Problem” <http://fortune.com/2014/06/30/big-data-dirty-problem/>

“The Dirty Little Secret Data Journalists Aren’t Telling You”

<https://www.washingtonpost.com/news/wonk/wp/2016/04/11/the-dirty-little-secret-that-data-journalists-arent-telling-you/>

*Headline of the Week #8 (per prompt) assigned*

**Day 2: Big Data**

What is Big Data? An explanation via PowerPoint/lecture

Students should come to class having read:

“What is Big Data?”

<http://www.forbes.com/sites/lisaarthur/2013/08/15/what-is-big-data/#75c8a2a13487>

### **Day 3: Data Journalist Presentation**

Invite a data journalist to class to present

*Headline of the Week 8 due*

## **Week 12: Applying the Ethical Principles to Science Writing**

### **Day 1: Intro to the Ethics of Science Writing**

The learner will begin to apply the ethical principles to basic research, statistic and data collection methods used in science writing.

Prior to class students will be asked to read the stories linked here to gain an understanding of how the scientific information was sourced, how it was utilized and what ethical concerns stemmed from those major stories. I will start the class by presenting on the stories and pointing out some concerns.

Students will then create a dyad (two students partner up, student A shares their thoughts for one minute, then student B shares their thoughts for one minute). Following that dyad, the partners will discuss and come to a conclusion about the thoughts they would like to share with the class. Then each group will report out, answering the following questions:

1. What ethical concerns to you have in regards to each story?
2. What ethical principles do you think the journalists considered when writing these stories?
3. What should they have considered?

STORIES:

“The Parmesan Cheese You Sprinkle on Your Penne Could be Wood”

<http://www.bloomberg.com/news/articles/2016-02-16/the-parmesan-cheese-you-sprinkle-on-your-penne-could-be-wood>

“Why People Are Freaking Out About ‘Wood Pulp’ In Parmesan Cheese”

[http://www.huffingtonpost.com/entry/cheese-wood-pulp-cellulose-parmesan\\_us\\_56c491c2e4b08ffac1271f1e](http://www.huffingtonpost.com/entry/cheese-wood-pulp-cellulose-parmesan_us_56c491c2e4b08ffac1271f1e)

“Ebola hysteria: An epic, epidemic overreaction”

<http://www.cnn.com/2014/10/20/health/ebola-overreaction/>

“Reporters got a lot wrong covering Ebola. We should do better next time.”

<http://www.vox.com/2015/5/12/8587843/ebola-reporting-lessons>

*HOMEWORK: Headline of the Week #9 (per prompt) assigned*

### **Day 2: Sources**

Finding ethical sources and learning to have a critical eye when it comes to the motivations and funding behind your data and research. Presented via PowerPoint.

### **Day 3: Science Writer Presentation**

Invite a science writer to class to present

*Headline of the Week #9 due*

## **Week 13: Ethics of Science Writing Continued and the Ethics of the Internet**

### **Day 1: Science Writing Wrap-Up**

Discuss the science writer presentation from last week and this article:

“That time a bunch of journalists confused an opinion piece for a study”

<http://arstechnica.com/science/2016/07/that-time-a-bunch-of-journalists-confused-an-opinion-piece-for-a-study/>

### **Day 2: Ethics of the Internet as they Apply to Science Writing**

Discuss the ethics of the internet and how they relate to science writing. Students should read the following article before class and listen to the NPR piece:

“The “Nasty Effect:” Online Incivility and Risk Perceptions of Emerging Technologies”

<http://onlinelibrary.wiley.com/doi/10.1111/jcc4.12009/full>

“The 'Nasty Effect': How Comments Color Comprehension”

<http://www.npr.org/2013/03/11/174027294/the-nasty-effect-how-comments-color-comprehension>

### **Day 3: Editing Review #2**

Students should come to class with a hard copy of one previous Headline of the Week article (select from assignments 6-10). Students will then exchange their paper with two other students who will edit the paper using proper editing marks. Students who edit the paper must sign the back so that I know each student has read at least two other student’s papers. Students will then be given time to make the edits. A final version of the Headline of the Week will be due Week 14, Day 1 at the start of class. This will count as the Headline of the Week, #10.

## **Week 14: Final Thoughts and Final Projects**

### **Day 1: Crowd Sourcing**

The Ethics of Crowd Sourcing, via PowerPoint. Students should come to class having read:

“Crowdsourcing done right”

[http://www.cjr.org/data\\_points/crowdsourcing\\_done\\_right.php](http://www.cjr.org/data_points/crowdsourcing_done_right.php)

*Headline of the Week 10 due*

**Day 2: The Ethical Funding of Journalism**

Students should read this article before class and be prepared to discuss:

“Crowdfunded Journalism: A Small but Growing Addition to Publicly Driven Journalism”

<http://www.journalism.org/2016/01/20/crowdfunded-journalism/>

**Day 3: Class Wrap-Up and Time to Work**

Time to work on Final Project and time to ask any final questions about the course

**Week 15: Final Projects**

**Day 1:** Final project due to me the night before Week 15, Day 1.

**Day 1-3:** In alphabetical order students will present their final project PowerPoints

## **Acknowledgements**

I would like to thank my committee, Professor John Bender, Professor Mary Kay Quinlan, Professor Sue Burzynski Bullard and Professor Matt Waite for their significant contributions to this project.

I would also like to thank all of the authors and researchers who are mentioned in this paper as well as those who contributed in other ways, specifically:

Eric Litke, *USA Today* Network (Appleton, Wisconsin)

Michelle Robinson, Race to Equity in Madison, Wisconsin

Kathleen Culver, Director, Center for Journalism Ethics at the University of Wisconsin-Madison

Sue Robinson, assistant professor of journalism at the University of Wisconsin-Madison

John Mackiel, professor of practice, College of Education and Human Sciences, University of Nebraska, Lincoln

Finally, I would like to thank my family for their continued support and for all of the ways they have positively contributed to my life. In particular, I would like to thank my parents and grandparents for showing me the value of hard work and the importance of ethics. I would also like to thank my husband for always believing in me and for supporting each and every one of my dreams.