UNLV | University Libraries
University of Nevada, Las Vegas

UNLV Theses, Dissertations, Professional Papers, and Capstones

August 2016

Reader Engagement with Data Journalism: Comparing the Guardian and Washington Post's Coverage of People Killed by Police

Dan Michalski
University of Nevada, Las Vegas, danm@pokerati.com

Follow this and additional works at: https://digitalscholarship.unlv.edu/thesesdissertations

Part of the Broadcast and Video Studies Commons, and the Journalism Studies Commons

Repository Citation
Michalski, Dan, "Reader Engagement with Data Journalism: Comparing the Guardian and Washington Post's Coverage of People Killed by Police" (2016). *UNLV Theses, Dissertations, Professional Papers, and Capstones*. 2795.
https://digitalscholarship.unlv.edu/thesesdissertations/2795

This Thesis is brought to you for free and open access by Digital Scholarship@UNLV. It has been accepted for inclusion in UNLV Theses, Dissertations, Professional Papers, and Capstones by an authorized administrator of Digital Scholarship@UNLV. For more information, please contact digitalscholarship@unlv.edu.

READER ENGAGEMENT WITH DATA JOURNALISM: COMPARING *THE GUARDIAN* AND *WASHINGTON POST*'S COVERAGE OF PEOPLE KILLED BY POLICE

By

Dan Michalski

Bachelor of Science in Journalism

Northwestern University

1994

A thesis dissertation submitted in partial fulfillment of the requirements for the

Master of Arts – Journalism & Media Studies

Hank Greenspun School of Journalism & Media Studies

Greenspun College of Urban Affairs

The Graduate College

University of Nevada, Las Vegas

August 2016

Thesis Approval

The Graduate College
The University of Nevada, Las Vegas

May 10, 2016

This thesis prepared by

Dan Michalski

entitled

Reader Engagement with Data Journalism: Comparing *The Guardian* and *Washington Post's* Coverage of People Killed by Police

is approved in partial fulfillment of the requirements for the degree of

Master of Arts – Journalism & Media Studies
Hank Greenspun School of Journalism & Media Studies

Julian Kilker, Ph.D.
Examination Committee Chair

Kathryn Hausbeck Korgan, Ph.D.
Graduate College Interim Dean

Stephen Bates, Ph.D.
Examination Committee Member

Greg Borchard, Ph.D.
Examination Committee Member

William Sousa, Ph.D.
Graduate College Faculty Representative

READER ENGAGEMENT WITH DATA JOURNALISM: COMPARING *THE GUARDIAN* AND *WASHINGTON POST*'S COVERAGE OF PEOPLE KILLED BY POLICE

ABSTRACT

The issue of people killed by police has become a focus of current political and social discourse related to criminal justice reform in the United States. Two data journalism projects attempting to track previously missing data have been central to a changing discussion. *The Guardian*'s *The Counted* and *The Washington Post*'s *Investigation: Police Shootings* have each attempted to create a running log of fatalities resulting from law enforcement activities. Such endeavors have added to a collective consciousness about the scope and commonality of deadly police encounters, and has provided empirical reference points for various legislative pushes related to police accountability. These two projects – one from an acknowledged leader in data journalism, the other by a legacy newspaper with a tradition of Pulitzer Prize-winning investigative journalism – presented a unique opportunity to compare and contrast two exemplars of data journalism in an active contemporary media context. This thesis conducted a comparative case study consisting of content analysis built on a framework of Coddington's (2015) typology. Findings showed two different approaches to data journalism, one of which won the Pulitzer Prize. In the end, this thesis proposes future research to consider adding a new dimension to Coddington – *vision of self*, which would assess different data journalism decisions as a binary choice between journalism seen primarily as a public service or hard news.

ACKNOWLEDGMENTS

This thesis was hardly a solo effort. I would like to thank many people for their contributions. First, thesis chair Dr. Julian Kilker for overseeing and guiding this project; he helped me push my bounds to see more with the betterment of research. Likewise, thank you to committee members Stephen Bates, Dr. Gregory Borchard, and Dr. William Sousa for their assistance, expertise, and patience. Additional thanks go to Hank Greenspun School of Journalism and Media Studies Director Dr. Larry Mullen and executive admin Diana Morrison for their help navigating administrative complexities. I would also like to thank my JMS classmates for inspiring me along the way – Jeni, Kaitlin, Azalee, Tiffany, Sara, Brittany, Jason, and Jason will always hold a special place in my internal autonomous organic cardiovascular mechanism. Speaking of hearts, I'd like to express a most special appreciation and admiration for Dr. Tricia Catalino for her support, encouragement, and sacrifice in walking with me through this adventure in academia. And last, I would like to extend a nod posthumously to Scott Chaffin (1960 – 2014), a self-described "web hippie" who labored by my side in endeavors leading up to my graduate school experience; he taught this journalist the ways of ethical database wranglers and so much more from a technologist's point of view, concepts that would emerge as central to this study. Every person mentioned here and others have been with me in this process more than they probably realize. It has been worthwhile struggle speckled with fits of joy and insight every step of the way.

TABLE OF CONTENTS

ABSTRACT iii

ACKNOWLEDGMENTS iv

LIST OF TABLES vii

LIST OF FIGURES viii

CHAPTER ONE: INTRODUCTION 1

What Is Data Journalism? 2

Purpose of Study 3

CHAPTER TWO: BACKGROUND 9

A Brief History of Data Journalism 9

Social Science Precision 15

Follow the Data 17

CHAPTER THREE: LITERATURE REVIEW 22

Defining Data Journalism for Research 22

Data in the Newsroom 28

Calls for Content Analysis 36

Summary of Literature 38

CHAPTER FOUR: METHODOLOGY 39

Comparative Case Study 39

Case 1: *The Counted* 44

Case 2: *Police Shootings* 47

Content Analysis 48

Research Steps 52

Summary of Methods 53

CHAPTER FIVE: RESULTS 55

1. Professional Orientation 58

2. Openness 61

3. Epistemology 65

4. Vision of Public 71

Additional Findings 82

2016 Changes 83

Summary of Findings 86

CHAPTER SIX: DISCUSSION 89

Conclusion 96

APPENDIX A: URLs Analyzed for Case 1, *The Counted* 98

APPENDIX B: URLs Analyzed for Case 2, Police Shootings 99

REFERENCES 100

CURRICULUM VITAE 111

LIST OF TABLES

Table Page

3.1 Eight traits of typical data-driven journalism 24

4.1 Two cases for data journalism 42

4.2 Coding sheet 51

5.1 Coding sheet operationalized for Coddington's typology 57

LIST OF FIGURES

Figure		Page
2.1	First issue of *The Manchester Guardian* (1821)	12
2.2	"The Mileage of Congress" (1848)	13
3.1	Coddington's typology	30
4.1	Comparative case study	40
4.2	Online traffic and audience growth	43
4.3	Providers of original source material	43
4.4	Pages for balanced case analysis	46
4.5	Human data frame analysis	48
5.1	Coddington comparisons	58
5.2	Title bar for *The Counted* articles	60
5.3	Reader control for private engagements with data	62
5.4	Data sharing enabled or constrained	64
5.5	Two types of paywalls	65
5.6	Narrative control	67
5.7	Default settings for data totals sorted by race	68
5.8	Narrative framing of race	69
5.9	Personified data in active and passive news environments	70
5.10	Comment-section disparities	73
5.11	Two-way conversation of Facebook	75
5.12	*The Counted*'s Facebook community	78
5.13	Personifying media personnel in the reader's personal space	79

5.14 *The Washington Post* Investigations team as a newspaper on Facebook 80

5.15 Conscious 2016 changes 81

5.16 Visible shifts in *Fatal Force* redesign 86

5.17 Synthesizing results with Coddington 87

6.1 Judges list for *The Washington Post*'s Pulitzer category 91

6.2 Videos and activism with police body cams and UGC 94

CHAPTER ONE: INTRODUCTION

"It's embarrassing and ridiculous," FBI Director James Comey said in October 2015 at a meeting with about 100 law enforcement officials and politicos who gathered to hear him speak about challenges to twenty-first century law-enforcement. "It is unacceptable that *The Washington Post* and *The Guardian* newspaper from the UK[1] are becoming the lead source of information about violent encounters between police and civilians. That is not good for anybody" (Davis & Lowery, 2015).

Comey was referring to two journalistic endeavors that have highlighted the federal government's data collection inadequacies surrounding people killed by police. Prior to 2015, no record of any valid measure existed to chronicle people in the US killed by police. A voluntary system of reporting applied to the country's 18,000 law enforcement agencies, an option that netted a response rate less than 6% – and those reports were filed with no standards, protocols or training to support any consistency in data reporting or give any validity to broader statistics. For the media, this lacuna in the public data became apparent in the wake of Michael Brown's death in Ferguson, Missouri, followed by a series of video recorded police killings to emerge in quick succession – Eric Harris, Tamir Rice, Freddie Gray, Walter Scott, to name a few – often telling a story different from what official reports conveyed. But as news organizations sought viable statistics on these matters and found none, journalistic inquiry led to these two efforts to fill the statistical void.

[1] This is a common misconception about operations that *The Guardian* is regularly trying to correct. The New York-based *Guardian US* office is staffed by Americans as well employees of other nationalities.

On June 1, 2015, *The Guardian* launched *The Counted*, a production of *The Guardian US*, the British daily's online-only American edition, headquartered in New York City. That same month *The Washington Post* launched *Investigation: Police Shootings*[2] – starting with a series of news articles eventually followed by a release of their own data collection. Both *The Guardian* and *Post* embarked on major efforts to uncover answers in the data the government could not provide. Some questioned the validity of these unofficial counts, particularly when on first glance their totals did not seem to agree. But in fact they gave credibility to the numbers, confirming the independence of their approaches. Both were telling the same story even if they were going about it a different way. In the end, for 2015, *The Guardian* chronicled 1,145 people killed by police – including deaths by gunfire, Tasers, vehicles, and other causes in custody – while the *Post* logged 990 people that police shot and killed only. Both numbers were more than double what the FBI had to suggest for official statistics (Gourarie, 2015).

By culling information from publicly available data on the Internet, social media, and tips from readers, then running collected information through journalistic verification processes (Silverman, 2015), these projects did more than just keep a log of deaths. By looking at the deceased individually and as a collective unit of shared experience, two publications showed that data journalism could use data to reveal bigger and deeper stories that have since brought criminal justice reform to new levels of salience in public discourse.

What Is Data Journalism?

Data journalism is more than infographics, top ten lists, and charticles -- though graphs, maps, and tables are often instrumental in effective displays of data. In *The Art and Science of*

[2] Since rebranded (in April 2016) as *Fatal Force*.

Data-Driven Journalism (Howard, 2014, p. 5), a guidebook produced by Columbia University's Tow Center that attempted to bridge scholarly explanation and instructional professional practices, author Alex Howard attempted to simplify. "Data journalism is telling stories with numbers, or finding stories in them." Brian Boyer of the *Chicago Tribune* describes data journalism as a matter of research methods, similar to photojournalism. "'Data journalism' only differs from 'words journalism' in that we use a different kit," he explains. "We all sniff out, report, and relate stories for a living. It's like 'photojournalism'; just swap the camera for a laptop" (Gray et al., 2013, p. 6). Aron Pilhofer of the *New York Times* concurred:

> Data journalism is an umbrella term that to my mind, encompasses an ever-growing set of tools, techniques, and approaches to storytelling. It can include everything from traditional computer-assisted reporting (using data as a 'source') to the most cutting-edge data visualization and news applications. The unifying goal is a journalistic one: providing information and analysis to help inform us all about important issues of the day. (Gray et al., 2013, p. 6)

Purpose of Study

Researchers are trying to understand data journalism as practitioners are actively advancing its reach without epistemological direction. At publications across North America and Europe, data journalism has taken on new levels of prominence, and as such is reshaping newsrooms. Sites such as *Pro Publica* have sprouted up to serve as a public source of data analysis, while other new, non-profit journalistic entities such as *Vox* and *Texas Tribune* have made a commitment to data journalism core to their journalistic enterprise (Gray et al., 2013). In Las Vegas, the *Review-Journal* recently committed to building a collaborative investigative

team, where one of the six people they are looking to hire will hold a new title of "data editor" (*Las Vegas Review-Journal*, 2015).

Research interests in data journalism spread across many disciplines, including sociology, social psychology, criminal justice, political science, public policy, and media and journalism studies. Journalism schools also have a great interest in expanding their knowledge of data journalism – particularly the ability to teach data journalism. A deficiency in incorporating data into journalism curricula has long been noted (Johnson, 1992), but recently the importance of integrating data journalism in journalism schools has taken on a new urgency (Krueger, 2014; Yu, 2014; Berret & Phillips, 2016) making it one of the most valued skills in present day media environs (Kang, 2015). Learning the techniques is not necessarily difficult, as many textbooks produced by practitioners are available online, in multiple languages, and for free by download (Gray, Bounegru, & Chambers, 2012; Silverman 2015). But beyond the technical skills came their intentional use. This research thus sought to understand better the dynamic relationships between journalists, readers and a particular set of data.

This study looked at two exemplars of contemporary data journalism practices. When both *The Guardian* and *Washington Post* began seeking data to back up stories being written about Ferguson, they discovered how little data existed and how poorly collected what it was. Both storied journalism outlets took it upon themselves to find the data, which presented a unique opportunity to explore the dynamics of data journalism relationships.

Definitions

Chapter Two explains with more specificity what is meant by "data journalism" (DJ) and "data-driven journalism" (DDJ) as forms of journalism in digital spaces. In this thesis, *data*

journalism refers to data use practices by established news operations to tell a story (Franklin, 2014). The term *media* refers to journalists and the tools they use, and "journalists" refer to agents of news operations in online spaces across multiple platforms. This thesis refers to "readers" and "users" almost interchangeably. Much of this thesis reflects the convergence of "technologists" – programmers, coders, designers, software engineers, among others – with journalists, and the cultural tensions between them, as one of the most common rifts – with users being the preferred term of technologists and readers the preferred term of journalists. Use of either term in this study is referring to people who engage with a digital news product. The term "audience" is more complex to define, as its meaning has evolved over time (Lewis & Westlund, 2015). Here, audience speaks to collective group of people acting as a unit. Users/readers make up an audience; while "public" is a broader term, referring to people who are plausibly exposed, but not making a conscious effort to consume media before them. With "newspaper," the reference is to digital pages produced and disseminated by Web operations of a traditional print publication though now operating increasingly online.

Research Plan

This thesis offers qualitative analysis of a generally quantitative phenomenon by exploring data journalism as conducted by two respected media outlets – one an acknowledged leader in data journalism (Stray, 2010; MIT Open Documentary Lab, 2015; Onouha, Pinder, & Schaffer, 2015), the other self-identifying pillar of investigative journalism (Kovach & Rosensteil, 2013). Two research questions helped guide me as I delved into data journalism as a method of practicing investigative journalism:

RQ_1 What mechanisms exist in *The Counted* and *Police Shootings* to potentially mediate engagement between journalists, readers, and data?

RQ_2 How do *The Guardian* and *Washington Post* use these mechanisms to enable or constrain reader participation? (And why?)

My overarching goal was to explore the dynamics at the convergence of data and journalism in contemporary society, particularly with today's data collection and processing abilities in some capacity influencing the role of data in news productions.

For improved scholarly analysis of data journalism as a sociotechnical construction, qualitative study becomes beneficial and essential for pattern matching and explanation building. In this thesis, I investigated data journalism constructs in the context of recurring current events in news stories related to police accountability and civilians they killed. If theoretical models are to adapt and evolve, it is important to understand how principles from investigative journalism's past apply to data journalism practices today. After reviewing the literature, I conducted content analysis that compared and contrasted *The Guardian*'s *The Counted* and *The Washington Post*'s *Investigation: Police Shootings*.

Police brutality has long been an issue of sociopolitical discourse (*New York Times*, 1973; Flynn, 2003) but the availability of data has changed the context with which these discussions now occur. An old adage in journalism contended that good journalists were there to report the news not make the news. But that paradigm was challenged long ago – when the press started using, and then conducting their own scientifically valid polls. Now data collection in and of itself can become news. There are multiple ways to practice data journalism, and the co-

existence of the *The Counted* and *Police Shootings* allowed me to compare very similar offerings from two different media institutions.

In March 2015, FBI Director Comey was defending the government's "sampling-based" statistics on police violence (McCarthy, 2015). In June, days after *The Counted* came out, two Democratic US senators introduced new legislation to mandate better federal tracking. The FBI still resisted demands to change their measures (Swaine & Laughland, 2015). By October 2015, a point when *The Counted* had grown from 480 to nearly 900, and *Police Shootings* had gone up from 440 to more than 800, Comey was acknowledging plans to change – with the US government ready to use open-source sourcing similar to the data journalism projects (Laughland, Lartey, & McCarthy, 2015). And in December 2015, they declared the government's intent to collect data similar to what both publications were collecting, starting in 2017 (Kindy, 2015). In the meantime, both continued their counts beyond 2015 into 2016.

Journalistically, these have been award-winning endeavors. Both data journalism productions have been on a regular short-list of awards for their work in 2015. *The Washington Post* won a Pulitzer Prize (for National Reporting) for *Police Shootings*. In addition to winning journalism's highest honor, the *Post* also won a 2015 Polk award, and both *Police Shootings* and *The Counted* were finalists for a 2015 Goldsmith Prize. The award honors investigative journalism that has had an "impact on public policy in the United States at the national, regional or local level" — with a prize of $10,000 for five of the six finalists, and $25,000 for the winner. For journalism conducted in 2015 (and a prize awarded in 2016), both *The Counted* and *Investigation: Police Shootings* were among the finalists. Looking at them together greatly enhanced the validity of observations looking at either. Comparing the two exposed their different design elements and data processes – quite relevant to early efforts in academic

research around new theoretical constructs in journalism and media studies. The co-existence of these two journalism productions presented a unique opportunity to identify and analyze two different approaches to data journalism – one from a media entity fully embracing the values of progressive technology, the other from a media entity enmeshed in a rich and successful tradition of "shoe leather" investigative journalism. This presented a fascinating means for exploration of the relationship between engaged readers and data as presented by journalists (McCombs et al., 1981; Franklin, 2014; Parisie, 2015).

The following chapters first cover relevant literature to this study – looking at the roots of data journalism and how the injection of ideals drawn from technology and academia have shaped journalistic practices, and then literature about conducting content analysis in the various Web spaces that made up my cases. Upon explaining in further detail how I conducted my comparative case study with content analysis, my findings present two distinct approaches to data journalism that have less to do with statistical uses of data, and more with its collection, presentation, and most important use as a reader engagement device. When considering results in conjunction with history of data usage in journalistic practice, an interpretation of these results came with implications that could shape future journalistic engagements with public and government data. Upon completion of this thesis, readers should better understand how instrumental components of data journalism mediate relationships between journalists, readers, and data. These include Web design features and journalistic processes alike – and include concepts such as crowdsourcing, social media engagement, framing, and gatekeeping. With advanced knowledge of these relationships, researchers, editors, educators, policy makers and others can better assess the real possibilities of data-driven journalism world before them.

CHAPTER TWO: BACKGROUND

This chapter examines the history of data journalism and argues that use of Web technologies by media entities have the potential to influence reader engagement with data, and thus with the story journalists are trying to tell, as it related to people in the US killed by police. In this chapter, I present a history of data and journalism together, from before the two terms were used in partnership. That provided a context for better understanding a theoretical mish-mash trying to attach structure to concepts as subjects of scholarly research, and reminded me that some new stories were not without precedent.

A Brief History of Data Journalism

Presenting visualized data to the masses stretches back centuries. In the 1700s, statisticians worked with engravers in to provide graphic pamphlets to make political arguments to be discussed in the town squares and public houses. By the end of the eighteenth century, the literate and even not so literate were beginning to understand line graphs, bar charts, and eventually pie charts of William Playfair, who introduced these concepts in his publications of 1786 and 1801 (Tufte, 2001). Before newspapers could efficiently incorporate graphics into their publications, they learned the power of telling stories with data by using agate type that visually stood out from surrounding text. *The Manchester Guardian* (now known simply as *The Guardian*) launched in 1821 with an example of this concept in their premiere issue, which featured a "to the editor" submission on the front page with a table comparing student enrollment and school spending across two British school districts as a way of exposing the reality behind academic privilege (Appendix C). Even then, the editor knew the potential for these data to spark

controversy, and probably saw this as a layer of protection from any backlash (S. Rogers, personal communications, March 22, 2016).

In the United States in 1848, Horace Greeley showed what was possible by incorporating fata into journalism. Greeley was publisher of the *New York Tribune* (and one of the founders of the Republican Party). Klein (2015) tells the story in *Pro Publica* of how he successfully parlayed his stature as an editor and publisher to build a public persona that allowed him to serve a three-month stint in Congress by filling a vacated seat that belonged to the Whig Party. During his tenure, Greeley made it his mission to challenge a Congressional reimbursement policy that he believed was wasting taxpayer money. The policy reimbursed legislators 40 cents per mile for travel back to their home districts. But it was a throwback to days before railways and steam ships had made travel less arduous, and to Greeley the policy seemed ripe for abuse should any member in a district far from D.C. choose to take a longer route home and get paid extra for it. Being in Congress gave Greeley access to government files on these reimbursements, data he requested as a measure of Congressional business. He then used a book of postal routes (gained from one of his reporters who used to work for the post office) to calculate the shortest path from each representative's district to the Capitol. Greeley's reporter was able to compare standard distances with reported reimbursements for each, which led to the *Tribune*'s publishing "The Mileage of Congress" as a front-page, three-column story. More than half the content consisted of a table listing each member by name with mileage money received vs. mileage the postal route would have granted him, and the difference in cost (*New York Tribune*, 1848).
The overall amount was not huge – about $60,000, or $2 million in 2015 dollars. Abraham Lincoln, serving his sole Congressional term, was an offender, overbilling the federal government $677 (about $20,000 in 2015). Greeley and the *Tribune* tried to frame the story not

as a scandal, but simply as an impetus for good governance. “Let no man jump at the conclusion that this excess has been charged and received contrary to law,” wrote Greeley in the text accompanying the *Tribune*’s chart. “The fact is otherwise. The members are all honorable men — if any irrelevant infidel should doubt it, we can silence him by referring to the prefix to their names in the newspapers” (Klein, 2015).

Yet that was not how the story was received. Despite the disclaimers, as clamor spread, Congressional members and their constituencies began claiming vociferously that the numbers were lies – error-riddled falsehoods fueled by a political agenda and rife with factual error, insinuation, and typos. Klein (2015) cited a heated floor debate chronicled in *The Congressional Globe* where members lined up to air their grievances with Greeley’s work. An example provided came from Illinois Rep. Thomas Turner, a Democrat with an overage of $998.40, who said Greeley had:

> ... either been actuated by the low, groveling, base, and malignant desire to represent the Congress of the nation in a false and unenviable light before the country and the world, or that he had been actuated by motives still more base — by the desire of acquiring an ephemeral notoriety, by blazoning forth to the world what the writer attempted to show was fraud. The whole article abounded in gross errors and willfully false statements, and was evidently prompted by motives as base, unprincipled, and corrupt as ever actuated an individual in wielding his pen for the public press. (Klein, 2015)

TO THE EDITOR OF THE MANCHESTER GUARDIAN.

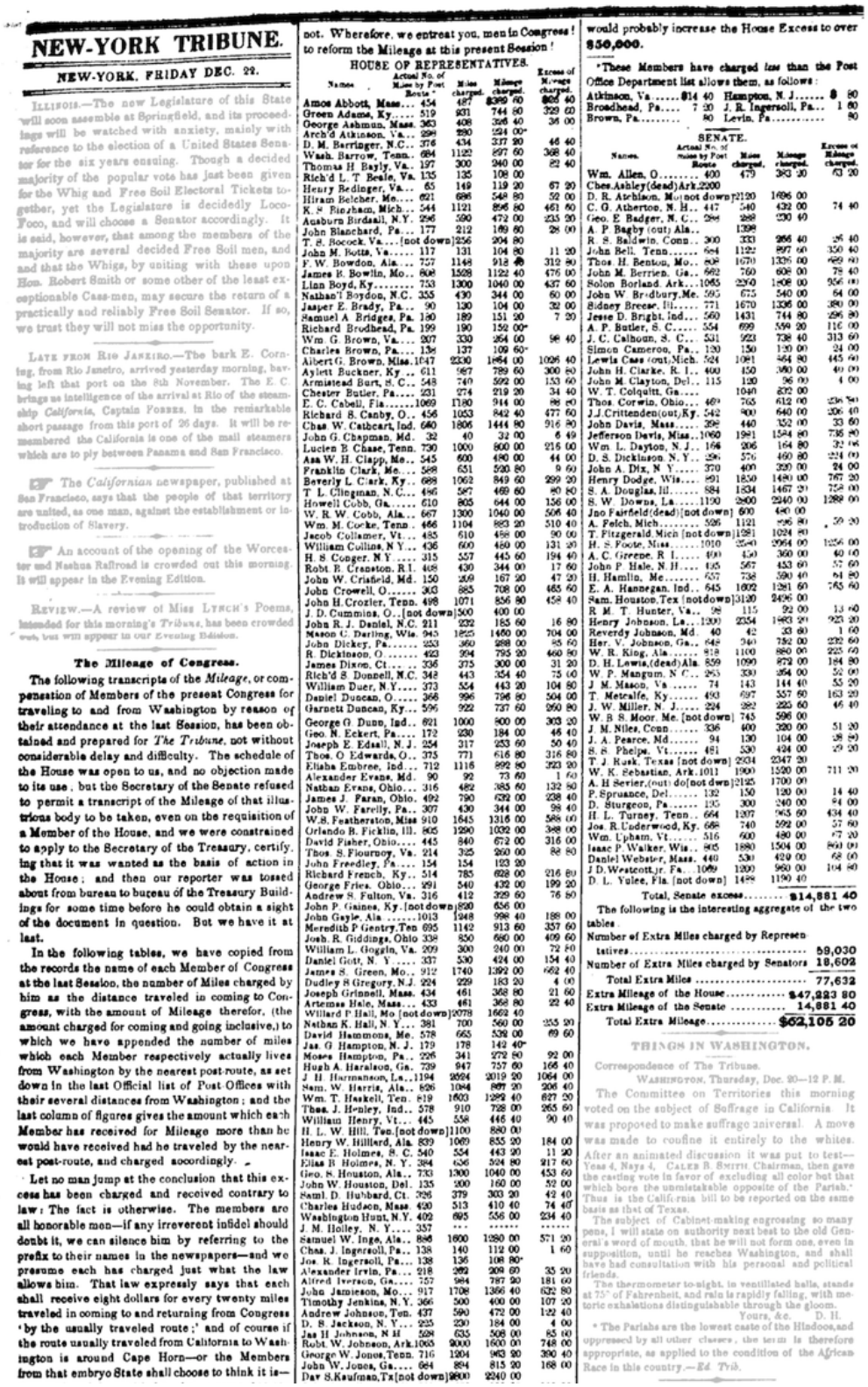
The annexed Report of the state of Charitable Education in this Town and the immediate Vicinity I communicate, in the expectation that you will consider it of sufficient importance and interest to entitle it to publication. No document of a similar nature has yet been laid before the public. At all times, such information as it contains is valuable; because, without knowing the extent to which education, and particularly the education of the labouring classes, prevails, the best opinions which can be formed of the condition and future progress of society must be necessarily incorrect. But, independent of general considerations, such a document seemed called for at the present juncture, by the legislative measures with which Mr. Brougham is threatening the country, and which appear to me to be a practical consequence of the ignorance which exists on this subject. The facts on which Mr. Brougham makes out the necessity for his Education Bills, are, as he is reported to say, "the numerical results of an inquiry, "carried on for years among the persons *best able* to report the state of education in each village and "hamlet of the island. Those persons have, by the *most minute details* of matters within their own "knowledge, enabled him to state the *exact numbers of schools actually existing, and the number of* "*children actually taught in each.*" The persons whose qualification and knowledge are thus extravagantly and (taken in connection with the real facts of the case) ridiculously praised, are the Clergymen of the Church of England, who make up the reports that compose the Education Digest. In the district to which the following report refers, I find the statements are attributed to four respectable Clergymen, who are, I believe, as competent as any of their brethren for the performance of such a task. Them it would be severe to censure for the imperfection of their returns; they were bound by the orders of the Education Committee to communicate only what they knew; and I am not aware that they were obliged, *ex officio*, to know even so much as they have communicated. I do not, therefore, hold them accountable for the great inaccuracy which appears in the Digest; but it may be reasonably objected to Mr. Brougham, that he ought not to have pledged himself quite so strongly in support of his informants, as "persons *best able* to report the state of education," and of their information, as containing "*the exact numbers of schools actually existing, and the numbers of children actually taught in each.*" That the public may judge, I have extracted the official returns that correspond with the district included in the following Report.—Vol. 1, p. 444. LANCASHIRE; TABLE.

Parishes or Chapelries.			Endowments.			Unendowed Day Schools.		Unendowed. Sunday Schools.	
Names.	Population in 1811.	Poor in 1815.	No.	Children.	Revenue.	No.	Children.	No.	Children.
Manchester	85,828	2899	1	155	£1800	3	2200	3	8000
			1	80	2000	4	83		
			1	50	200				
			1	10	100				
			1	50	40				
Ardwick ...	5344	129	1	none	8			1	200
Salford	19,114	791				4	600	8	700

From this it appears, that there are only seven Sunday Schools, containing 8900 scholars, in Manchester, Salford, and Ardwick. Such is the *statement* given by four resident Clergymen of the Church of England, who are, as Mr. Brougham thinks, "*best able* to report the state of education;" while the *fact* is, that the number of the CHURCH Sunday Schools alone, is 19; and of the Dissenters, 46; making together, 65; and affording gratuitous education to 23,000 children. If to this number be added the day-scholars and boarders of the public and private schools, it would appear, that, with respect to this town and neighbourhood at least, there is no necessity for any legislative proceeding whatever; and I am persuaded, that there is as little necessity in any part of the manufacturing districts of the country. In the agricultural districts, the case may be different. With respect to them, some legal provision might be useful; but Mr. Brougham's bills are too obnoxious to the liberal spirit of the times, to produce any thing but evil in any part of the kingdom. As *disqualifying* acts, they ought not to pass into laws; and they will not, unless the Dissenters negligently disregard the protection of their own rights. My object, however, is not to discuss the provisions of Mr. Brougham's bills; but merely to call attention to the errors which exist in the Education Digest, and to express an earnest wish, that similar reports may be made up for other large towns, to the publication of which, I trust, your columns would be open. N. H.

DAY SCHOOLS.—*Establishment.*	*Boys*	*Girls.*	*Total.*	*Ann. Exp.*	*Remarks.*
Grammar School	155	....	155	1800	..... (Endowments.)
Blue Coat ditto	80	....	80	2000	Taught, clothed and boarded. (Endowments.)
Green Coat ditto	50	....	50	200	Taught and clothed. (Endowments.)
Collegiate Church ditto	....	50	50	40	And offertory money: do. do. (Endowments.)
Strangeways ditto	10	....	10	100	..... (Endowments.)
St. Mary's ditto	12	12	24	40	(Suppose)—Taught and clothed. Funds arising from Sacramental Offerings.
St. John's ditto	9	....	9	40	(Suppose)—Expences raised by voluntary Subscription.
St. Paul's ditto	20	....	20		
Ladies' Jubilee	....	30	30	250	Taught, clothed and boarded. by voluntary Subscription.
Back King-street	21	....	21	40	(Suppose)—Taught and partly clothed. This School is supported by the benevolence of a single individual.
NATIONAL SCHOOLS, Granby-row	194	119	313	600	Voluntary Subscription, and Collections at Churches.
Bolton-street, Salford	300	170	470		
	851	381	1232	£5410	

Figure 2.1. First issue of *The Manchester Guardian* (1821). The publisher used visually compelling white space to draw attention to a report about school spending, which was presented as a letter to the editor.



NEW-YORK TRIBUNE.

NEW-YORK, FRIDAY DEC. 22.

ILLINOIS.—The new Legislature of this State will soon assemble at Springfield, and its proceedings will be watched with anxiety, mainly with reference to the election of a United States Senator for the six years ensuing. Though a decided majority of the popular vote has just been given for the Whig and Free Soil Electoral Tickets together, yet the Legislature is decidedly Loco-Foco, and will choose a Senator accordingly. It is said, however, that among the members of the majority are several decided Free Soil men, and that the Whigs, by uniting with these upon Hon. Robert Smith or some other of the least exceptionable Cass-men, may secure the return of a practically and reliably Free Soil Senator. If so, we trust they will not miss the opportunity.

LATE FROM RIO JANEIRO.—The bark E. Corning, from Rio Janeiro, arrived yesterday morning, having left that port on the 8th November. The E. C. brings us intelligence of the arrival at Rio of the steamship *California*, Captain FORBES, in the remarkable short passage from this port of 26 days. It will be remembered the California is one of the mail steamers which are to ply between Panama and San Francisco.

☞ The *Californian* newspaper, published at San Francisco, says that the people of that territory are united, as one man, against the establishment or introduction of Slavery.

☞ An account of the opening of the Worcester and Nashua Railroad is crowded out this morning. It will appear in the Evening Edition.

REVIEW.—A review of Miss LYNCH'S Poems, intended for this morning's *Tribune*, has been crowded out, but will appear in our Evening Edition.

The Mileage of Congress.

The following transcripts of the *Mileage*, or compensation of Members of the present Congress for traveling to and from Washington by reason of their attendance at the last Session, has been obtained and prepared for *The Tribune*, not without considerable delay and difficulty. The schedule of the House was open to us, and no objection made to its use; but the Secretary of the Senate refused to permit a transcript of the Mileage of that illustrious body to be taken, even on the requisition of a Member of the House, and we were constrained to apply to the Secretary of the Treasury, certifying that it was wanted as the basis of action in the House; and then our reporter was tossed about from bureau to bureau of the Treasury Buildings for some time before he could obtain a sight of the document in question. But we have it at last.

In the following tables, we have copied from the records the name of each Member of Congress at the last Session, the number of Miles charged by him as the distance traveled in coming to Congress, with the amount of Mileage therefor, (the amount charged for coming and going inclusive,) to which we have appended the number of miles which each Member respectively actually lives from Washington by the nearest post-route, as set down in the last Official list of Post-Offices with their several distances from Washington; and the last column of figures gives the amount which each Member has received for Mileage more than he would have received had he traveled by the nearest post-route, and charged accordingly.

Let no man jump at the conclusion that this excess has been charged and received contrary to law: The fact is otherwise. The members are all honorable men—if any irreverent infidel should doubt it, we can silence him by referring to the prefix to their names in the newspapers—and we presume each has charged just what the law allows him. That law expressly says that each shall receive eight dollars for every twenty miles traveled in coming to and returning from Congress 'by the usually traveled route;' and of course if the route usually traveled from California to Washington is around Cape Horn—or the Members from that embryo State shall choose to think it is—not. Wherefore, we entreat you, men in Congress! to reform the Mileage at this present Session!

HOUSE OF REPRESENTATIVES.

Names.	Actual No. of Miles by Post Route.	Miles charged.	Mileage charged.	Excess of Mileage charged.
Amos Abbott, Mass...	454	487	$389 60	$60 40
Green Adams, Ky....	519	931	744 80	329 60
George Ashmun, Mass.	363	408	326 40	36 00
Arch'd Atkinson, Va...	298	280	224 00*	
D. M. Barringer, N.C..	376	434	337 20	46 40
Wash. Barrow, Tenn..	684	1122	897 60	368 40
Thomas H Bayly, Va..	197	300	240 00	22 40
Rich'd L. T Beale, Va.	135	135	108 00	
Henry Bedinger, Va...	65	149	119 20	67 20
Hiram Belcher, Me....	621	686	548 80	52 00
Kingsley Bingham, Mich...	544	1121	896 80	461 60
Auburn Birdsall, N.Y.	296	590	472 00	235 20
John Blanchard, Pa...	177	212	169 60	28 00
T. S. Bocock, Va....	[not down]	256	204 80	
John M. Botts, Va.....	117	131	104 80	11 20
F. W. Bowdon, Ala....	757	1148	918 40	312 80
James R. Bowlin, Mo..	808	1528	1122 40	476 00
Linn Boyd, Ky.....	753	1300	1040 00	437 60
Nathan'l Boydon, N.C.	355	430	344 00	60 00
Jasper E. Brady, Pa...	90	130	104 00	32 00
Samuel A. Bridges, Pa.	180	189	151 20	7 20
Richard Brodhead, Pa.	199	190	152 00*	
Wm. G. Brown, Va....	207	330	264 00	98 40
Charles Brown, Pa....	136	137	109 60*	
Albert G. Brown, Miss.	1647	2330	1864 00	1026 40
Aylett Buckner, Ky ..	611	987	789 60	300 80
Armistead Burt, S. C..	548	740	592 00	153 60
Chester Butler, Pa....	231	274	219 20	34 40
E. C. Cabell, Fla.....	1069	1180	944 00	88 80
Richard S. Canby, O..	456	1053	842 40	477 60
Chas. W. Cathcart, Ind.	660	1806	1444 80	916 80
John G. Chapman, Md.	32	40	32 00	6 40
Lucien B Chase, Tenn.	720	1000	800 00	216 00
Asa W. H. Clapp, Me..	545	600	480 00	44 00
Franklin Clark, Me....	568	651	520 80	9 60
Beverly L. Clark, Ky..	688	1062	849 60	299 20
T. L. Clingman, N.C...	486	587	469 60	80 80
Howell Cobb, Ga.....	610	805	644 00	156 00
W. R. W. Cobb, Ala...	667	1300	1040 00	506 40
Wm. M. Cocke, Tenn .	466	1104	883 20	510 40
Jacob Collamer, Vt...	485	610	488 00	90 00
William Collins, N Y...	436	600	480 00	131 20
H. S. Conger, N Y	315	557	445 60	194 40
Robt B. Cranston, R.I.	408	430	344 00	17 60
John W. Crisfield, Md.	150	209	167 20	47 20
John Crowell, O.....	303	885	708 00	465 60
John H. Crozier, Tenn.	498	1071	856 80	458 40
J. D. Cummins, O..[not down]		500	400 00	
John R. J. Daniel, N.C.	211	232	185 60	16 80
Mason C. Darling, Wis.	945	1825	1460 00	704 00
John Dickey, Pa.....	253	360	288 00	85 60
R. Dickinson, O	423	994	795 20	460 80
James Dixon, Ct.... ..	336	375	300 00	31 20
Rich'd S. Donnell, N.C.	342	443	354 40	75 00
William Duer, N.Y....	373	554	443 20	104 80
Daniel Duncan, O.....	366	996	796 80	504 00
Garnett Duncan, Ky...	596	922	737 60	260 80
George G. Dunn, Ind..	621	1000	800 00	303 20
Geo. N. Eckert, Pa....	172	230	184 00	46 40
Joseph E. Edsall, N. J.	254	317	253 60	50 40
Thos. O. Edwards, O..	375	771	616 80	316 80
Elisha Embree, Ind...	712	1116	892 80	323 20
Alexander Evans, Md.	90	92	73 60	1 60
Nathan Evans, Ohio...	316	482	385 60	132 80
James J. Faran, Ohio.	492	790	632 00	238 40
John W. Farelly, Pa..	307	430	344 00	98 40
W. S. Featherston, Miss	910	1645	1316 00	588 00
Orlando B. Ficklin, Ill.	805	1290	1032 00	348 00
David Fisher, Ohio....	445	840	672 00	316 00
Thos. S. Flournoy, Va.	214	325	260 00	88 80
John Freedley, Pa....	154	154	123 20	
Richard French, Ky..	514	785	628 00	216 80
George Fries, Ohio...	291	540	432 00	199 20
Andrew S. Fulton, Va.	316	412	329 60	76 80
John P. Gaines, Ky.	[not down]	820	656 00	
John Gayle, Ala	1013	1248	998 40	188 00
Meredith P Gentry, Ten	695	1142	913 60	357 60
Josh. R. Giddings, Ohio	338	850	680 00	409 60
William L. Goggin, Va.	209	300	240 00	72 80
Daniel Gott, N. Y	337	530	424 00	154 40
James S. Green, Mo..	912	1740	1392 00	662 40
Dudley S Gregory, N.J.	224	229	183 20	4 00
Joseph Grinnell, Mass.	434	461	368 80	21 60
Artemas Hale, Mass...	433	461	368 80	22 40
Willard P Hall, Mo	[not down]	2078	1662 40	
Nathan K. Hall, N. Y...	381	700	560 00	255 20
David Hammons, Me.	578	665	532 00	69 60
Jas. G Hampton, N. J.	179	178	142 40*	
Moses Hampton, Pa..	226	341	272 80	92 00
Hugh A. Haralson, Ga.	739	947	757 60	166 40
J. H. Harmanson, La..	1194	2524	2019 20	1064 00
Sam. W. Harris, Ala..	826	1044	897 20	206 40
Wm. T. Haskell, Ten.	619	1603	1282 40	627 20
Thos. J. Henley, Ind..	578	910	728 00	265 60
William Henry, Vt...	445	558	446 40	90 40
H. L. W. Hill, Ten.	[not down]	1100	880 00	
Henry W. Hilliard, Ala.	839	1069	855 20	184 00
Isaac E. Holmes, S. C.	540	554	443 20	11 20
Elias B. Holmes, N. Y.	384	636	524 80	217 60
Geo. S. Houston, Ala..	733	1300	1040 00	453 60
John W. Houston, Del.	135	200	160 00	52 00
Saml. D. Hubbard, Ct.	726	379	303 20	42 40
Charles Hudson, Mass.	420	513	410 40	74 40
Washington Hunt, N.Y.	402	695	556 00	234 40
J. M. Holley, N. Y....	357	...	.....	.....
Samuel W. Inge, Ala..	886	1600	1280 00	571 20
Chas. J. Ingersoll, Pa..	138	140	112 00	1 60
Jos. R. Ingersoll, Pa....	138	136	108 80*	
Alexander Irvin, Pa....	218	262	209 60	35 20
Alfred Iverson, Ga....	757	984	787 20	181 60
John Jamieson, Mo...	917	1708	1366 40	622 80
Timothy Jenkins, N. Y.	366	500	400 00	107 20
Andrew Johnson, Ten.	437	590	472 00	122 40
D. S. Jackson, N. Y...	225	230	184 00	4 00
Jas H Johnson, N H	528	635	508 00	85 60
Robt. W. Johnson, Ark.	1065	2000	1600 00	748 00
George W. Jones, Tenn.	716	1204	963 20	390 40
John W. Jones, Ga....	664	894	815 20	168 00
Dav S. Kaufman, Tx	[not down]	2800	2240 00	

would probably increase the House Excess to over $50,000.

*These Members have charged *less* than the Post Office Department list allows them, as follows:

Atkinson, Va	$14 40	Hampton, N. J.....	$ 80
Broadhead, Pa....	7 20	J. R. Ingersoll, Pa...	1 60
Brown, Pa.....	80	Levin, Pa.....	80

SENATE.

Names.	Actual No. of Miles by Post Route.	Miles charged.	Mileage charged.	Excess of Mileage charged.
Wm. Allen, O.....	400	479	383 20	63 20
Ches. Ashley (dead) Ark.	2000			
D. R. Atchison, Mo	[not down]	2120	1696 00	
C. G. Atherton, N. H..	447	540	432 00	74 40
Geo. E. Badger, N. C..	288	288	230 40	
A. P. Bagby (out) Ala..		1398		
R. S. Baldwin, Conn..	300	330	264 00	26 40
John Bell, Tenn.....	664	1122	897 60	350 40
Thos. H. Benton, Mo..	808	1670	1336 00	699 60
John M. Berrien, Ga..	692	760	608 00	79 40
Solon Borland, Ark...	1065	2250	1800 00	956 00
John W. Bradbury, Me.	595	675	540 00	64 00
Sidney Breese, Ill.....	771	1670	1336 00	380 00
Jesse D. Bright, Ind...	560	1431	744 80	296 80
A. P. Butler, S. C.....	554	699	559 20	116 00
J. C. Calhoun, S. C...	531	923	738 40	313 60
Simon Cameron, Pa..	120	150	120 00	24 00
Lewis Cass (out), Mich.	524	1081	864 80	445 60
John H. Clarke, R. I..	400	450	360 00	40 00
John M. Clayton, Del..	115	120	96 00	4 00
W. T. Colquitt, Ga....		1040	832 00	
Thos. Corwin, Ohio...	469	765	612 00	236 80
J. J. Crittenden (out), Ky.	542	800	640 00	206 40
John Davis, Mass.....	398	440	352 00	33 60
Jefferson Davis, Miss..	1060	1981	1584 80	735 80
Wm. L. Dayton, N. J..	166	206	164 80	32 80
D. S. Dickinson, N. Y..	296	576	460 80	224 00
John A. Dix, N Y.....	370	400	320 00	24 00
Henry Dodge, Wis....	891	1850	1480 00	767 20
S. A. Douglas, Ill.....	884	1834	1467 20	758 00
S. W. Downs, La.....	1190	2600	2240 00	1288 00
Jno Fairfield (dead)	[not down]	600	480 00	
A. Felch, Mich.....	526	1121	896 80	59 20
T. Fitzgerald, Mich	[not down]	1281	1024 80	
H. S. Foote, Miss.....	1010	2580	2064 00	1056 00
A. C. Greene, R. I.....	400	450	360 00	40 00
John P. Hale, N. H....	495	567	453 60	57 60
H. Hamlin, Me.....	657	738	590 40	84 80
E. A. Hannegan, Ind..	645	1602	1281 60	765 60
Sam. Houston, Tex	[not down]	3120	2496 00	
R. M. T. Hunter, Va..	98	115	92 00	13 60
Henry Johnson, La...	1200	2354	1883 20	923 20
Reverdy Johnson, Md.	40	42	33 60	1 60
Her. V. Johnson, Ga..	648	940	752 00	232 60
W. R. King, Ala.....	818	1100	880 00	225 60
D. H. Lewis, (dead) Ala.	859	1090	872 00	184 80
W. P. Mangum, N C..	265	330	264 00	52 00
J. M. Mason, Va	74	143	144 40	55 20
T. Metcalfe, Ky.....	493	697	557 60	163 20
J. W. Miller, N. J.....	224	282	225 60	46 40
W. S. Moor. Me.	[not down]	745	596 00	
J. M. Niles, Conn.....	336	400	320 00	51 20
J. A. Pearce, Md.....	94	130	104 00	38 80
S. S. Phelps, Vt.....	481	530	424 00	29 20
T. J. Rusk, Texas	[not down]	2934	2347 20	
W. K. Sebastian, Ark.	1011	1900	1520 00	711 20
A. H Sevier, (out)	do [not dwn]	2125	1700 00	
P. Spruance, Del.....	132	150	120 00	14 40
D. Sturgeon, Pa.....	135	300	240 00	84 00
H. L. Turney, Tenn..	664	1207	965 60	434 40
Jos. R. Underwood, Ky.	668	740	592 00	57 60
Wm. Upham, Vt.....	516	600	480 00	67 20
Isaac P. Walker, Wis..	805	1880	1504 00	860 00
Daniel Webster, Mass.	440	530	424 00	68 00
J. D. Westcott, jr. Fa..	1069	1200	960 00	104 80
D. L. Yulee, Fla.	[not down]	1488	1190 40	

Total, Senate excess..... $14,881 40

The following is the interesting aggregate of the two tables.

Number of Extra Miles charged by Representatives.....	59,030
Number of Extra Miles charged by Senators	18,602
Total Extra Miles	77,632
Extra Mileage of the House.....	$47,223 80
Extra Mileage of the Senate	14,881 40
Total Extra Mileage.....	$62,105 20

THINGS IN WASHINGTON.

Correspondence of The Tribune.

WASHINGTON, Thursday, Dec. 20—12 P. M.

The Committee on Territories this morning voted on the subject of Suffrage in California. It was proposed to make suffrage universal. A move was made to confine it entirely to the whites. After an animated discussion it was put to test—Yeas 4, Nays 4, CALEB B. SMITH, Chairman, then gave the casting vote in favor of excluding all color but that which bore the unmistakable opposite of the Pariah.* Thus is the California bill to be reported on the same basis as that of Texas.

The subject of Cabinet-making engrossing so many pens, I will state on authority next best to the old General's word of mouth, that he will not form one, even in supposition, until he reaches Washington, and shall have had consultation with his personal and political friends.

The thermometer to-night, in ventilated halls, stands at 75° of Fahrenheit, and rain is rapidly falling, with meteoric exhalations distinguishable through the gloom.

Yours, &c. D. H.

* The Pariahs are the lowest caste of the Hindoos, and oppressed by all other classes; the term is therefore appropriate, as applied to the condition of the African Race in this country.—*Ed. Trib.*

Figure 2.2. "The Mileage of Congress" (1848). The *New-York Tribune* used Congressional data about travel reimbursement, presented with visually engaging vertical striping that stood out in agate type.

The *Tribune* ran multiple corrections in subsequent weeks to fix mistakes and numbers resultant from reasonable explanations that were not included in the ledgers he accessed – contending they hardly subtracted from the value of the overall work – yet the public still contested both its validity and virtue. The counter-narrative challenge, however, seemed not to matter much as eventually the House and Senate did agree on legislation to reduce the rate to 20 cents per mile. But they did leave in phrasing about a "usually travelled route," which even at just 20 cents a mile still left travel as a financially rewarding Congressional perk. Though Greeley's methods of using political position to gain source material might be considered ethically questionable today, at the time there was no Freedom of Information Act, which began guaranteeing in 1966 that journalists, and any person really, had rights and access to certain government data.

Klein (2015) identified five lessons from Greeley's efforts mixing data with journalism that he said still apply today: (1) Open records are important and essential; it all begins with access to data. (2) Journalists should anticipate the reaction to their data release; when numbers do not lie, it is inevitable some people will not like them. (3) Raw data is never raw; by definition any data set conceived has already begun being processed, and thus is impossible to be delivered completely without bias. (4) It was important to verify data with human subjects; too often there existed a reasonable explanation that required individualized attention to clarify. (5) "Bulletproofing" should be an essential component of data's pursuit; data are messy and errors are inevitable, but journalists have the skills of verification to keep their analysis clean and credibility intact.

Social Science Precision

A big leap toward data journalism came nearly 100 years after Greeley, in the same year the US passed and signed the Freedom of Information Act into law. In 1967, journalist Philip Meyer had recently completed a Knight Fellowship in which he studied the application of social science methods to journalism when riots in Detroit put his new outlook to the test. Six days of looting, shooting and fires had led to more than 7,000 arrests and 43 dead. Working for the *Detroit Free Press* and in partnership with the Urban League, Meyer deployed a team of 30 surveyors to African-American neighborhoods where the riots occurred in an effort to get to the root causes of the violence. With a scientifically valid survey, Meyer disproved certain theories that were being bandied about – ideas suggesting that the rioters were likely recent migrants from the South, for one, or were "riff raff" with no money and little education. With his survey analysis at its center, the *Free Press*'s coverage of the riots would go on to win the Pulitzer Prize (for local news coverage), and largely would change journalism moving forward.

In the short term, his statistical work changed the news conversation surrounding the riots, which eventually led to various social justice reforms. In the longer term, Meyer was setting a new standard for journalism and its use of data. These data did more than just supplement an existing story – they became the story. Meyer's reporting was revolutionary in how it showed academic rigors of social science could be applied to deadline-driven demands of a daily newspaper. His work essentially replicated a 1965 study conducted by USC evaluating the Watts riots; but while their work took two years, Meyer and his team, using a hybrid manual and computerized approach, took just one month (Meyer, 1988; Rosegrant, 2011).

In 1973, Meyer released *Precision Journalism: A reporter's introduction to social science methods*, which provided a more generalizable and instructional outlook on social

science applied in reporting situations. With the elections of the 1970s, precision journalism would show its strength as a matter of polling. In the elections of the 1970s, news operations were not just running stories about the polls, but were conducting their own polls. To some extent, it created competition to be statistically more valid. But despite a brief surge in statistical literacy among reporters and editors, soon editors began to accept polls at their word and no longer were they challenging data like they would any other sources (Wormer, 2008; Kovach & Rosenstiel, 2010).

Precision journalism was not just a matter of techniques but also was a theory of news similar to agenda setting and framing. McCombs, Cole, Stevenson and Shaw (1981) explained how they came around to such a conception when realizing the impact polling had on news of the day, and also in seeing the stories that could be done with computer-assisted content analysis of document troves. Changing the methods changed the stories, which changed the public's reaction to stories that subsequently emerged, they contended. Their study looked at three social science techniques – sample surveys, content analyses, and field reports -- and explained how they applied to newsgathering of the day.

> While most journalists may not recognize terms such as content analysis and controlled field experiment, most are comfortable with traditional reporting techniques of digging through documents or becoming part of the event long enough to get a story. All three of these methodologies represent techniques for shifting the journalistic focus toward direct, structured observation which is key to [precision journalism as a] new theory of news. (McCombs et al., 1981, pp. 27-28)

Beyond polls, precision journalism became instrumental in investigative journalism, particularly on behalf of social justice issues. In one case, the *Miami Herald* had made open-records requests for 3,000 criminal records over a 12-month period, and entered into a computer database details about arrests, charges, and outcomes, and attaching different lawyers and judges to each case. They found specific courtrooms where certain criminals got light sentences, frequent acquittals, or unusually stiff penalties. The Philadelphia *Inquirer* conducted a similar records search, and ran cross-tabulations on the data to reveal racial and gender biases in sentencing. (African-American men received the longest sentences, on average) (McCombs et al., 1981).

Follow the Data

Another key factor in precision journalism was its general reliance on computers. Meyer had used an IBM mainframe to crunch has statistical data from the Detroit riots survey. Computers took on increased importance as political polling became part of precision journalism, and by the 1970s they proved essential to the news cycle for the ability to report election results. The computerized data usage drove new forms of competition in news, with the quality of hardware and software sometimes determining which entities were able to break certain election news. When considered as a theory of news, precision journalism began to change the way some reporters and editors thought, advocating for journalists to begin to think like a computer to find government data, or how a database thinks about it. "[The reporter] no longer needs to sift through drawer after drawer of records but instead must now be able to conceptualize questions in terms appropriate to computerized government files" (McCombs et al., 1981, p. 30).

As computing became a bigger part of stories connected to it, precision journalism would evolve into computer-assisted reporting (CAR). The term CAR became quite accepted as it got integrated with traditional investigative approaches. With CAR representing a form of newsroom specialty, an infrastructure supporting it for investigative reporting built up around it[3] and by the 1990s the primary purpose of CAR was investigative journalism – using databases for example, to find disparities and crack through stories where sources were not talking (Anderson, 2013). This was before the days of widespread Internet use, so storage and ownership of databases became premium, with access to certain data becoming proprietary.

But that would begin to change in the early days of the twenty-first century, further accentuating the split and tension between technologists and writers, data and journalism. A new generation of tech-savvy workers came into the newsroom, but they were brought up on open-source philosophies (Coleman, 2004). Open data and public data were becoming key to its future, and what pulled data in new journalistic directions. For the UK, efforts to secure access to data began in 2000, but legal wrangling that would benefit data journalism would carry on over the next 10 years. With a new data law known as The 2000 Act (noting the year it started), Great Britain began a legal push to include free and open data governable under Freedom of Information principles. The law would be mostly implemented by 2005, affecting more than 100,000 public entities. The Act provided a "general right of access" to public information in 2005 and placed a duty on public authorities to develop publication schemes for the regular release of agency data[4]. But it was not just releasing the data that mattered to journalists, it was

[3] National Institute of Computer-Assisted Reporting (NICAR) became key to the investigative reporting organization IRE and currently serves as a repository for federal databases. IRE now gives several awards, including the Phillip Meyer Award for incorporation of social science methods.

[4] Former UK Prime Minister Tony Blair has since called passing the Act "one of the biggest mistakes of my career … For political leaders, it's like saying to someone who is hitting you over the head with a stick, 'Hey, try this instead,' and handing them a mallet. The information is neither sought because the journalist is curious to know, nor

releasing it in usable form. Then it became a matter of what information was still protected, a cause for which *The Guardian* would fight for liberalization, leading the "free our data" movement in 2006.

The Guardian has a special role in the history of data journalism. They have been noted in mainstream media and in academic literature as leading the charge in "freeing the data" and is generally well-regarded in the literature for aggressively taking a lead in establishing professional practices for data journalism among both journalists and non-journalists alike (Stray, 2010; Franklin, 2014; De Maeyer et al., 2015; Domingo et al., 2015; Gourarie, 2015). The British paper got credit for the coining the term "data journalism" and helping it gain stickiness over computational journalism and other sub-disciplines competing to be the heir apparent to CAR.[5]

In the spring of 2010, *The Guardian* had already launched its Datablog, on which it was making all its databases public. By 2010, the transition to data journalism was fully underway, as *The Guardian* had begun to familiarize the public and public officials with the concept of crowdsourcing. *The Guardian*'s 2009-10 project on expense claims of UK members of parliament (MPs) showed what was possible with crowdsourcing when the government released nearly 458,000 pages of newly public records, and *The Guardian* asked readers to fo through them for their members and flag questionable claims, resulting in investigative reports and visualizations that led many MPs to pay back many monies. *The Guardian* had shown its success in data journalism early, with a scandal reminiscent of Greeley's. The difference between the

given to bestow knowledge on 'the people.' It's used as a weapon," accusing journalists of going on "fishing expeditions" for salacious stories. (Rosenbaum, 2010).

[5] In 1999, Philip Meyer called for an end to the term "CAR," saying journalists were embarrassing themselves being the only profession that still held up use of a computer as something special (Meyer, 2004; Gynnild, 2014).

Guardian's work and Greeley's was how they crowdsourced the act of parsing through documents to see what matched up with representatives to the public, with specific instructions of how to go through the data, on which they revealed who was leaking money (Rogers, 2012). But what really changed things were the release of WikiLeaks – more than 92,000 rows of data about a military event in Afghanistan out the story of an entire war in code-able form. Part 1 of the Wikileaks War Logs was what made Simon Rogers editor moved from downstairs in a basement (with the graphics people) to upstairs with editorial was handle a new trove of documents being released by Wikileaks. "News organizations are all about geography—and proximity to the news desk," Simon Rogers explained (p. 36) "If you're close, it's easy to suggest stories and become part of the process; conversely, out of sight is literally out of mind."

In 2010, the ongoing fight for data access in Europe continued. *The Guardian* and the BBC together led the charge for not just access to data, but access to usable data – both as a matter of form and privacy. In their case, previously protected personal data such as data about all amputations was restricted because of personal medical privacy, but these news organizations fought for a mechanism to have it properly anonymized so they could use it. The government initially resisted, but the side of journalism, bringing Great Britain's Freedom of Information laws more in line with the US's FOIA (Stray, 2010; Parisie, 2015). In addition, they fought to get rid of the PDF a format excellent for delivering certain text-based documents, but for data "where data go to die" (Stray, 2010), acknowledging that the government should be assisting reporters with getting data, not encumbering them. *The Guardian* model became a data journalism standard, scholar and editor Liliana Bounegru explained in an interview:

> *The Guardian* Datablog model can be described as 'journalism as a trusted source of data.' In this model the act of journalism focuses on the provision of selected datasets related to the news of the day in a structured format, to an audience envisioned as engaged data-savvy citizens who actively participate in uses of public interest through production of data visualizations and their own analyses of the data. (Ršumović, 2013)

Challenges arose as more programmers found their way into newsrooms. As they had been part of open-source communities and front-line fighters for open-source advocacy, this level of involvement and participation proved uncomfortable for epistemological positions of journalists that never would have advocated that level of involvement. But more progressive news outlets, as *The Guardian* declared itself, were starting to accept that objectivity may have been less important than transparency in the new sociotechnical environments of digital media (Parisie & Dagiral, 2012).

"You're more accountable if you're transparent," explained Katherine Viner, deputy editor of *The Guardian* and editor-in-chief of *The Guardian* Australia at the time, and now editor-in-chief of *The Guardian* US. And she explained how this open approach is still the way to get scoops -- reaching out to the public for artifacts and evidence, running images, videos and narratives through verification procedures (Silverman, 2015), and otherwise benefiting from the speed and expansive outreach afforded by the Web. "In a world in which we are all flooded with information, readers also want to know how you arrived at a story, and how you account for any errors you may have made. This is why readers' editors who are independent from editors are so vital, and why the 'show your workings' approach is a powerful tool" (Viner, 2013).

CHAPTER THREE: LITERATURE REVIEW

Data journalism is shaping the present and future of news, but only recently has the concept begun to emerge as a distinct subset of literature in journalism and media studies. The literature I looked at explored a convergence of media technology and journalism to show how two precepts have much in common, but also differences that are defining tensions in evolving news spaces[6]. This included literature to help define the concept of data journalism, literature that helped operationalize current theoretical concepts related to Web media, and literature for conducting content analysis.

Defining Data Journalism for Research

As with many emergent phenomena, data journalism poses initial challenges for researchers simply in defining its terms. Without definitional consensus, applying empirical insight toward meaningful theoretical constructs becomes near impossible, and likewise data journalists seeking resources from their news organizations have found difficulty making their case for additional financial commitments (Fink & Anderson, 2015). A recent surge in literature has made strides toward adding empiricism to definitions within data journalism workflows, starting with efforts to more clearly define its pieces and processes (Ausserhoffer et al., 2015).

In 2012 the hashtag #DDJ emerged on Twitter in to stand for "data-driven journalism" as a means of grouping together different examples of datafied narrative visualization. Noted scholars using the #DDJ hashtag on Twitter include Edward Tufte and Albert Cairo, who use it to share and aggregate examples of DDJ and DDJ research. The phrase spawned websites

[6] "News spaces" refers to both the real-life analog, brick-and-mortar construction of news rooms, and the virtual expression of that destination in Web spaces.

datadrivenjournalism.com and Twitter feeds @datadrivenjournalism, indicative of the role data was beginning to establish in the journalism mainstream. The phrase put an umbrella over different varieties of the practice when they focused on determining how news organizations were practicing data journalism. Loosen, Reimer, and Schmidt (2015) reviewed literature focusing on #DDJ output in an attempt to put quantitative bounds on journalism being grouped under this hashtag. Loosen's research team analyzed submissions from 2013 and 2014 to The Data Journalism Awards, an annual judged competition produced by the Global Editors Network. From a pool of $N > 820$ submissions, they selected $n=119$ award-nominated data-driven journalism pieces to examine. Only 15 (12.6%) were winners. With a sample of winners and also-rans, they conducted content analysis and identified common traits of a data-driven journalism (p. 17). (See Table 3.1 for more quantitative detail about the findings.) In sum, their definition found eight determinants of "typical" data journalism – all applicable to the cases of this thesis. These included: (a) published by a newspaper; (b) covered a political topic; (c) relied on public data from official sources; (d) built on financial or geolocational data; (e) based on simple units of analysis (such as single persons); (f) compared values to shows differences and similarities; (g) combined more than two types of visualizations; and (h) allowed user engagement through interactive functions.

All eight standards applied to the two cases in this thesis, confirming their standings as exemplars of the practice. The only questionable reason was (c) relied on public data from official sources. Beyond these eight "typical" traits, Loosen et al.'s (2015) put quantitative form to #DDJ qualities. They found methods that incorporated newly gathered data sets with

Table 3.1. Eight traits of typical data-driven journalism. (from Loosen et al., 2015)

1. *Published by a newspaper.* Legacy newspapers with an historic commitment to journalism and access to wide-scale resources accounted for 42.5% of submissions. Only 5.8% came from sources that were online only. A growing number (14.3% in 2013 to 25.0% in 2014) came from investigative journalism organizations.

2. *Covered a political topic.* Politics was the topic of 48.3% of DDJ samples. Sports came in at only 2.5%, which surprised researchers as low considering sports are built on statistics, and visualizations of sports-related data journalism had been common for decades.

3. *Relied on public data from official sources.* 67.5% of data come from official sources, 44.2% from non-commercial organizations such as NGOs or foundations, and 41.7% from publicly available data. Only 16.7% came from private companies (journalists maintained high skepticism over objectivity and reliability of such commissioned data). Only 3% relied on leaked data. All winning entries, 100%, declared source, compared to 60% overall.

4. *Built on financial and/or geodata.* Financial data most popular (45.4%) followed by geographic data (42.9%). The use of personal data is on the rise (from 21.8% to 32.8%) as is use of metadata (12.7% to 20.3%), referring to anonymous data about collected data. Also on the increase is sensor data (34.5% to 43.8%), which describes geographical data measured with sensors.

5. *Based on simple unit of analysis, such as single persons*. Applied to 60.0% of sample, with more complex units, such as nations or companies, comprising 46.7% of entries. Only 10.8% used aggregate units of analysis such as a household, social class, or team.

6. *Compares values to show differences and similarities between objects of study*. Allowing readers to compare quantitative values was present in 85.0% of sample, 46.7% showed changes over time, 34.2% showed connections and flows, 11.7% used devices such as Top Ten rankings to show hierarchies.

7. *Combines more than two types of visualizations.* 100% of sample used visualizations in the form of pictures (60.0%), charts (54.2%), maps (49.2%), tables (26.7%), diagrams (18.3%), or animations (5.8%) – with more than two on average (M = 2.24, SD = 1.05).

8. *Allows user engagement through interactive functions.* Most popular engagement features were details on demand (55.8%), filtering (51.7%), internal search (26.7%), personalization (18.3%), or "playful interaction" (2.5%). 18.3% had no interactivity, but only 6.7% of winners were void of interactive features.

Note: Results from n =119 samples of #DDJ nominated for the Data Journalism Awards in 2013 and 2014.

traditional journalistic storytelling methods. Most submissions were created by a team with an average size of 5 or 6 members. External collaboration with partners was present in 35% of submissions. All were indicative of newsroom changes with the onset of data journalism. In the case of people killed by police, it was the absence of that public data from an official source that compelled the newspapers to find it from other public sources, sometimes official, sometimes not, but on a more granular level. While there was no official record of people shot and killed or just killed by police, they got it from multiple smaller official designees, as well as non-official people.

They too sought a new definition, and in the process highlighted and explained the role of crowdsourcing to some but not all practitioners of data journalism. Appelgren and Ngyren (2014) looked deeper to describe newsroom-based data journalism practices and began to establish that there were more than one way to conduct its practices. In 2012, Sweden had embarked on a nationally sponsored open-data initiative, and that left the news operations in the midst of transitioning newsrooms to accommodate data journalism – adding staff and equipment to support new commitments to adding technology. Appelgren and Ngyren sent online surveys to 186 journalists and developers from seven Swedish news media companies. The survey asked 13 questions about professional practices, to allow researchers to divvy up the pool of respondents into four groups according to experience – ranging from "not very familiar with data journalism techniques" to "practices data journalism techniques daily." Researchers were sure not to provide a definition of data journalism in any of the questions to avoid bias on the concluding question, which sought a definition from the respondent, from which researchers based interview questions seeking more qualitative explanation of thought processes about data journalism from seven management-level editors.

Researchers were able to deduce a common denominator to give a preliminary definition: "Data journalism involves a set of work methods used to make journalistic sense of raw data" (Appelgren & Ngyren, 2014, p. 403). What those work methods were could not be settled universally, as many say data journalism, like investigative journalism, was "just journalism." Appelgren and Ngyren found, however, that editors with more experience in data journalism, more experience requesting large data sets from government agencies, for example, were more likely to contend they required additional skills in journalistic pursuits based on a unique toolkit that revealed different types of stories than pre-data journalism (Pearson's $r = .426$ [$p < .01$]).

Appelgren and Ngyren (2014) also found among editors declaratively practicing data journalism two different experiences with crowdsourcing and visions of its utility. Crowdsourcing refers to the practice of seeking information by reaching out to the masses, to see who comes forward. Onuoha, Pinder, and Schaeffer (2015) in the Tow Center's *Guide to Crowdsourcing* define it as more than just mining for information. They say, "Journalism crowdsourcing is the act of specifically inviting a group of people to participate in a reporting task – such as newsgathering, data collection, or analysis – through a targeted, open call for input; personal experiences; documents; or other contributions" (p. 8). This could occur, they found, in unstructured call-outs – an open invitation for engagement – or structured callouts, which presented a more targeted outreach with specific requests.

Of the eight editors Appelgren and Ngyren (2014) interviewed, half reported a lack of enthusiasm for crowdsourced projects because "they were too easily hijacked by rogue elements of the public" (p. 402). However, other editors reported the opposite experience – saying the easier they made it to participate the easier it was for them to tap into the "wisdom of the crowd" (p. 402). Appelgren and Ngyren made it clear that this relationship with the reader is

instrumental. The audience chose whether to participate, according to *ease of participation* and whether or not a topic was *inspiring*.

The notion that journalism itself was changing with the emergence of what would become known as user-generated-content (UGC) became apparent to Bivens (2008) as cell phones became more capable. She recognized that new levels of audience engagement would change journalistic practice, as well as the relationship with the reader amid evolving news values. The key factor of UGC, however, as she saw it, was in its power to change the conversation because of the additional eyes becoming present. She gave Saddam Hussein's execution as an early example of how UGC revealed a story different from what the media was telling. Initially, the story told by the news outlets was about a solemn undertaking until video taken by one of the guards revealed a far crasser undertaking. With video proof, news outlets could not ignore what all were seeing, and thus the story had to change.

In the ongoing story of people killed by police, it was the death of Walter Scott, the man in Charleston, South Carolina who was seen shot in the back *after* Officer Randy Slagle had already filed a report that would be backed up by other police but did not jibe with what a citizen-captured video showed. His killing emerged in the wake of Tamir Rice and Freddie Gray, while Michael Brown was still in memory. Each death advanced additional narrative with Brown's telling a story of plausible injustice, Rice's showing police willing to act before even a child has a chance to surrender, and Gray's telling a story of additional officers, not even initial responders, being potentially complicit. But Scott's death in April 2015 was the one where the media realized they had possibly gotten many stories all wrong, some self-reflection knowing how they would have told the story without the citizen-captured video to tell another story. My initial research into this area found decades of a similar stories. Was it possible the media had

been prematurely buying a police script without doing the necessary due diligence to get to truth? This killing was significant because it was where the media became aware of how wrong they may have been getting the story. For years and decades, police have been providing a story, corroborated by officers during an investigation. But at the same time, if that were true, this shooting happened before any databases had been released, and indeed there had not yet been much consideration of video as data. Researchers have long understood the power of video as an emotional trigger (Hedley, 2013) since Rodney King. So if videos were enough to rally activists and change the conversation, really, while the death total may be a statistic, the database put faces on the data, and provided the deceased's network a virtual rallying point.

Data in the Newsroom

Transparency came from a synergistic commitment to transparency from both the programmers' culture and those trained in general journalism. Long held as a journalistic bailiwick, objectivity in new media constructions is being challenged by a notion that transparency might be the key – fueled by new levels of reader engagement. We should have seen it coming, said Meyer (2004), who declared "The end of pseudo-objectivity" was upon us, but there would still be a need for the pursuit of it through new data methods. Lewis and Usher (2013) was one of the earlier studies to address data journalism in its current postmodern form and begin to provide a structure for understanding it as a sociotechnical construction. Essentially, they noted that as a new generation of technologists came into newsrooms, they entered having been brought up in an open-source movement – which blended well with journalists brought up in a "culture of verification" (p. 611) and brought open-source mentalities in the newsroom to the fore. Journalists long strove to be open and honest. Through this overlap of culture, he identified

transparency, *iteration*, *tinkering*, and *participation* as four principles that were emerging from a "hacker's" ethos – four shared characteristics in governing philosophy between technology specialists and newspeople. With readers more involved in helping get things right. "transparency more important than objectivity", reputation for integrity through transparency. But these standards certainly had not found their way to the highest offices of every newsroom.

> These values of iteration, tinkering, transparency, and participation, each embedded in the open-source ethic, can be brought into the newsroom as architecture *and* culture – as a structural retooling of news technologies and user interfaces, and as a normative re-articulation of what journalism means in a networked media setting. (Lewis & Usher, 2013, p. 615)

As these ideas became better understood, there still was enough uncertainty in determining what constituted data journalism that Coddington (2015) tried to offer some definitional consistency. With Coddington (2015) researchers had the beginnings of a framework to assess multiple characteristics of data journalism that were changing over time. "The principles of open source have been an important common ground for bringing together 'hacks' (journalists) and 'hackers' (technologists)" (p. 333).

Coddington (2015) was bothered by conflicting definitions that were getting in the way of a coherent body of literature from taking shape. In an effort to establish normative dimensions for data journalism on which theoretical research could be built, Coddington sifted and sorted through hundreds of texts to use stacks to identify four particular dimensions on which he found polarity that differentiated three types of quantitative journalism forms – computer assisted

reporting (CAR), computational journalism (CJ), and data journalism (DJ). Then through close readings of 90 texts from professional and academic discourse, he placed them along four dimensions (p. 332) to develop a typology for analysis. The typology he created (Figure 3.1) presented these four dimensions as a binary range between two opposing poles, with each of these dichotomies highlighting different outlooks.

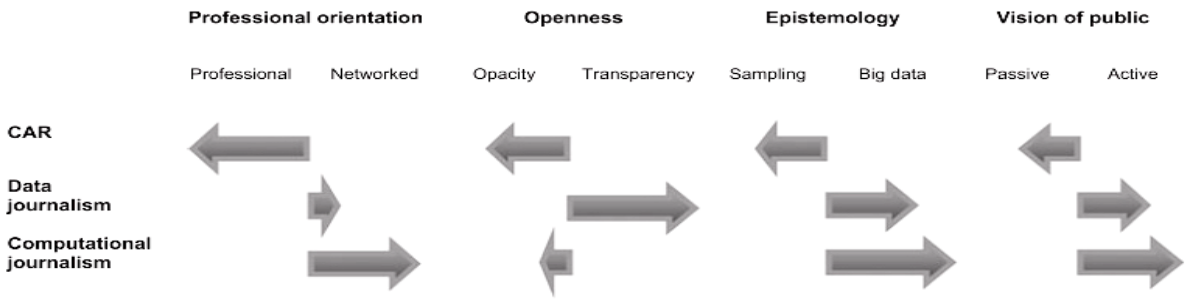


Figure 3.1. Coddington's typology. Data-dependent journalistic forms showed a preference in journalistic outlook across four binary dimensions determining decisional characteristics.

From left to right, professional orientation considered a journalistic methodology that looked either to a closed professional circle of well-seasoned sources and a broad Rolodex of the perfect connection for specific needs, or to a wide-open network of professionals and non-professionals, experts and non-experts, reaching outside usual circles to reach the masses.

Coddington (2015) evaluated this dimension on a spectrum of Professional Expertise versus Networked Information (p. 338). The professional leaning revealed a traditional journalistic culture that found value in "a semi-secret Rolodex of informed connections" but otherwise deferred to official sources and gave authority to anecdotal example. Indicators of a professional expertise orientation included use of: *personal anecdotes*, *expert analysis*, and reliance on *credentials for credibility*. On the other side of that spectrum were techniques that drew their journalistic credibility from Networked Information – the ability to draw journalistic

narrative from a broad range of sources and rely on a network of experts and non-experts alike, professionals and non-professionals, as an act of collaborative verification (Hermida, 2012).

"Transparency has been an ascendant journalistic value over the past decade, one characterized as a crucial element to establishing credibility with an increasingly mistrustful public" (Coddington, 2015, p. 340). This openness in journalism stemmed from open-source traditions in software and academia to suggest a skeptical reader should be able to see the journalist's work. The idea of openness was built around expanding *access to data*, as any effort to improve the reader's access to data – in the spirit of having nothing to hide – provided credibility (Lesage & Hackett, 2013). But providing access to *raw data* was not necessarily the same thing. Openness ventured into an important subset about *transparency*. Indicators of transparency that relate to credibility included *sharing data*, *showing methodology*, *revealing sources of data*, and *providing additional documentation* for participatory analysis. Opaqueness was not inherently bad – proprietary software code has become just as important to development as undercover and protected sources had been in journalism. "Data's meaning and value arguably stems from the extent to which it is said to be objective. But if objectivity's place within journalism is itself the source of much debate then we must also question how data is imbued with the quality of objectivity within journalism" (Lesage & Hackett, 2013, p. 40).

As to epistemology, even venerable journalists turned out to have had little trouble embracing new ideas of "big data." The scientific processes built in precision journalism long challenged the profession (Messner & Garrison, 2007; Wormer, 2007), so it was not a surprise the less precise methods of big data held great appeal. Epistemology on Coddington's typology referred to how data was gathered and analyzed, and determined what impact collection processes had on meaning of associated journalistic stories. The other side of this binary choice

was *selective sampling*, referring to academic traditions that anchored Meyer's prescription for precision journalism, and its subsequent evolution to CAR. Technology has allowed scientists to move away from the need for sampling in favor of pursuits guided by the theoretical concept of *N*=all (Mayer-Schönberger & Cukier, 2013). As technology allows scientists to move away from the rigor of sampling, data journalism has been able to show less concern for causality in favor of correlation. But this move has created epistemological tensions for investigative reporters from the CAR tradtions, particularly as associated stories move from hypothesis-driven analysis to data-driven reporting (Parisie, 2015).

Coddington explains the view of the public in journalistic tradition, as a singular mass with occasional lone standouts in the form of letters to the editor and op-eds. This dimension is instrumental in evaluating the relationship between reader and journalist. Coddington noted the shift from passive to active, in perception and functional reality. Indicators here include giving the reader the ability to manipulate data through *filtering, hovering, ordering, data on demand* (Segel & Herr, 2010), and *interactive dashboards* (Cairo, 2015). Where the difference becomes clear throughout is in vision of the public – seen either as active and passive. This is the big changes, and the area important to my research questions. One of the most significant areas of change that Coddington noted was the view of public. And in that are many of the changes coming to journalism. "It's not that long ago journalists were the gatekeepers to official data. Now that dynamic has changed beyond recognition. Our role is becoming interpreters, helping people understand the data," writes Rogers (2012, p. 36). "Numbers without analysis are just numbers, which is where we fit in."

The Tow Center for Digital Journalism at Columbia University has defined crowdsourcing and people are talking about it, but it suggested something more about news in

the context of a changing relationship with the reader (Onouha, Pinder, & Schaffer, 2015). Crowdsourcing was a means of attaching meaning to reader engagement. Even for readers not directly participating, *The Counted*'s use of crowdsourcing put forth a perception of how it happens, that this was a group project. "*The Counted* lives at the intersection of crowdsourced data collection and traditional reporting methods" (p. 14).

One key difference between *The Counted* and *Investigation: Police Shootings* was an apparent openness to crowdsourcing. Tow Center put out a white paper on crowdsourcing that also included a case study of *The Counted*, (among eight total) as it pertained to crowdsourcing. This consisted of explanatory interviews about what had become one of their strengths:

> Still the journalistic leader in tasking volunteers with crowdsourced requests has been *The Guardian*. Since 2009, it has tapped into the power of its audience base, expertly finding ways to work collaboratively with its large and active audience. As time has passed, *The Guardian* has learned how to target specific communities with clear tasks, while simultaneously broadening the types of answers it is open to receiving. The result is a seamless, back-and-forth interaction that benefits all parties. (Onuoha, et al., 2015)

What makes *The Counted*'s reader interactions succeed were the same factors that made them resource intensive. *The Counted* from the start was about audience engagement. It was globally hyperlocal, dealing with readers on an individual level, acknowledging every tip, counting everyone as significant until proven otherwise (Onuoha, et al., 2015). One discovery in the process was different mediums had different roles, and thus media needed to be on multiple platforms. Different groups of people were getting news from Facebook or engaging on Twitter,

this collection of case studies on crowdsourcing found. Some were on their smart phones while commuting and others were at the office. In this community, readers engaged in the spirit of working toward collective accuracy, an engagement that gave them a sense of purpose (Domingo et al., 2008), which brought the crowd into crowdsourcing in a way that extended the publication's reach.

Unsettled definitions of openness in journalism and an uncertain authority relationship with the reader challenged the objectivity paradigm that has for so long guided journalism (and science, and precision journalism). What was really at stake was the relationship between author and readers. An essential part of data journalism was a shift in how data journalists saw the reader as an active participant – whether that be from crowdsourcing engaging with the site to cast votes, correct facts, contribute videos, or participate in collaborative verification (Lesage & Hackett, 2015). Coddington (2015) gave examples, but said there was much more to draw from. To establish some means of understanding, a theoretical mishmash also had to be considered. Tandoc (2014) suggested much of it could be all for naught, as the mainstream media have long rejected audience feedback and are more inclined to go to numerical analytics that were not influenced by the audience.

Tandoc (2014) did a case study looking at three examples of news operations seeking audience feedback, revealing how with UGC, the audience did have control of a story – not complete but forcing the online media to contend with its content and implications. But not all media is embracing this, as part of a quest to maintain autonomy and because journalism as an institution had long not really wanted to listen to its audience. Journalism as an institution had long rejected audience feedback. But what is different now, too, is that some are being driven by the desire for views and clicks – artificial engagement.

If the boundaries between audience and professional journalist were blurring with development of the Web, advances in mobile technology exacerbated that condition (Bivens, 2008). But it would not be long before user-generated-content (UGC) would show its capability for changing the shape of news conversations. Domingo et al. (2008) called it "participatory journalism." He conducted an empirical study that looked at audience participation in 16 online newspapers around the world to determine what that really meant. This research team found first that there were many ways readers engaged with websites in ways where they contributed as a participant to the journalism being offered. This could include interactive tools such as hovering, filtering, and data-on-demand, but what was more determinative, Domingo and his team found, were the areas where such engagement between readers and the media source occurred. He defined these spaces as *news production spaces*, *commentary and debate spaces*, *networking or distribution spaces*. This study showed multiple ways readers were able to engage with a website – by voting, submitting content, and what they enable users to do, which has a role in helping them build a community, where the group takes responsibility and lead by example. But what mattered more than these tools were the areas they engaged as a matter of journalistic task:

Access/observation

Selection/filtering

Processing / editing

Distribution

Interpretation (comments)

These were the areas where the journalists determined the roles the readers could engage in, and as such retained a traditional gatekeeping role. Domingo et al. (2008) concluded that citizens never really get the power. The journalists still controlled the narrative, and that was acceptable for committed contributors who recognized their role in contributing to the community. Data are supposed to be the great democratizer. But gatekeeping functions evolve. And the media still maintain control (Gray et al., 2012; Rogers, 2012).

Calls for Content Analysis

While literature review set examination and analysis on the right path, effectively answering the research questions required content analysis. Calls for such have arisen in an emergent corpus of literature on data journalism and user participation (Domingo et al., 2008; Lewis 2012; Lesage & Hackett, 2013; Ausserhoffer et al., 2014; Coddington, 2015). But content analysis of the Web can be quite difficult. For one, the Internet is constantly changing, across many different dimensions. Websites change over time – whether because of programmatic or artistic tinkering, or due to the continuous addition new content. Similarly, different people connect to different components in different ways, in different locations on different devices at different times – affecting how the reader (or coder) receives content to be analyzed. Maintaining intracoder reliability and intercoder validity becomes seemingly impossible (McMillan, 2000; Kim & Kuljis, 2010; Lee, Lewis, & Power, 2012; Lewis & Usher, 2013).

Schneider and Foot (2004) contended the best way to combat the ephemerality of Web spaces was to focus on the structure of a site, conscious of the platform on which the content sat. This sort of structural analysis, they contended was what underlay any network or web sphere to be analyzed. Herring (2010) agreed, but called for an expanded paradigm – one that does not

ignore insights from textual analysis, visual analysis, link analysis, exchange analysis, and all sorts of other assessments at any one given time. She said that because of the interconnectivity of HTML, there was no reason not to conduct research cognizant of the scattered forms of analyses.

From a data journalism perspective, all of the above represent input into a media operation producing data journalism. The output similarly consists of scattered theoretical explanations and thought. Hermida (2012) explained its relevance in ambient journalism, how news stories of the day exist in something of a cloud, spreading with different veracity across different Internet platforms and different devices (Yuan, 2011). That is what the ouput looks like. That is what we know for online interactions, and Web engagements. And now, the next phase becomes assessing how data exists and influences existence within that ambient mix.

To untangle this complex "web" of scholarly inquiry, Herring (2010) informed the conceptualization of a research design to best assess how decisions setting up a Web environment potentially mediate interactions with readers and the public.[7] Herring provided foundational guidance in explaining how content analysis for the Web (WebCA) "considers content to be various types of information 'contained' in new media documents, including themes, features, links, and exchanges, all of which can communicate meaning" (p. 245). Through text, pictures, and other content elements, Herring (2010) provided an update to McMlllan (2000) on applying content analysis to the Web for a framework with which to analyze how data journalism conveyed transparency and supported reader participation in news construction process. By creating interchangeable analytical parts, her framework supports a search for holistic understanding through analysis of data journalism components, which

[7] As a matter of definition, the "public" represents tangible human masses in digital online spaces – a subset of all of humanity that have the ability to connect with a journalistic output, from which a further subset become "readers" individually, or collectively the "audience".

included text, links, settings, features, images, and graphics. Thus, analysis of structural features made possible by HTML coding language offered new opportunities for insight, and Herring's directional urgings explained how these mechanisms work simultaneously across multiple components.

Summary of Literature

As explained in the next chapter, I used insights gained from this literature on precision journalism, data journalism and framing to compare and contrast *The Counted* with *Police Shootings* to see what they revealed about essential mechanisms with data that are possibly changing journalism in the twenty-first century. As research made inroads defining data journalism as a practice, and careful study revealed data journalism to be changing newsrooms, but some with more direction and determination than others do. At this crossroads of journalists and technologists, they were finding mutual benefit in sociotechnical change.

There has been in the past few years much literature scattered about on data journalism. Ausserhoffer et al. (2015) tried to make sense of it, adding to works by Loosen et al. (2015) and Appelgren and Ngyren (2014). But Coddington (2015) presented the research that offered a framework, which I would be able to apply using techniques from Segel and Herr (2010) and Herring (2010). This would allow me to begin applying the insights from Domingo et al. (2008, 2012), about the changing nature of participation in journalism, and Bivens (2008), who assessed how UGC was changing the news conversation, forcing journalistic entities to contend with certain amateur creations as an epistemic condition of contemporary news environments.

CHAPTER FOUR: METHODOLOGY

My research consisted of a case study using content analysis of *The Counted* and *Investigation: Police Shootings*, triangulated with news articles and interviews. Comparing and contrasting these two currently active data journalism projects built from the same essential data and contexts provided a unique opportunity for discovering how structural and processual components of data journalism might influence user engagement. The "liveness" of the two cases would prove challenging. I sought to identify Web features and describe their relevance to data journalism, and their role in mediating interactions between readers, journalists, and data – situations that had to be observed in near-real time to be effectively analyzed.

Comparative Case Study

My research design followed the model prescribed by Yin (2003) for multiple cases, with *The Counted* as Case One and *Police Shootings* as Case Two, and embedded subcases to explain social media extensions, reader comments, and other media that exists in the ambient cloud surrounding these digital artefacts as an ongoing story (Hermida, 2012). I drew context from news archives, podcast discussions, interviews, and email exchanges, but the primary means of study was content analysis – a systematic review guided by Krippendorf (2013), using tools provided by Herring (2010). I built a framework according to Coddington (2015), and supplemented his work with insights from additional literature to give particular attention to theoretical concepts surrounding crowdsourcing and gatekeeping. My goal was to identify components in these two cases and operationalize concepts relevant to the data journalism practices. Thus I looked to Herring's (2010) Web content analysis (Web CA) for guidance on

using Web components to analyze these two cases in relation to concepts gleaned from observation and literature review.

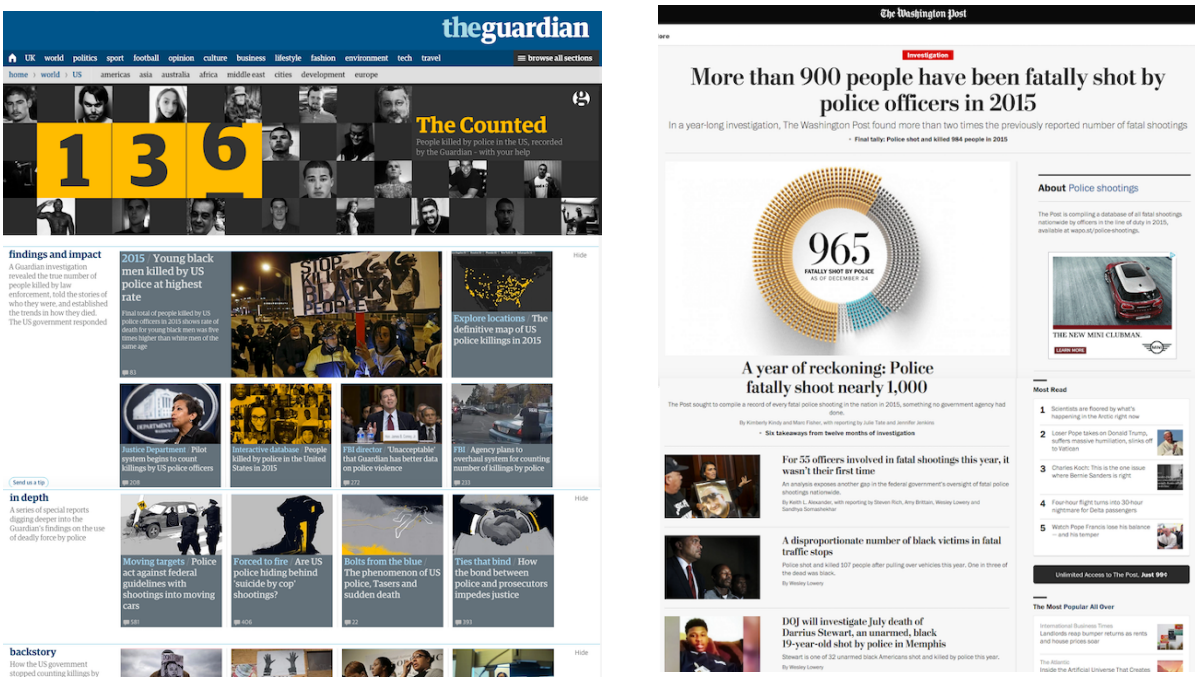


Figure 4.1. Comparative case study. *The Counted* (left) and *Police Shootings* (right) presented two different approaches to telling stories with similar data sets.

The comparison between these two consisted of examining *The Guardian*, identified repeatedly as a global leader in data journalism (Anderson & Fink, 2013; Ausserhoffer et al., 2014; Gynnlid, 2014; Coddington, 2015), and *The Washington Post*, which has incorporated data into its journalism, but position it as merely a part of a decades-long commitment to investigative journalism. Each case consisted of a data set visualized with interactive features, news articles related to that data set, pages explaining the acquisition and maintenance of that data set, and social media surrounding that data set. Hermida (2012) said that journalism exists in an ambient realm. Likewise, Yuan (2011) noted these media are consumed across all different devices,

which can have an impact on how they are received. Thus the complexity of these interconnected iterations of the work must also be considered.

Figure 4.1 sets the stage for how the two cases matched up. These cases were informed by more than 100 Web pages that were either part of or a few clicks away from these two data sets and the journalism surrounding them. I eventually settled on 24 unique URLs – 12 from each publication – as my fundamental units of analysis. They were chosen because they provided a comprehensive look at the totality of these projects, and presented a balanced approach in terms of the pages' purpose and usage on each site. The groupings I developed to categorize them aligned with Domingo et al. (2012), who suggested interactions in participatory journalism had to be considered separately in news production spaces, interpretation spaces, and distribution spaces. This justified my set of URLs, as I looked at data analysis and news articles, comments, and social media for each.

Table 4.1 highlights a basic explanation of their self-attributed data sourcing. Both acknowledge they started with databases that have been tracking this for a while – KilledbyPolice.net and FatalEncounters.org (Figure 4.3).[8] These citizen journalism efforts each have slightly different criteria for what counts on their list. It is of note that their numbers were different at the two publications in my study, and this was because they took different standards. But they started from these sites, as well as tips, Google searches, email news alerts, telephone calls, and other means of verification. *The Guardian* called their process crowdsourcing and reached out to the readers for help, while the *Post* credited their own digging.

[8] Two original source data sites. KilledbyPolice.net stopped operating at the end of 2015. Fatal Encounters is trying to log every police fatality back to 2000, and gives its community tools for custom visualizations.

Table 4.1. Two cases for data journalism.

	***The Guardian* US**	**Washington *Post***
Basics	Online-only US edition of UK-based newspaper Founded in 1821, US edition in 2011	DC-based newspaper with online presence Founded in 1877
Headquarters	New York City	Washington, DC
Funding	Non-profit: Funded by donations and advertising, subsidized by Scott trust	For profit: Purchased by Jeff Bezos of Amazon.com in 2013 for $250 million
Reputation	Populist leader in data journalism	Establishment leader in investigative journalism
	The Counted	Investigation: *Police Shootings*
Launched	June 1, 2015	May 30, 2015 (narrative) June 30, 2015 (data)
Tracking	"deaths arising directly from encounters with law enforcement [including but not limited to] people who were shot, tasered and struck by police vehicles as well those who died in police custody"	"shootings in which a police officer, in the line of duty, shot and killed a civilian — the circumstances that most closely parallel the 2014 killing of Michael Brown in Ferguson, Mo., which began the protest movement culminating in Black Lives Matter and an increased focus on police accountability nationwide"
2015 deaths counted	1,145 (3.14/day)	990 (2.71/day)
Database sources	"combine[d] Guardian reporting with verified crowdsourced information"	"news reports, public records, Internet databases, and original reporting"[10]
Data team size (est.)	21-24 people	13 primary, up to 70 total
2015 Awards[9]	- Finalist, Goldstein Award - Winner, Data Award, FiveThirtyEight - Winner, SPJ Sigma Delta Chi Award	- Finalist, Goldstein Award - Winner, Polk Award - Winner, Sigma Delta Chi Award - Winner, Pulitzer Prize, National news
Pay gates	Bottom 5% of screen covered with optional donation request, with page explaining its contribution to Guardian-style journalism	Paywall covers 100% of screen after 10 views per calendar month; cost to lift for national (not local) edition $.99/wk to $10/mo

[9] A representative but not exhaustive list of accolades and honors.
[10] In April 2016, the *Post* replaced "Internet databases and original reporting" with "social media and other sources"

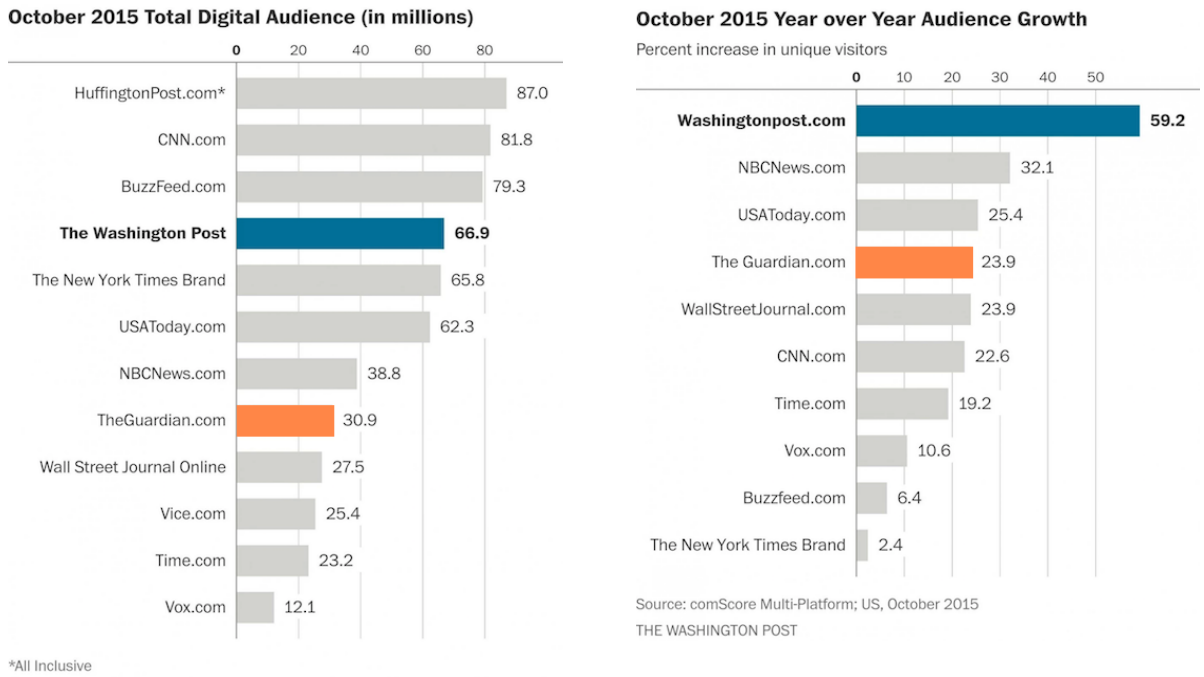


Figure 4.2. Online traffic and audience growth. *The Washington Post* in blue and *The Guardian* in orange. (WashPostPR, 2015)

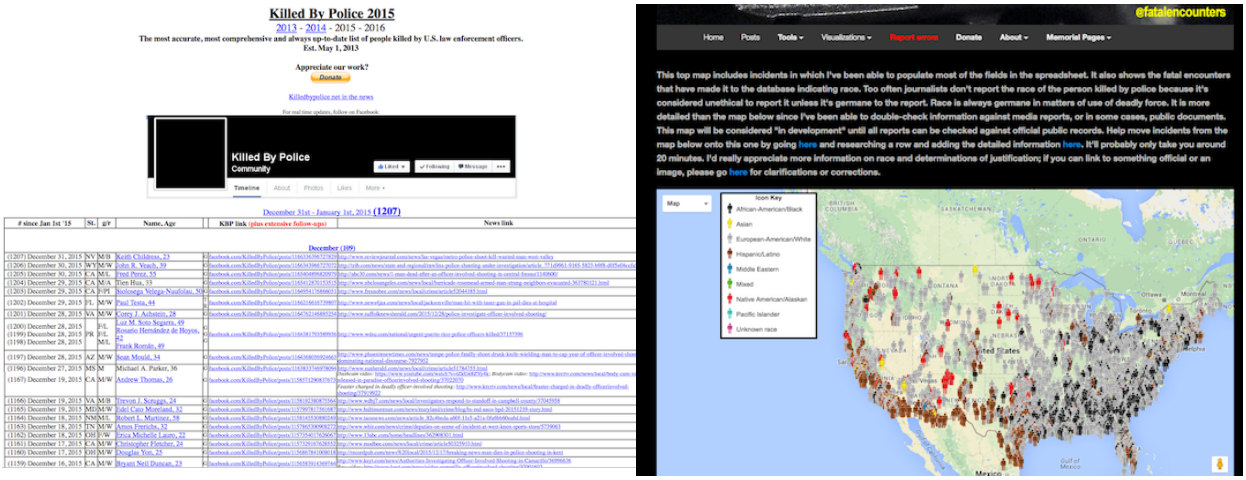


Figure 4.3. Providers of original source material. Both cases in this study credit KilledbyPolice.net (left) and FatalEncounters.org (right) for their help, but in different ways.

Figure 4.2 showed the reach by taking traffic data presented by *The Washington Post* and augmented it visually to show how both of these two cases exist in publications that have some of the largest online news consumption audiences online. Figure 4.3 showed the two originating data sources. This was important to know to understand the basis of different criticisms, as well as for understanding media behaviors toward shared and public data.

Case 1: *The Counted*

The Guardian has a long history as an acknowledged leader in data journalism. Parcel to their mission is a commitment to open-source and open-data ideologies (*The Guardian*, 2016) – a concept from coders that transitioned easily into journalistic practice. *The Guardian* has been on the front lines to fight for open data in Europe, and are verbally committed to establishing crowdsourcing, and data-driven journalism.

This is special as it identifies as a crowdsourced endeavor focus is on "you." Data editors at *The Guardian* first conceived of *The Counted* after traditional journalistic inquiry in the wake of Trayvon Martin and Michael Brown's shooting deaths revealed a lack of available and reliable statistics. This led to the ongoing development of a crowdsourced endeavor that attempted to connect directly with their readers. See Table 4.1. *The Counted* "is a project by *The Guardian* – and you – working to count the number of people killed by police and other law enforcement agencies in the United States throughout 2015 and 2016, to monitor their demographics and to tell the stories of how they died" (Retrieved from www.theguardian.com/us-news/ng-interactive/2015/jun/01/about-the-counted).

The Counted launched on June 1, 2015, and appears on the US edition of *The Guardian*'s website. *The Guardian* launched their US edition in 2011 as an online-only edition. The team

started with data about people killed during the first six months of the year and continued to add each new fatality – a rate that would become 3.1 a day. Though *The Guardian* is a British company, this project comes from *The Guardian US*, a 23-person team based in New York City. Four primary journalists led the effort – Oliver Laughland, Jon Swaine, Jamiles Lartey, and Ciara McCarthy. Assisting them have been researchers, developers, writers, artists, and editors. And, they repeatedly let readers know, also on the team is *you* – emphasized by them in email.

The Counted starts with a visual image of an odometer, providing where the count stands, with the last digit half-ticked, visually implying a certain inevitability that the next one is coming soon. The page visually and collectively highlights each person killed by police, with a link to their data. The project's public display includes four key pages identified on the top Navbar, all of which were fully analyzed. "Send a Tip" and "About" are text-based pages revealing much about the project's intention. The "Database" page is *The Counted*'s primary Web destination, and is a place where the reader interacts with the data to reveal additional data troves from each person. Each one of these squares corresponds to the data set for each individual killed by police. *The Counted* consisted of four main pages that make the scope of this project understandable. One is the database display, one is a collection of background articles that explain the project, one is the "About" page, and the other is "Send a tip." All four of these exist under the same Navigation bar; also included among them is a "Join Us" button that connects the reader to *The Counted*'s social media feeds on Twitter and Facebook. I examined each one in depth. For a complete list of URLs for these pages, see Appendix A.

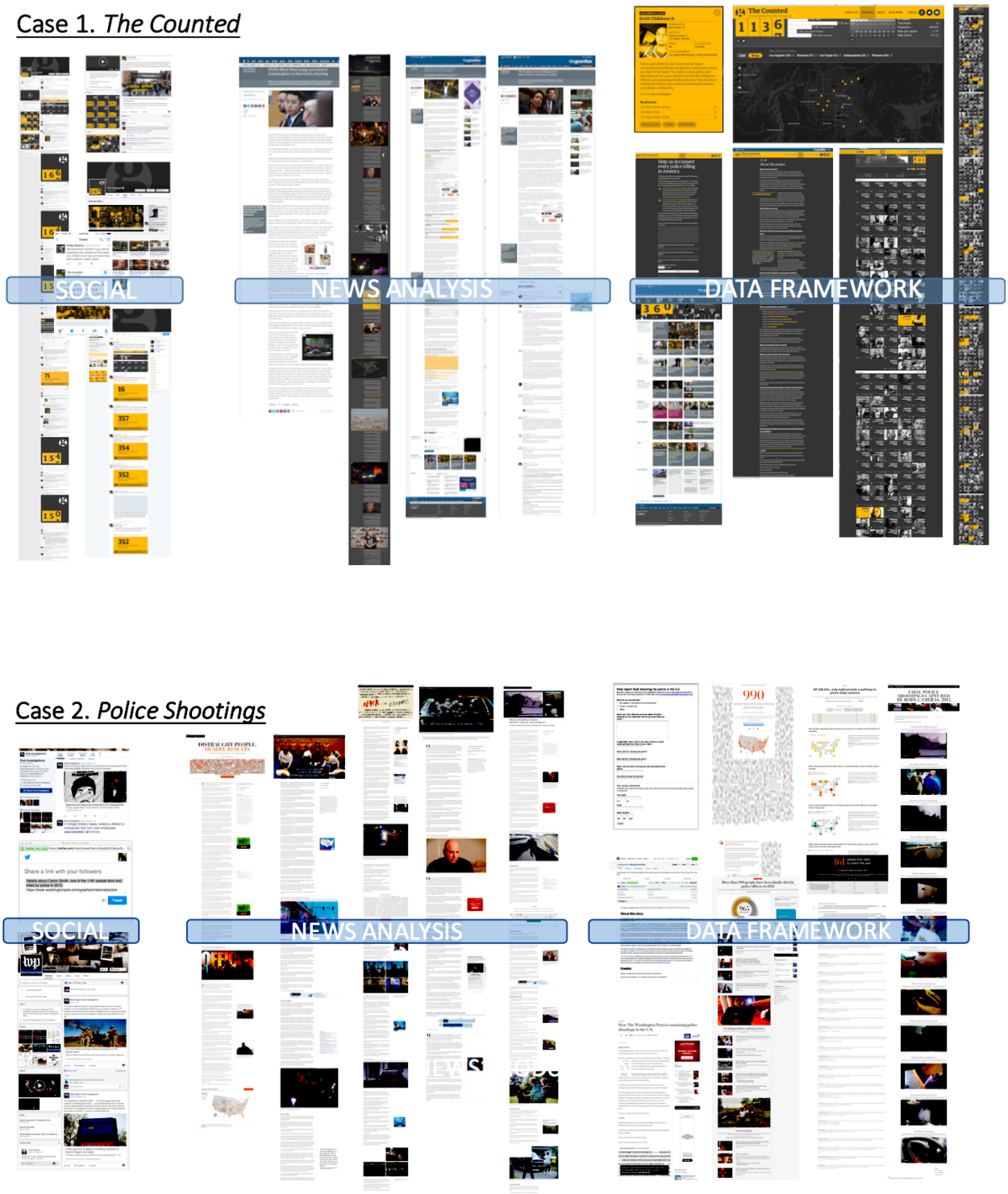


Figure 4.4. Pages for balanced case analysis. Different pages for each case comprising principal units of analysis were grouped by category to give representative, balanced looks at the two sites.

Case 2: *Police Shootings*

The Washington Post has a storied history of investigative journalism, having won 47 Pulitzer prizes among countless other awards and honors, with its heyday still celebrated for its coverage of Watergate as an example of investigative journalism at its finest. They set a standard for dogged investigative journalism. Their Woodward-and-Bernstein method of "follow the money" and checking sources relentlessly would become a standard for investigative journalism if not all journalism taught in journalism schools for decades that followed. And the idea that it was never one big story, but a series of smaller stories, building up to an inevitable whole – that is the standard *The Washington Post* tries to adhere to. When Amazon founder and CEO Jeff Bezos bought the *Post* in 2013, he heightened its digital focus, making the *Post* the fourth most visited newspaper online in the world – with 50.5 million visitors a month to the Website, and 37.5 million monthly unique mobile users (Brunyee, 2015). Bezos met with Bob Woodward to discuss bringing the foundations of *The Washington Post* to the digital age. While Bezos has made data, and capitalizing on Big Data, core to the *Post*'s new mission, data journalism has not been its focus (Meyer, 2014).[11]

The *Post* unveiled *Investigation: Police Shootings* twelve hours before *The Guardian* on June 1, with stories about what their analysis was finding. They had been running small stories in the runup to it. But their data would not be released for public consumption until June 30. "*The Washington Post* is compiling a database of every fatal shooting in the United States by a police officer in the line of duty in 2015." Their team consisted of 14 staffers – six people listed as

[11] The *Post* was doing notably well with data journalism through its Wonkblog under the leadership of Ezra Klein. Crafted in the spirit of Simon Rogers' *Guardian* Datablog, Wonkblog became a successful traffic driver about policy, loaded with data visualizations and analysis, as well as a profitable sales vertical. But when Klein, then 29 years old, proposed a $10 million upgrade plan for staff and equipment, the paper's leadership balked, which prompted Klein to leave the paper for a new startup at Vox media (Byers & Gold, 2014).

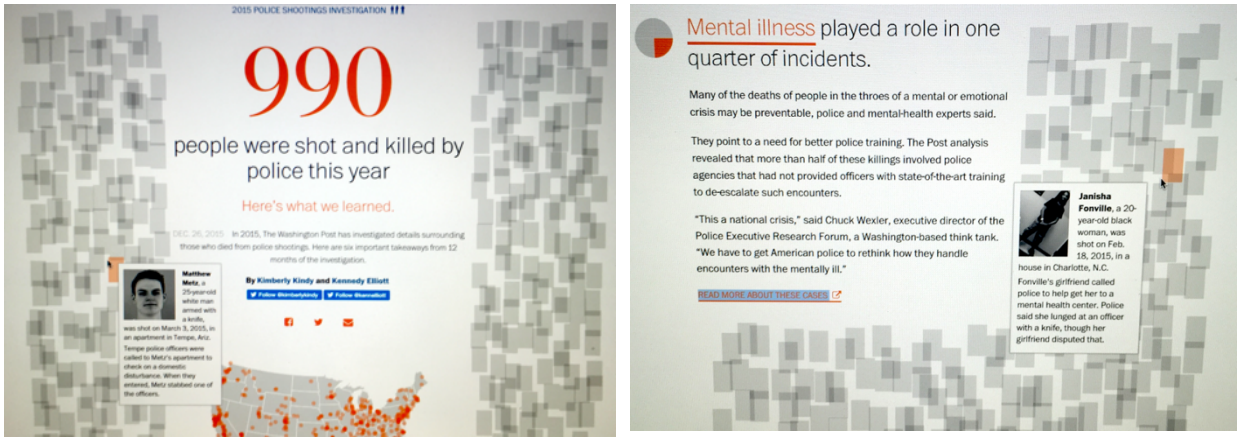


Figure 4.5. Human data frame analysis. *The Washington Post* provided interactive visualizations as framing context for their journalistic stories surrounded by the data.

"research and reporting," four for "production and presentation," two for photo, and two for video (*Washington Post*, 2015).

Police Shootings consisted of an assortment of more than 75 articles of varying types – features, news, visualizations – with no obvious central page, but a node connected to several different points on a distributed network within the paper, each with different assortments of links and text to the visualization as well as the stories related to fatal *Police Shootings*. The visualization employed a similar dashboard system for filters, with each shooting attached to a human shape lit up according to the filters, displaying additional information about the individual case. The *Post* also includes articles explaining the project's methodology, which I have included as part of this case. Additionally, article comments were part of my analysis of reader engagements, as were social media usages of the data.

Content Analysis

Being able to compare and contrast these two cases based on content provided the most direct means of answering RQ_1 and RQ_2 (what mechanisms online were churning the data

journalism machinery, and what sort of impact did they have on relationships with the reader?). With Herring's (2010) guidance on content analysis for the Internet, I was able to break down pages into relevant components. These included *text*, *images*, *links*, *videos*, *features*, *graphics* – all six of which work interconnectedly to offer insight across various disciplines. To understand how these content components mediated interactions in functional and transactional relationships between journalist, reader, and data, I operationalized a series of conceptual themes critical to explanation of the data journalism phenomenon as studied by scholars and practitioners.

With a case study according to Yin (2003), and content analysis tools from Herring (2010), I built a framework according to Coddington (2015), and added thematic elements from other literature (Domingo et al., 2008; Lesage & Hackett, 2013; Tandoc, 2014). This painted a rather comprehensive look at data journalism connections that potentially affected relationships with the reader.

I examined from each of these two cases a generally parallel and representative sample of interactive graphic pages presenting the data, and text-heavy pages explaining the data collection process (Framework), articles analyzing the data or reporting news associated with them (News Analysis), and pages found on Facebook and Twitter (Social). These three categories aligned with Domingo et al.'s (2012) findings of audience participation varying in news production (Framework and News Analysis), news interpretation (comments), and news distribution (social) spaces. I visited these Web destinations (URLs) on multiple times on multiple devices during eight months between September 2015 and April 2016. Overall, I scanned hundreds of Web pages in an effort to identify overlap, and eventually settled on 12 particular Web destinations for each case based on parallel existence between the two cases in function and purpose. (See Appendices A and B for a list of URLs examined in depth). In some situations, where I was

evaluating a template, I scoured multiple pages until redundancies made the structure of the template obvious, which I was able to later confirm through journalistic fact checking with primary sources. Most thematic elements in the codebook were developed inductively from the literature, with pages determined by explorative analysis of a subsample from 2015. With the presentation of results, I attempted to provide further insight into the analytical relevance of structural components in relation to thematic elements and categories.

This typology identified mentalities underlying data journalism as a process. To operationalize this inquiry, I drew from both Coddington (2015) and Segel and Herr (2010) to synthesize concepts. In total, I coded for more than 45 thematic elements at different points of overlap and reassessment, noting either presence (2), partial or conditional presence (1), or absence (0). I went through multiple rounds of categorizing and recategorizing until settling on 32 thematic elements that operationalized the dichotomous categories established by Coddington (2015). Using his categories and sub-categories, operationalized with four thematic elements each, I created a simple scale to compare *The Guardian* and *Post* as distinct forms of data journalism, the same way he compared data journalism, computational journalism, and computer-assisted reporting as distinct forms of quantitative journalism.

Table 4.2. Coding sheet. On the right, 32 thematic elements conceptualized on Coddington's (2015) typology (right) to operationalize his dimensions (left), with summary of new binary decisions.

Dimensions	**Elements**
Professional Orientation **Prof. Expertise vs. Networked info** Coddington (2015), Lewis (2015), Lesage & Hackett (2013)	- Anecdotal storytelling - Expert analysis - Official sources - Story > Data - Participatory openness - "Wisdom of the crowd" - Social media sourcing - Platform > Narrative
Openness **Opacity vs. Transparency** Coddington (2015), Lewis & Usher (2013) Lesage & Hackett (2013), Lewis (2012)	- Downplays data process - Journo access > public access - Little explanation of data sources / methodology - Avoids linking outward - Shows methodology - Provides access to data - Helpful link transparency - Provides additional documentation
Epistemology **Sampling vs. Big Data** Coddington (2015), Tufte (2001), Klein (2015), Mayer-Schonberger & Cukier (2013)	- Targeted sampling - Detailed explanation of methods / limits - "Bulletproof" facts / function - Claims statistical or journalistic rigor - Robo-friendly data - Algorithmic accountability - Skilled use of computational tools - Pursuit of N=all
Vision of Public **Active vs. Passive** Coddington (2015), Segel & Herr (2010) Domingo et al. (2008), Lesage & Hackett (2013), Cairo (2015)	- Highlight violations of moral order - Journalist held as authority - Primarily 1-way engagement - Readers seen as unitary public - Participatory newsgathering - Participatory sharing (distribution) - Personal user utility tools - 2-way engagement with accessible journalists

Research Steps

Using Coddington's typology as a framework, I operationalized concepts from different research trying to apply theories to data journalism where I noticed overlap. Table 4.1 shows how I coded thematic elements beneath his four dimensions (professional orientation, openness, epistemology, vision of public), each characterized by a dichotomous tension. The code sheet transformed into this spreadsheet provided a checklist. Below is a 5-step manual coding process, for each page in each case (Appendices A and B).

Step 1: Identifying component parts

I scanned each page to assess its primary purpose. My analysis focused usually on text or images first, then on links, then on features, etc., until I had looked at and noted the usage of each of the six components drawn from Herring (2010).

Step 2: Considering thematic elements

With knowledge of each page and its components, I assigned a 0, 1, or 2 to each thematic element on each page unit. The coding sheet provided my checklist. I returned to these codings several times to maintain intracoder consistency. With the numerical code on the code sheet, I also took note in salient examples to describe the character of these mechanisms.

Step 3: Marking up the coding sheet

On Feb. 15, 2016, I conducted a pretest marking up the code sheet to see if I could replicate the ability to glean insights similar to Segel and Herr (2010). I tried a handful of

pages, some from *The Guardian*, some from *The Washington Post*, as well as one from Twitter to identify coding elements that were proving difficult in operationalization. This tri-shaded approach also provided another route for adding validity to my study should I chose to bring in a second coder to verify my results.

Step 4: Recode and recategorize

After going through the above process multiple times, I returned to my notes and wrote up a single paragraph or bulleted list summarizing the key points of the page, highlighting codable words for use according to Table 3.1. For a sample notetaking on a coded page, see Appendix E.

Step 5: Revisit and note changes

Changes to Websites are inevitable in the course of research (Lewis and Usher, 2013; Schneider and Foot, 2004, McMillan, 2000), and regular changes through iteration and tinkering can be problematic. But limiting to a very specific and isolated time span could create an unreal picture, thus it was important to revisit and note key changes.

Summary of Methods

These methods provide a case study validated by Yin (2003) that uses content analysis tools that stem from Herring (2010). I looked to Coddington (2015) for a framework, which I was able to operationalize from his own literature along with works explaining crowdsourcing (Onouha et al., 2015) and gatekeeping (Bivens, 2008; Domingo et al., 2008) and participatory

means of audience engagement (Tufte, 2001; Lesage & Hackett, 2013; Mayer-Schonberger & Cukier, 2013).

There were limits to my study due to the enormity of content and complexity of connections. I recognized the challenges to reliability due to the personalization of Web experiences, as well as the irregular pace for the regular changes to content being evaluated. I addressed this limitation by making particular note of changes as I noticed them, and making these changes part of the study. Where this limitation proved most troubling was in my own screen grabs – where I would have benefited from having a more systematic approach to grabbing screen captures for evaluation and presentation in this thesis.

Results explicated in the next chapter revealed structures that exist in the digital realm, and showed more precisely how they were handled by two distinctive publications in ways that either enabled or constrained reader engagement through these mechanisms. These results provided points for discussion in my closing chapter about the future landscape of data journalism as a means of engaging readers and challenging policymakers while holding government to account.

CHAPTER FIVE: RESULTS

Having two respected journalistic organizations simultaneously pursue the same essential data set presented a unique opportunity to study their approaches. Using methods described in the previous chapter, this study successfully identified similarities and differences between *The Counted* and *Police Shootings*, which helped answer the research questions to provide a clearer picture of data-journalism dynamics. My research sought to address a dearth of empirical knowledge about digital structures in data journalism that potentially mediated relationships between news enterprises and the public. Analysis found that these two cases shared data sources, borrowed visual concepts, and made comparable tools available – suggesting they simply replicated data collection. Yet upon close reading, evidence showed their approaches to be quite different. Despite similarities in their data journalism as an end product, differences stood out in their data journalism when considered as a process.

In addressing RQ_1 (*What mechanisms exist in* The Counted *and* Police Shootings *that potentially mediate engagement between readers, journalists, and data?*) and RQ_2 (*How do these two exemplars of data-driven journalism enable or constrain reader participation in news construction processes? And why?*), content analysis allowed unobtrusive observation of how news environments were shaping engagements between readers and data. By adapting the coding sheet (Table 4.3) to a spreadsheet format (Table 5.1), I was able to answer these research questions using a color-coded notation borrowed from Segel & Herr (2010) to visually distinguish codes on quick glance.[12] These codes also provided a numerical basis for creating a

[12] Segel and Herr (2010) deconstructed visual representations of data in newspapers to instruct technical specialists about graphic design for narrative structures. Their design-space analysis proved particularly beneficial in its method for charting usage of different functions within a Website. They used a +, -, or 0 system for pattern matching to reveal insights into commonalities and oversights.

scale that compared relative presence of concepts according to Coddington's (2015) dichotomous categorical dimensions (Figure 4.1).

As coding got underway, more than 50 different thematic elements accounted for tangible concepts identified in literature from both scholars and practitioners surrounding crowdsourcing, gatekeeping, and user participation. After many passages of multiple pages, different overlap and redundancies became apparent, leading to an eventual settling of 32 thematic elements that created a balanced assortment. on a spreadsheet with a 0-axis under each dimension, splitting the binary choices with negative numbers on the left side of each dichotomy, and positive numbers on the right. By using the 0-1-2 coding system, then treating these totals as positive or negative depending on which side of the dimensional spectrum they fell, a numerical average (calculated in the gutter of each binary in Table 5.1) could be translated to a graphical image of relative proportion. To stay consistent with Coddington (2015), positive numbers became arrows pointing rightward, negative numbers pointing leftward, with arrow size determined by the relative magnitude of the number.[13] Though not statistically precise, these visualizations were acceptable for explaining an outlook according to what the use of different elements revealed.

Squares in Table 5.1 also helped guide the content analysis, with repeated passes over different pages helping improve the validity of the study. Note that thematic elements can continually be improved. A benefit of Herring's Web CA is that it allows for a plug-and-play analytical approach. Additional coders would add more precision and granularity to the research, but as it stands the thoroughness of the approach taken here provided a validity to these results by attempting to minimize the influence of any one particular error of omission or subjectivity.

[13] Though methodological detail is limited, the size of Coddington's arrows came through what was effectively stack analysis of the 90 different texts he assessed.

Table 5.1. Coding sheet operationalized according to Coddington's typology.

	Dimensions		Professional Orientation									Openness									Epistemology									Vision of Public								
	Binary choices		Prof. expertise					Networked info				Opacity					Transparency				Sampling					Big Data				Passive					Active			
Page	Page type	THEMATIC ELEMENTS	Anecdotal authority	Defer to official source	Expert analysis	Story > Data		Participatory openness	Tools to personalize	"Wisdom of the crowd"	Platform > narrative	Downplays data process	journo access > public access	Presents external data with little	Avoids linking outward to sources		Shows methodology	Provides access to "rawer" data	Transparent Links	Additional documentation	Targeted Sampling	Detailed explanation of methods and	Bulletproof data prez	Reveals Statistical / Journo Rigor		Robo-friendly data	Algorithmic accountability	Skilled use of computational tools	Pursuit of *N* = all	Highlights violations of moral order	Journalist given authority	1-way engagement	Seen as unitary public		Participatory newsgathering (UGC)	Participatory news distribution	Tools for personal User-utility	2-way engagement
GUA DB	Framework		0	0	0	1	7	2	2	2	2	0	1	0	0	5	2	1	2	1	1	0	2	1	3	2	1	2	2	1	0	0	0	5	1	1	2	2
GUA Datum	Framework		2	1	0	1	0	1	1	1	1	1	0	0	0	6	2	1	2	2	1	1	2	0	3	2	1	2	2	1	1	1	0	2	1	2	1	1
GUA Send a Tip	Framework (form)		1	0	0	1	3	2	1	1	1	0	1	0	0	5	2	0	2	2	0	2	2	1	0	1	1	1	2	1	1	1	1	2	2	2	1	1
GUA About	Framework		0	1	0	1	3	1	1	2	1	0	1	0	0	7	2	2	2	2	0	2	2	1	1	1	1	2	2	1	1	1	1	1	2	1	1	1
GUA Read Articles	Framework		1	2	2	1	0	1	2	2	1	1	1	0	1	3	1	1	2	2	0	0	1	1	2	1	0	2	2	2	1	2	1	-3	1	1	1	0
GUA Race Frame	News Analysis		0	0	1	0	4	1	1	1	2	0	1	0	1	4	2	1	2	1	1	1	1	1	4	2	2	2	2	2	2	1	0	-2	1	1	0	1
GUA The County	News Analysis (video)		2	1	1	1	-1	1	1	2	1	0	1	1	0	4	1	1	2	2	1	1	1	0	2	1	1	2	1	2	1	0	1	0	1	1	1	1
GUA Op-Ed	News Analysis		2	2	1	1	-3	1	0	1	1	0	0	0	1	3	1	0	2	1	0	1	1	1	3	1	1	2	2	1	1	1	1	-2	1	1	0	0
GUA News	News Analysis		2	2	0	1	-2	1	0	1	1	0	1	0	1	1	1	0	2	0	0	1	2	0	2	1	1	1	2	1	1	1	1	-2	1	1	0	0
GUA FB	Social		1	0	0	1	2	1	1	1	1	2	1	0	1	1	2	0	2	1	0	1	1	1	3	1	2	1	2	2	1	0	0	4	2	2	1	2
GUA FB Vid	Social (video)		1	1	0	1	2	1	1	2	1	0	0	0	1	7	2	1	2	2	1	0	2	0	1	0	2	1	1	2	2	1	0	1	2	2	1	1
GUA Twitter	Social		2	0	0	0	3	2	2	1	0	1	1	0	0	3	1	1	2	1	0	1	1	1	3	1	2	1	2	2	1	0	0	4	2	2	1	2
							1.5									4.1									2.3									1.2				
							-2.8									0.5									1.3									-2.6				
WP HP	Framework		2	1	1	2	-5	0	0	0	1	2	2	0	2	-2	1	1	1	1	1	2	1	1	2	2	2	1	2	2	2	2	2	-6	1	1	0	0
WP About	Framework		0	2	1	2	-2	1	0	1	1	1	2	0	2	2	2	2	1	2	0	2	1	1	0	0	1	1	2	1	2	2	2	-4	0	1	1	1
WP Form	Framework (form)		1	0	0	1	1	0	1	1	1	1	1	1	1	2	2	1	2	1	0	1	2	0	2	2	1	1	1	1	2	1	1	3	2	2	1	1
WP Viz	Framework		1	1	0	1	5	2	2	2	2	1	2	1	2	-2	1	1	1	1	0	1	1	0	6	2	2	2	2	1	1	1	0	1	1	1	2	0
WP Body Cam Vids	Framework (video)		2	2	1	1	-3	1	0	1	1	1	1	0	1	4	2	1	2	2	0	2	2	1	0	1	1	2	1	2	1	1	1	-2	1	1	1	0
WP Body Cam Vids	News Analysis (video)		2	2	2	1	-3	2	0	1	1	1	1	1	1	1	2	1	2	0	0	2	2	0	0	0	0	2	2	2	2	2	2	-5	1	1	1	0
WP Race Frame	News Analysis		2	0	2	1	1	2	1	2	1	0	0	0	1	6	2	2	1	2	1	1	0	1	4	1	2	2	2	2	1	2	1	-2	1	0	2	1
WP Mental Illness	News Analysis		2	2	2	1	-5	0	0	1	1	1	2	1	0	-1	1	1	1	0	1	2	0	1	2	2	2	1	1	2	1	2	2	-5	1	1	1	0
WP Profile of Killer	News Analysis		2	2	2	2	-6	1	0	1	0	1	2	0	1	-2	1	0	1	0	1	1	1	1	1	1	2	1	1	2	2	2	2	-4	1	1	2	0
WP Police Training	News Analysis		2	2	2	2	-7	1	0	0	0	2	1	0	0	-1	2	0	1	1	1	1	0	1	-2	0	0	0	1	1	2	2	2	-4	1	1	1	0
FB WP Investigatio	Social		2	2	2	2	-8	0	0	0	0	1	1	1	1	0	1	1	1	1	0	1	0	1	0	0	1	0	1	1	2	2	1	-2	1	1	1	1
TW Post Investigate	Social		1	1	0	2	-2	1	1	0	0	2	1	1	1	-1	1	1	1	1	0	1	0	1	1	1	1	0	1	1	2	1	1	-1	1	1	1	1

Note: GUA= *Guardian*, WP = *Washington Post*, FB = Facebook, TW = Twitter. White lines separate dimensions. Gutters between each binary choice represent a 0-axis, with numbers added and averaged to create a comparative scale. For a full list of URLs, see Appendices A and B.

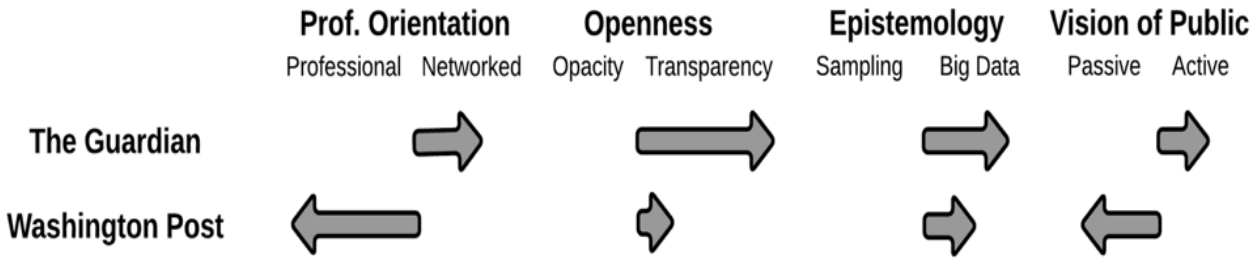


Figure 5.1. Coddington comparisons. Visualizing results of Table 5.1 revealed *The Counted* as a data journalism standard-bearer while the Post maintained commitments to CAR in professional orientation and a vision of the public as passive.

The arrows in Figure 5.1 revealed that *The Guardian*'s *The Counted* as near identical to Coddington's definitional depiction of data journalism (Figure 3.1). The *Post*'s *Police Shootings*, meanwhile, showed CAR-like tendencies. Coddington had attributed such leanings in CAR to personal philosophies on computing formed in an earlier era, when CAR was thought of as an investigative tool in the newsroom, handled by a privileged few. Below are explications of further comparative findings along each dimension, leaving the discussion chapter in this thesis (Chapter 6) to surmise how and why these two publications have such orientations and what that means for journalism and society. In this report, annotated screen captures provide a visual guide to similarities and differences between the cases. My goal was to highlight digital mechanisms, default settings, and process decisions that were both exemplary and problematic in how they enabled or constrained a reader's engagement with both the data and the journalism being produced about people killed by police. I chose examples of feature, visual, and textual structures from among dozens of possibilities to provide a broad but representative sample of identifiable traits within the content comprising each case.

1. Professional Orientation

Discerning this dimension borrowed from Coddington (2015), Lewis (2012), and Lewis and Westlund (2015) to assess the source of each publication's credibility. *Police Shootings*

established credibility through credentialed connections, while *The Counted* claimed credibility through connectivity to a broad swath of readers.

The Power of "You"

The Guardian distinguished itself by making this data journalism project an experiment in reader engagement (Smith, 2015). *The Guardian*'s most salient distinction highlighted in multiple places throughout *The Counted* – "you" – indicated its crowdsourcing approach and appeal. On the front page of articles, in the top-right, beneath the glow of *The Guardian*'s logo in the dark, "*The Counted*," in bold and cautionary yellow, "People killed by police in the US, recorded by *The Guardian* – with your help" (Figure 5.2). This mantra was repeated throughout many (but not all) Counted-branded pages. On the About page, on social media pages, and from the authors in interviews (Onuoha et al., 2015), they made it clear that this was a reader collaboration product. They made *The Counted* a community effort and stated as much, with visible actions online to back up their commitment, stated first on the "About" page:

> Our intention is to progress to a verified crowdsourced system. We want you to inform us as soon as possible if you witness a killing by law enforcement officers or learn of one that has taken place. We want to hear from you if you have further information about a case already included in *The Counted*. (*The Guardian*, 2015)

The intangible spirit of cooperation showed itself in tangible form through textual choices. *The Counted* chose to prominently display active verbs encouraging participation – "join us," "send a tip," "join our community" were some. In addition to the text, its placement was permanently affixed as part of the primary navigation (Figure 5.3). By contrast, the *Post*'s interest in crowdsourcing was less pronounced – with the phrase "crowdsourcing" never mentioned on

Police Shootings Web pages, despite offering use of essentially the same tools as what *The Counted* offered its readers. But the Post places the ability to "send us tips and additions" in light gray, in a scrolling area of the page where readers are expected to keep scrolling.

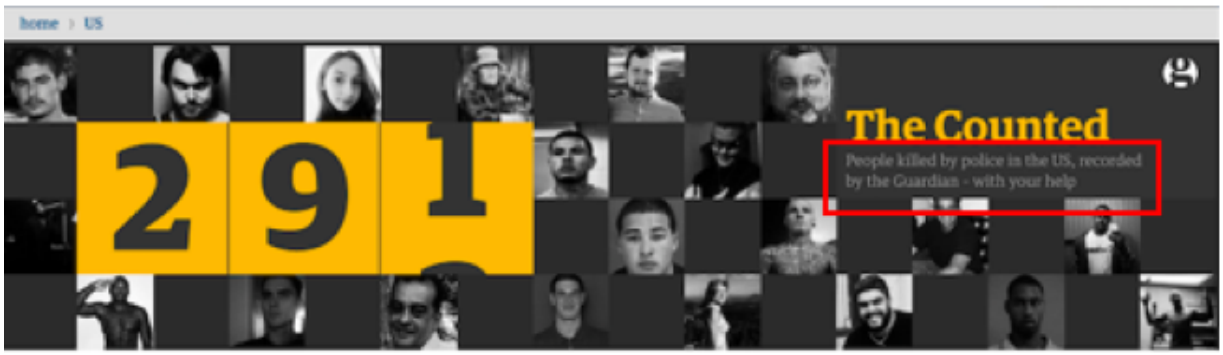


Figure 5.2. Title bar for *The Counted* articles. Showing readers that they were key to this process was a topmost priority for *The Guardian* US, with indicative text outlined in red.

Personalized Data Experience

My analysis found parallel areas worthy of study beginning with the front page of each publication's opening presentation of data. Scrolling and clicking through the representation of the data set gave information about the criminal justice story being told, and the data journalistic means of telling it. Both provided similar tools to their audience to manipulate the data. These dashboards were interactive. In these mechanisms, I found structural similarities and partial visual similarities. The reader-data relationship was mediated with filtering, hovering, comparative selections, and data-on-demand, these tools made for a pleasant visual and interactive experience, and allowed the reader to personalize the experience of engaging with the data. A reader might look, for example, on a map to see where cities relevant to me stood by comparison, and otherwise make the data more personally relevant (Hurrell & Walton, 2013). Such design tactics gave the reader more control.

Personalization refers to the act of customizing data to the reader's liking. This can be zooming into one's own area code, or entering one's demographic info, to name just a couple.

But in both the *Times* and *Guardian* cases, using data visualization tools to customize the data presentation did not require the user to provide any personal information, unless they chose to through registering with each publication. One apparent difference was with their use of social buttons. The semiotic assemblages of pixels looked virtually the same in both cases. But *The Counted*'s social buttons were surrounded by encouraging phrasing – "join us," with links to Facebook and Twitter pages for *The Counted*'s community. *The Post* used the same buttons, however, they did not take them to any new community or destination; they just provided tools in the form of a pop-up window allowing tweets or shares of *The Washington Post*'s work – but with no incentive or benefit to beyond any self-perception of smartness attached such sharing.

Personalizing the data was different than *personifying* the data, which also was present in both publications. The journalists personified the data by presenting them reconfigured in digital form as representative of a real person. "Personify" is what Loosen et al. (2015) were describing as personal details. Personifying data refers to elements that make the data set an individual and human person. They have a name, an age, a hometown. All these allowed readers to begin to understand the data on a people-based level, which is what journalism has always done with anecdotal examples. Examples of this personified data presented in personalized experiences were apparent in Figure 5.3, which showed how customizable dashboards were instrumental in such engagements.

2. Openness

Openness as a concept has driven data and journalism. The open-source movement was a favorite for technologists, and easily embraced by journalists (Lewis & Usher, 2013). This was a change from the days of opacity, where additional information got in the way of reader understanding data. But the Internet has changed that, as transparency itself is becoming a key to credibility. The apparent disparity seen between the two on this dimension could be a little

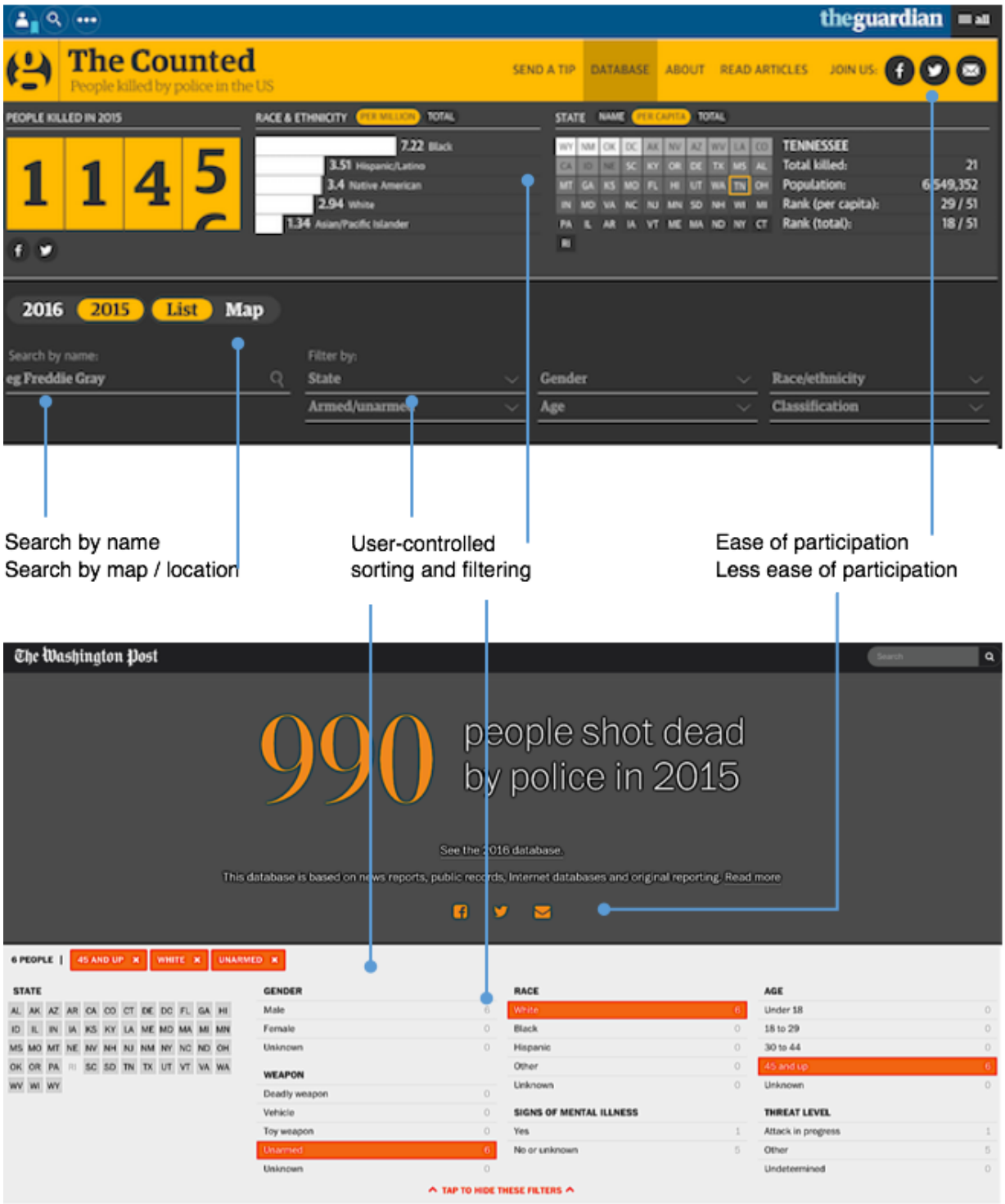


Figure 5.3. Reader control for private engagements with data. Interactive Dashboards personalized the data experience, where the news operation chose the initial set of data readers received. From there they were able to customize their interaction with data. (See Figure 5.11 for more details.)

misleading, as my scoring system penalized non-transparent links. Additionally, even though the *Post* was being transparent with its data, it maintained opacity behaviors in several places despite showing otherwise a commitment to transparency.

Mutually Transparent, Differently Enabled

Both cases showed dedicated commitment to what Lesage & Hackett (2014) identified as "transparency work" – a concept they defined as "making publicly available the sources, interests and methods that might influence the information presented, so that notionally readers/viewers (as rational subjects) can take potential bias into account" (p. 41). This was translated as sharing "raw" data, sharing methodology, improving access to data, and providing additional documentation, which included a willingness to link outward. My examination found mutual agreement on support for transparency as they conveyed it. *The Guardian* showed transparency right from its "About" page, as did the *Post*. *The Guardian* offered these characteristics clearly on *The Counted's* "About" page, which was easy to find among the top-level navigation (Figure 5.4). This page highlighted a one-click download of their "raw" data. A single click on "download the data" offered immediate access to a zip file, which was loaded with the latest version in easy-to-use CSV, Excel, and TXT files.[14]

The *Post* similarly attempted to be transparent and share their dataset. But they did so through third-party Website Github, which, before any downloading of data could begin, required additional clicks, membership signup, and other rules to follow that were complex and not clear without focused assessment of a separate communities' norms and vernacular. Likewise, files as presented were cumbersome and uninviting. Both "About" pages credited

[14] In the spirit of open-source, an example of someone doing good with shared data was FiveThirtyEight, which took *The Counted*'s raw data and overlaid them with crime statistics for different zip codes to identify potential correlations along geosocial lines.

sources, including KilledbyPolice.net and FatalEncounters.org. But *The Post*'s transparency did not transcend to openness when they did not include outbound links, leaving it to the reader to find the locations and visit. *The Counted*'s structures made it easy for engaged readers/users to do what they might want to do, while the *Post*'s structures provided complicated but navigable paths for readers/users most committed to accessing rawer data.

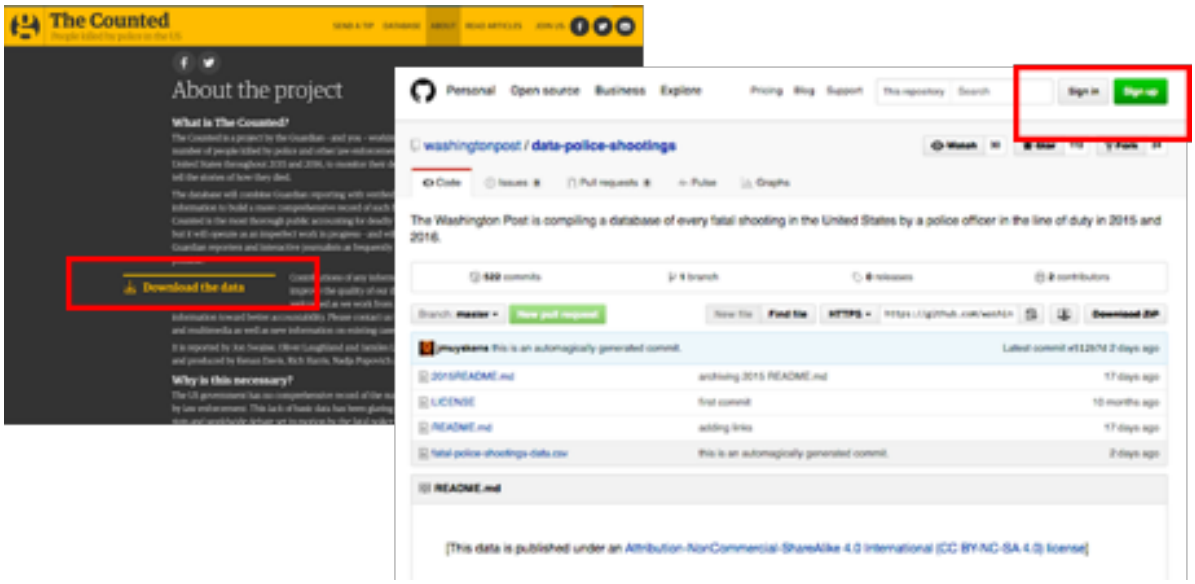


Figure 5.4. Data sharing enabled or constrained. Both sources make their datasets accessible, but *The Counted* (left) was significantly more accessible in a way that was beneficial to data researchers.

Beyond perceived credibility associated with transparency (Lesage & Hackett, 2014), in comparing these two projects as products and processes of data journalism, credibility of the data had to be considered. The *Post* claimed their data came "by culling local news reports, law enforcement websites and social media and by monitoring independent databases such as Killed by Police and Fatal Encounters. The *Post* conducted additional reporting in many cases" (external links excluded). *The Guardian* claimed: "So far, we count with traditional reporting on police reports and witness statements, by monitoring regional news outlets, research groups and open-source reporting projects such as the websites Fatal Encounters and Killed by Police" (external links included).

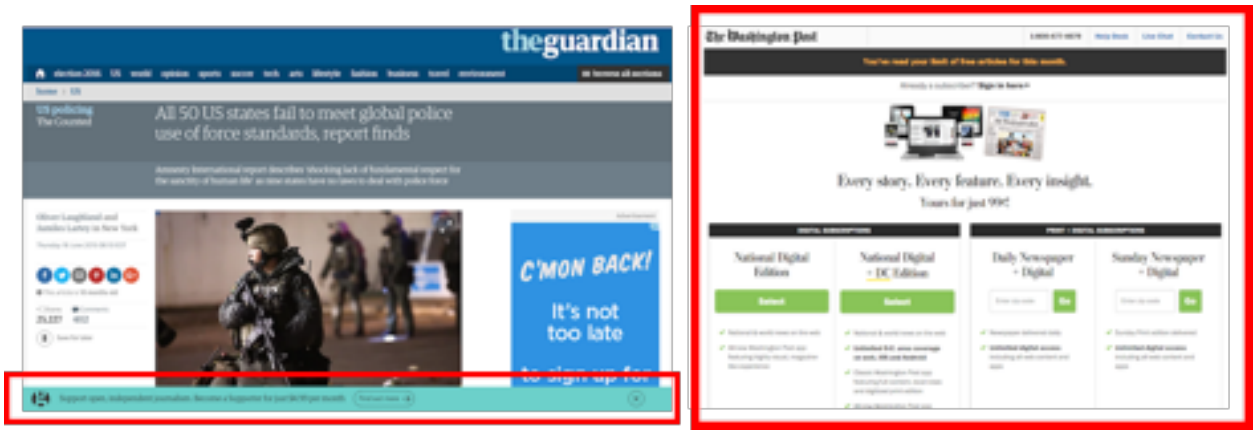


Figure 5.5. Two types of paywalls. *The Guardian* (left) offers an optional donation request, with an easy-to-dismiss link or more information about *The Guardian*'s business model. The *Post* (right) enforces a limited free sampling on most pages.

Paywall Opacity

As media businesses, particularly legacy newspapers, try to figure out how to monetize costly journalism, two different business models existed in these cases, with one literally on the side of opacity and the other on the side of transparency. *The Guardian* ran on a non-profit model, supported by the Scott trust. They put a brightly colored bar at the bottom of a page – hard to miss but easy to dismiss – that invited readers to contribute. It took up 5% of the screen (measured in pixels) and used common semiotics, an x to easily close, and an arrow to go to, which took clickers to a page explaining *The Guardian*'s commitment to watchdog journalism. This was very different from *The Washington Post*'s paywall, which took up 100% of the screen, with 100% opacity (as a setting), which created a gate in and of itself.

3. Epistemology

This dimension showed both cases in many ways embracing emergent ideologies and methodologies surrounding big data. This included a belief in the value of pursuing *N*=all (Mayer-Schönberger & Cukier, 2013) – which these databases are by their purpose and nature –

relying on correlations over statistical causality for stories, and showing skilled use of computational tools.

Structured Storytelling

Parasie and Dagiral (2012) saw as part of the shift to reader empowerment included a change in how stories are told. In a datafied world, stories maintained some author structure, but gave them additional control. At *The Guardian*, the reader is able to choose what to look at. This included the ability to hide stories. Whereas the *Post* still presented overall narrative in linear fashion. Figure 5.6 showed *The Guardian* on the left side, with stories placed in data squares by manual choice not algorithmic full, just as on the right the *Post* shows the stories that they chose to determine the narrative. Note, neither requires readers to click one over the other, but they revealed a subtle difference for control the publication maintains over narrative consumptions.

Narrative Framing Devices

Racial disparity or lack thereof was central to media coverage of people killed by police. This was true in editorial workspaces, in comment discussions, and in default presentations of data. *The Guardian* has been unapologetic in its presentation of racial frames. On the dashboard, *The Guardian* chose to make race a setting by showing killings by race in a per-capita context, whereas the *Post* let just numbers speak. (Figure 5.7). A certain racial framing was a constant in *The Guardian*, in the form of links, images, text, and feature settings. The *Post* showed a shifting relationship with race – starting with a feature specifically intent on personifying data about race ("Black and Unarmed," June 30, 2015), highlighting race in the top-level entry, and championing external study about race from their own data (Figure 5.8).

Figure 5.6. Narrative control. Visual analysis showed how *The Guardian* (left) presented story narrative as blocks, not randomly assorted, but easily readable by reader's choice. The *Post* (right) was more linear in its narrative rollout.

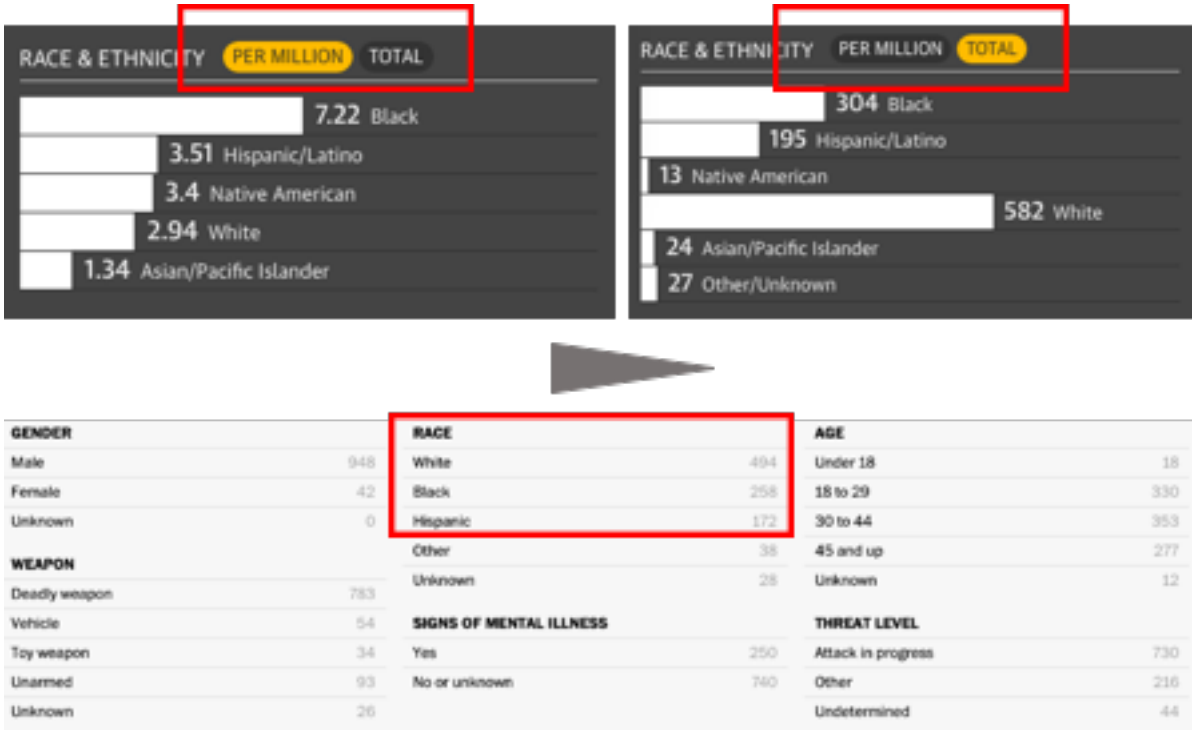


Figure 5.7. Default settings for data totals sorted by race. The view that greeted visitors of *The Counted* (top left) highlighted the situation as a matter of racial disproportion, telling a different story than the other option (right); whereas *Police Shootings* (bottom) presented numbers in a more raw form.

Data Personified

The building blocks of this data set were people. Individuals of all ages, races, mental condition, innocence, and guilt. This had a potential impact on narrative worthy of note. It also makes certain numbers more real. Figure 5.9 shows how each case transformed data sets into a real person, real individuals. Both revealed data being personified through images and worlds and interactive features, and both presented the totality such that it became a matter of data framing data, which seemed to have a strong potential for affecting both journalistic narrative and reader commentary narratives.

Figure 5.8. Narrative framing of race. The racial frame was key in *The Guardian*'s coverage (top), and was the first feature of data analysis in the *Post* (bottom left) – a frame they highlighted throughout much of 2015 (middle) but stepped away from in 2016, but not completely (bottom right).

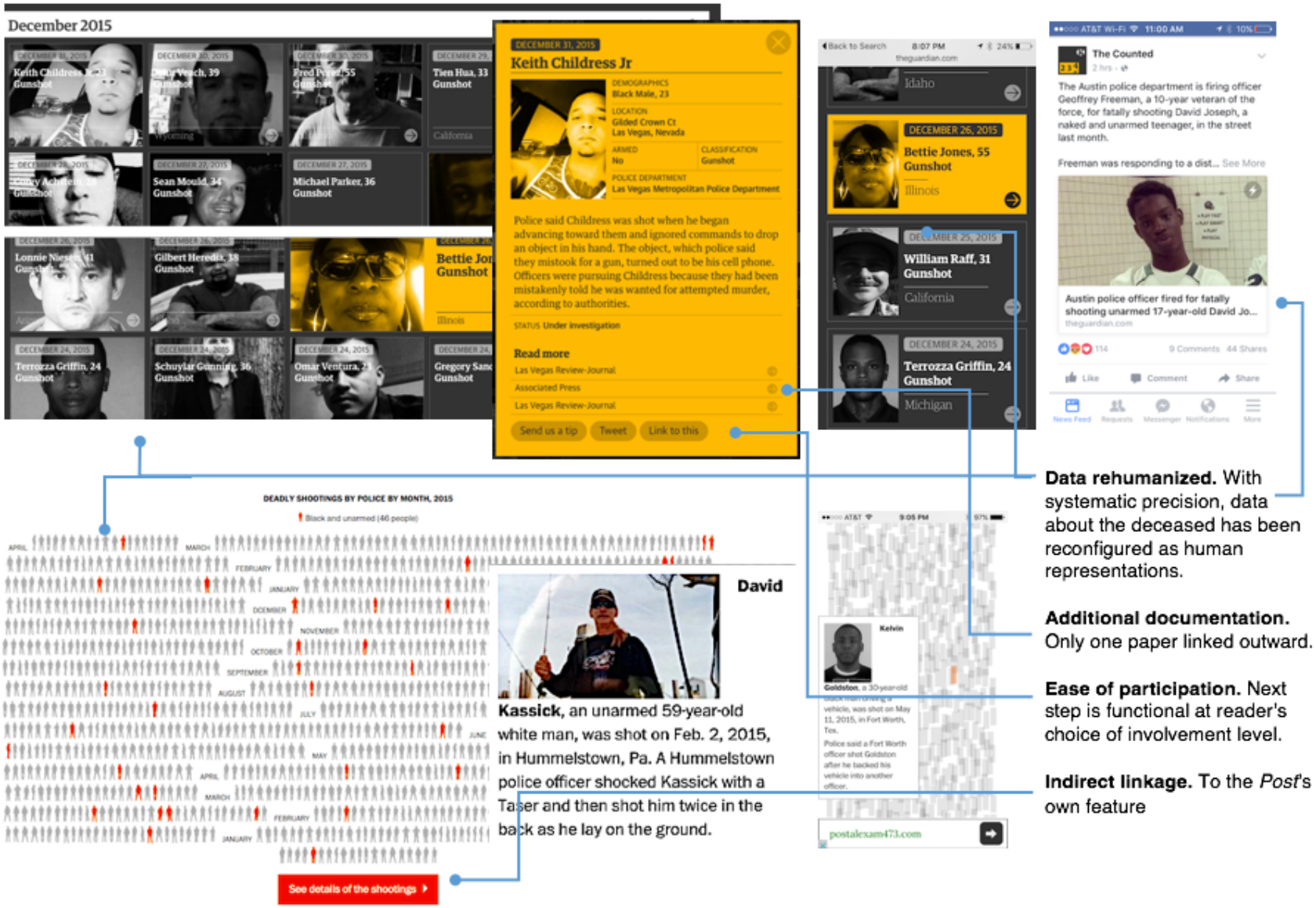


Figure 5.9. Personified data in active and passive news environments. *The Counted* (top) provided readers with tools to take actions, while *Police Shootings* (bottom) allowed readers to engage with data but not each other.

70

The *Post*'s narrative formula was replicable, with a name, age, race, armed, information about who shot, all sortable as data elements. With the addition of new data points came the potential for addition of new narrative turns. Interestingly, the *Post* gave readers more control over their private engagement with the data, but less control over the narrative. Mousing over a personified data set revealed additional differences. The *Post* gave links that took the reader/user to more of their work on the story. *The Guardian* engaged in similar self-promotion of their epistemology, but gave links that made the act of clicking to provide additional information a benefit to the reader seeking to engage in such a task.

4. Vision of Public

Exploring this dimension provided the starkest difference between the two publications – with *The Guardian* showing an active view of the public, and the *Post* seeing the public as more passive. These elemental codings were apparent on the publications' proprietary websites, but also extended to comments and social media pages. Public engagement tools were similar, but the engagements themselves were quite different. This vision of the public played out in multiple areas, with a difference between one-way and two-way communications, most notably visible in comments and then social media spaces. Overall, the publications' visions of the public became apparent as either passive or active from different thematic elements seen on their own sites' content, in comment spaces on their own sites, and in different social media engagement spaces.

Open and Closed Forums

Notable differences in participation occurred with comments, starting with the pages open to them. *The Guardian* seemed more discriminating about pages that provided readers a place to interact with the publication and each other. News articles that did have comments in most cases stayed open only a few days. (Three days for some, four days for others.) No

framework pages – the ones presenting data or explaining *The Guardian*'s methodology – provided a space on the same base URL for public discussion. *The Guardian* notably only left articles open to comments – but the data pages themselves as well as the About pages explaining the methodology were not open to public discussion. *The Post*, however, did leave many of these framework pages open for discussion, including its "About" page (for 10 days). The *Post* claimed by policy comments stayed open for two weeks. However, on most articles that appeared as part of their *Police Shootings* investigation, the *Post* violated its own policy and left comments open, as visible comment numbers indicative of engagement continued to climb.

Comment Participation

To participate as a commenter, both publications required registration. But what that meant differed for each. At *The Guardian*, all contributors were required to have a user profile, which was publicly accessible and by policy supposed to be attached to a real person (though pseudonymous profiles were tacitly allowed). Standard prohibitions against abusive commenter behavior applied, and readers were given tools not just to vote on the quality of comments but also to report violators of community standards.

Visible Moderation

When violations occurred, comment moderation occurred in public view, with links to contributor profiles maintained, and links to forum rules showing *The Guardian* actively enforcing standards. *The Washington Post* disallowed abusive conversations, but their required user profiles were not publicly accessible, meaning commenters could operate essentially anonymously. Anecdotally speaking, there were no swear words or racial slurs at the *Post* suggesting there was moderation, just not visible.

"There is more public awareness, and I do think that in the past few years the veracity of police officers is being questioned more, after their statements were shown to not be consistent with video evidence," - Damn right.

Report

RobDavy 31 Dec 2015 12:34

This comment was removed by a moderator because it didn't abide by our community standards. Replies may also be deleted. For more detail see our FAQs.

Maj_jak2509 31 Dec 2015 12:34

This comment was removed by a moderator because it didn't abide by our community standards. Replies may also be deleted. For more detail see our FAQs.

Danny885500 31 Dec 2015 12:35

This comment was removed by a moderator because it didn't abide by our community standards. Replies may also be deleted. For more detail see our FAQs.

Chuckman 31 Dec 2015 12:35

This comment was removed by a moderator because it didn't abide by our community standards. Replies may also be deleted. For more detail see our FAQs.

poechristy 31 Dec 2015 12:37

This comment was removed by a moderator because it didn't abide by our community standards. Replies may also be deleted. For more detail see our FAQs.

CommonSenseWelcomed 31 Dec 2015 12:37 23

Institutionalised racism... The African-American community have been shouting about this for decades and nobody listened....

America, you claim to be the "best", it's about time you joined the rest of the Western world in the modern age and actually tackled institutionalised racism, your whole society is riddled with racism. You cannot call yourself a "decent" nation while you allow Cops the ability to slaughter innocent, unarmed black people.

Report

All Comments

Newest First

Pause live updates

GBrouter

4/3/2016 1:34 PM PDT

Cops have near total discretion on use of their weapon.
DAs have near total discretion on who is charged with a crime.
Judges have near total discretion in many areas (you think criminal is corrupt, look at civil. Judicial discretion grows exponentially in civil court).
Attorneys - license to steal.

I vote Judicial as most corrupt branch of government. "Bench and Bar Buddies Club" protects cops, attorneys, and judges to the maximum limit allowed under the law, plus some more.

Like Report Reply Share

enderrorism

[illegible] PDT [Edited]

We need more investigators and DAs and some that work primarily on officer homicides.

There are simply far too many unsolved murders in the US, not to mention, the uninvestigated largest mass murder in the history of the US, the collapse of the WTC.

There is no labor shortage in the US, and yet we have probably half the number of DA's and DA investigators required to deal with caseload.

https://www.youtube.com/watch?v=v-27FGbpBk4

Like Report Reply Share

Hunter_Bluff

4/3/2016 5:56 AM PDT

Somehow other countries manage not to blow away their citizens at the same rate. Of course, those countries aren't awash in guns either. Coincidence? No.

Figure 5.10. Comment-section disparities. *The Counted* (left) showed *The Guardian's* community-based approach, which included visible moderation – apparent and attachable to real user profiles. In The *Post*'s *Police Shootings* comment threads, moderation went unseen and pseudonymous user profiles not publicly viewable created virtual anonymity.

73

Social Engagement

Clear differences were apparent in the social media feeds for *The Counted* and *Police Shootings*. For *The Counted,* the social experience was fundamental, with *The Counted*'s Facebook and Twitter pages going live June 1, 2015, the same day *The Guardian* US released its data set and initial stories from it. These pages set a tone from the start. The *Post* did not create separate accounts for *Police Shootings*, instead incorporating the special project into previously set social media streams for its Investigations unit. Close examination of social streams produced valuable data indicative of the relationship the publication sought to maintain with its readers. For the *The Counted* it became a hub of their community. For the *Post*'s Investigations home for *Police Shootings* activity on Facebook they ended up abandoned.

Two-way Conversation Control

Two-way conversation was most evident on social media. On both Facebook and Twitter, *The Guardian* showed a greater commitment to creating two-way engagement with their readers, with staffers speaking as humans, but under the institutional moniker. The *Post* again provided the tools for the same conversation – email addresses directly – but there was no mandate to respond.[15] *The Guardian* used Facebook as something of a community hub for *The Counted*, serving more than 20,000 fans who opted in via "like" to have Facebook include posts from *The Counted* regularly appear in their personal feeds. Their Facebook "About" page: "Think of this page as a direct line to *The Guardian* US' journalists." And while that may have been the initial intent, it showed many signs of enabling such engagement, it did not create a truly direct, or

[15] Though little more than anecdotally suggestive, this researcher got no response from multiple attempts (using multiple provided tools) to establish contact with Washington Post journalists regarding *Police Shootings*. *The Guardian* was immediately responsive on all accounts.

unencumbered route to *The Guardian*'s journalists – establishing a two-way conversational thread and a place for readers to participate in crowdsourcing activities. Also, this created a place to establish certain gates, as engagement came from *The Counted* as an institution, not as a named individual. *The Counted* maintained a standing that "you," the readers – particularly active readers – were a top priority. They also were highly responsive – in comments and direct message – establishing personal connections that conveyed transparency (Lesage & Hackett, 2014). When communicating institutionally, readers could expect their personalized response to be signed by "*The Counted* team." Their Facebook location, promoted on their Twitter feed (but not mentioned on *The Guardian*) put forth a perception of engagement with the individual authors, even though the direct engagement was open to occur for only a limited period of time. By holding a Question-and-Answer session, Facebook became a community destination for individuals interested in *The Counted* itself, as a process. This showed an incentive for being a more involved. Engaged followers had a chance to communicate more directly. Collectively, it created an institutional "gate" between *The Counted* and readers, allowing the journalists to maintain a certain elevated stature and authority. *The Guardian* had systems in place to eventually put community members

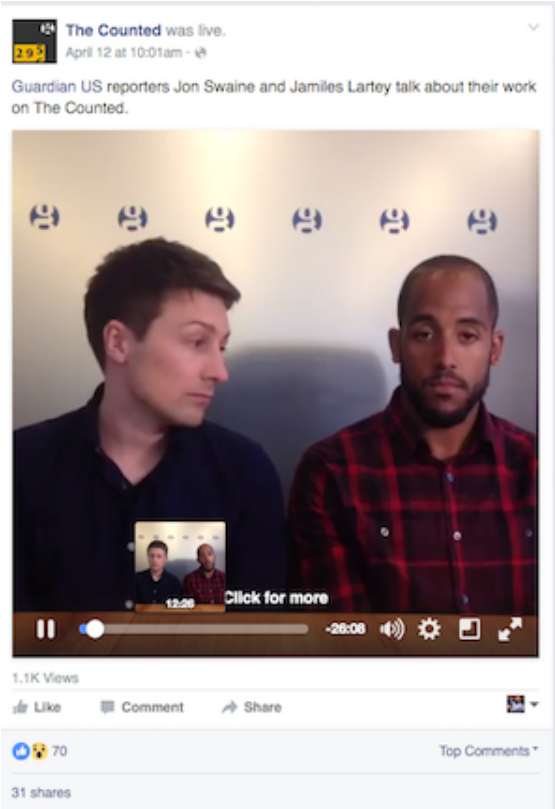


Figure 5.11. Two-way conversation on Facebook. Jon Swaine and Jamiles Lartey appear on video to take questions about *The Counted* from their audience.

in touch with individuals at *The Guardian US* office, similar to a traditional front-desk receptionist or an executive admin gatekeeping role.

As for community rules, Facebook has its own Terms and Conditions and enforcement standards, but *The Guardian* restated its own expectations in a separate "About" Page, prominently linked to above-the-fold on their Facebook Page, reminding the community to take care of itself (Figure 5.12). They also showed limited but consistently present two-way communication, with an institutional incarnation of "*The Counted* Team" expressing gratitude to support and encouragement, personifying the journalists while maintaining an institutional gate on neutral territory, and also capable of showing sensitivity while maintaining political distance when dealing with families of the deceased. All the while *The Guardian* staff reinforced branding, showing consistent responsiveness with expressions of appreciation and concern, while also responding to criticism and working toward truth in partnership with the readers.

Conscious Interactive Neglect

The *Post*, by contrast, did not have separate social media spaces for *Police Shootings*. Instead, the *Post* included data journalism about people killed by police as part of their "Investigations" team section, which identified itself as a "newspaper" not a community. The page was kept "alive" by automated postings of new content into their timeline, which kept a handful of commenters engaged, but there was no re-engagement, and otherwise the page spoke more to the *Post*'s inadequacies in social spaces, despite indicators of transparency. Ultimately this page was populated by an automated news story feed, non-engaging comments.

Signs of abandonment were all over this page. For example, the video section showed an idea not materialized nor cleaned up. Their page displayed two videos, each showing zero views. This suggests the setting is to not show views (a Facebook video option), as my personal viewing

from multiple devices across multiple times would have ticked the view count. The primary video consisted of a roundtable discussion, featuring Bob Woodward and other *Post* editors. The video is labeled as an "exclusive," which proved to be true as this researcher's efforts to locate the conversation elsewhere on the Web turned up no other copies. The second video, posted "over a year ago" according to Facebook's date-stamp algorithm, was likely a placeholder, marked by an unrelated, non-investigative instructional Texas Hold'em video. In the Photos section, two random photos from the *Police Shootings* feature show an attempt to mimic what *The Guardian* was doing, but discontinued after two images (Figure 5.14).

Below the fold, a notes section featured postings from 2009, with a handful of articles cut and pasted from the *Post*, and on the front page one article titled "Obama's Quandry" (sic) and one titled "Featured Advertiser," which takes clickers to a blank page within the Facebook posts framework. Below that were visitor posts, which were indeed up-to-date. However, here they provided the tools for readers to engage, but without direction or engagement to shape that reader experience. and visitor posts that had little to do with investigative journalism instead became more of a bulletin board on a public kiosk.

The *Post*'s experience was similar on Twitter. In a nod to transparency, David Fallis is named as maintaining it, but there was little engagement beyond automated feeds. And these feeds when made shareable by others from their own *Police Shootings* Web pages actually put out incorrect data for more than three months into 2016. When the year turned, an algorithmic error in the *Post's* social sharing feature from individual entries on its main data page had any sharing promulgating inaccurate information by attributing 2016 deaths to 2015, and therefore adding to the total number of 2015 deaths inaccurately. This lasted for at least three months.

Figure 5.12. *The Counted's* Facebook community. A "living" location with active engagements more frequent and regular than stories being added to the news articles database.

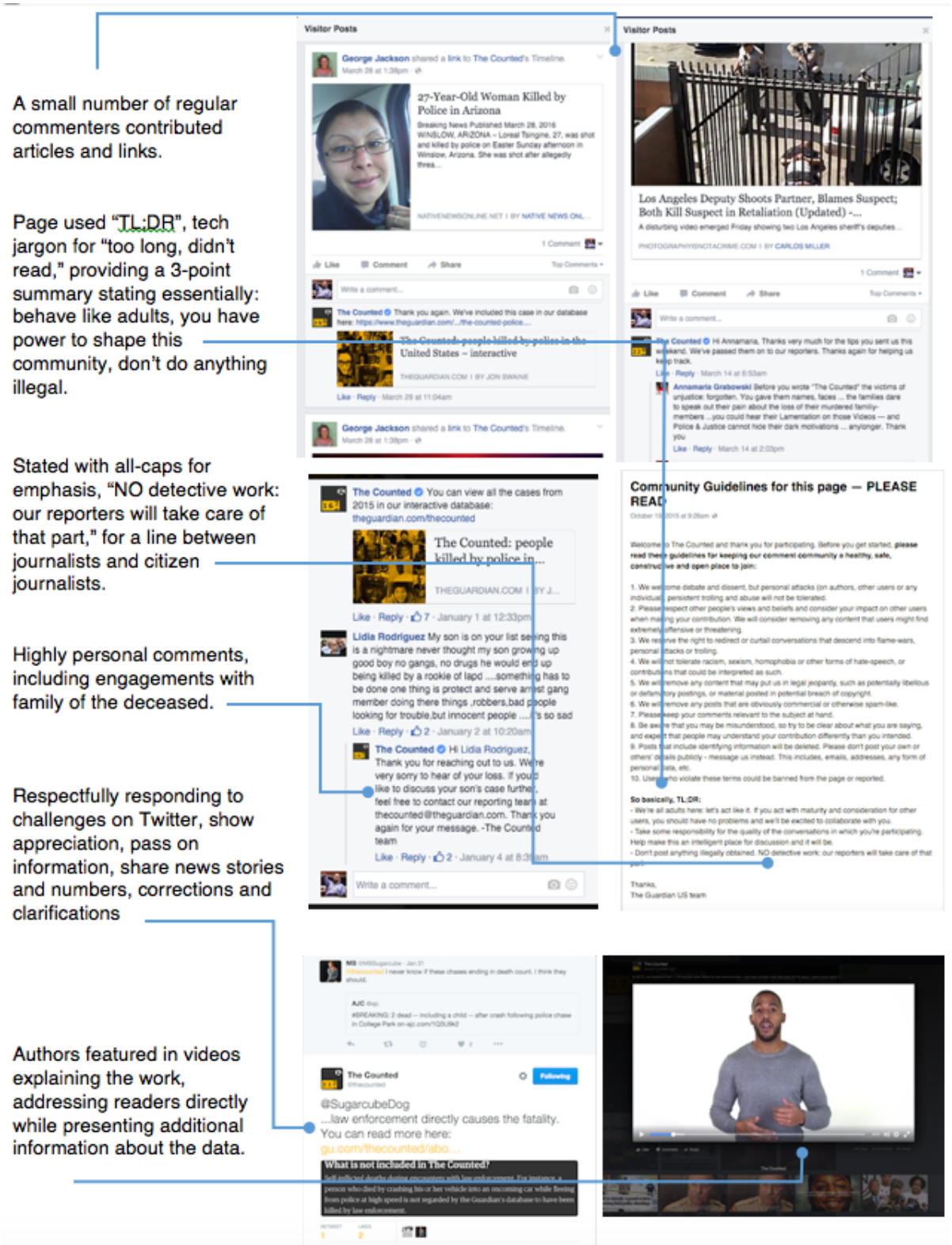


Figure 5.13. Personifying media personnel in the reader's personal space. *The Counted's* social media engagements explained to readers what they were trying to create using rules of the platform to maintain it, rewarding readers who participated as they desired with two-way engagement

Figure 5.14. *The Washington Post* Investigations team as a newspaper on Facebook.

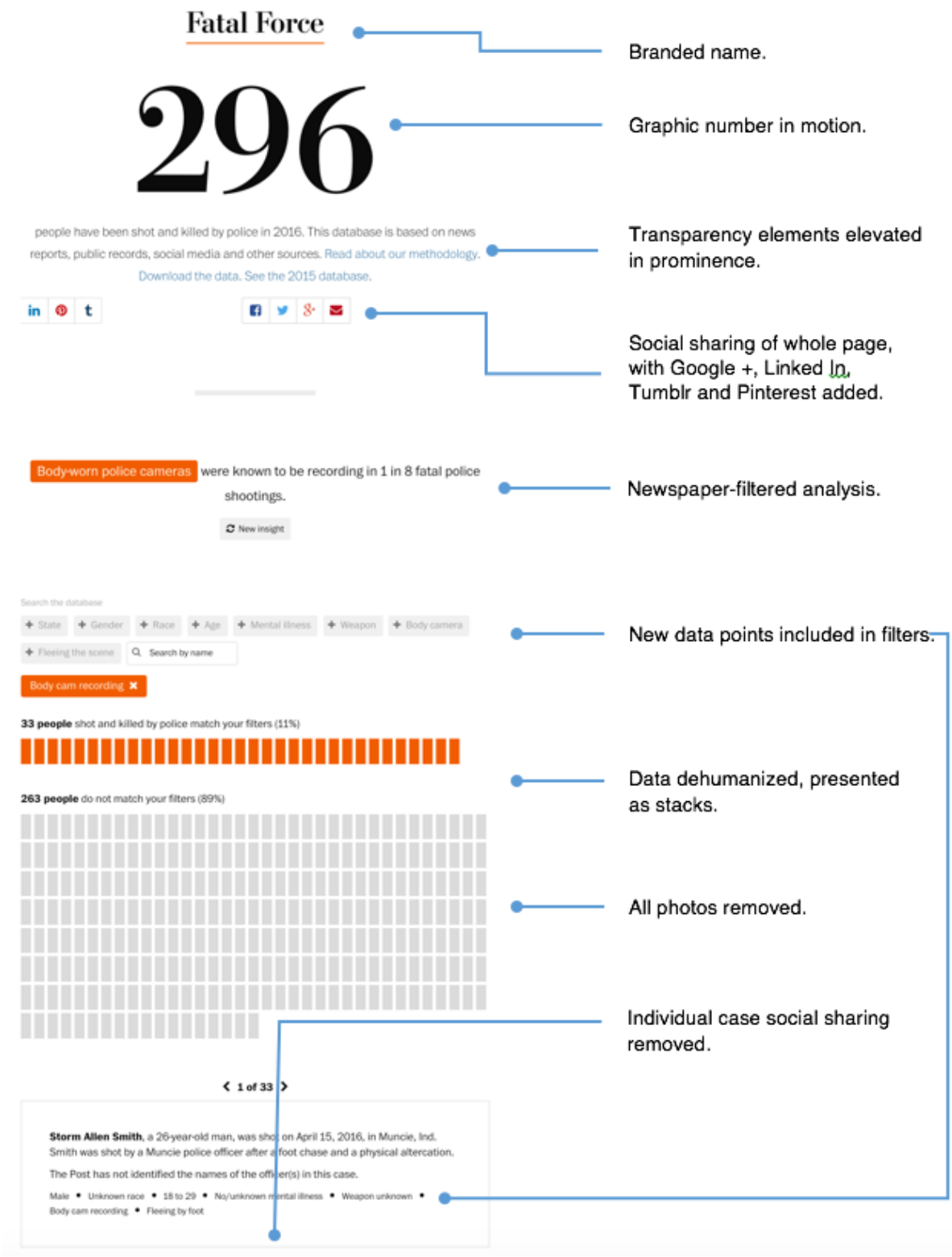


Figure 5.15. Conscious 2016 changes. Revisions to *The Washington Post's* primary data visualization page revealed several intentional directional shifts.

Additional Findings

Additional findings not accounted for in this study's adaptation of Coddington (2015) stood out as meriting special notice because of how they contributed to answering the research question(s). In the course of conducting content analysis it became increasingly clear that not all engagement was the same. Two concepts for future researchers to consider when studying the idea of reader/user engagement:

Public vs. private participation. What is seen when looking for indicators of engagement, engagement opportunities, or participation structures, in both cases there were examples of two very distinct engagements with data and related journalism. Researchers should ask if the engagement they are noting is occurring in private – meaning just an interaction between the user and the data site, without leaving behind visible data trails (such as when uses one of the filters), or is it taking place in public – such as liking something , leaving behind a digital trail.

Personal vs. institutional gatekeeping. On *The Washington Post* Investigations team's Facebook page, they transparently attached authority and responsibility to the person of editor David Fallis. But the institution had no presence beyond a small branded logo in the header. *The Guardian*, contrarily, featured their writers and editors on very personal levels (Figures 5.11 and 5.13), but the most responsive were not the personal accounts of *Guardian* writers, but rather an official entity, branded and responsive, speaking on behalf of not just one person but the entire *Counted* team. It seemed to allow them to be responsive, and give the perception as much which gives them credibility.

2016 Changes

Examining changes that occurred in 2016, these questions were important because they began to address the research question about why. These 2016 changes began to answer the last part of RQ_2, providing a new set of data points allowing researchers to hypothesize about why a publication does certain things to engage or constrain reader participation. Lewis and Usher (2013) explicated the origins of iteration and tinkering in online media, and explained thusly why researchers should expect data journalism to invariably change over time – not just in concept but individual data journalism projects and products. Noting these changes as they happened was important to this study, which initially intended to limit its bounds to content produced in 2015, but altered that plan when efforts to maintain context witnessed notable changes several months into 2016.

Substantively, *The Guardian* decreased the number of articles that were tagged as part of *The Counted*, and the few that were added to the list did not include any open comment days. The data set still existed, and was being made more central to its focus, with commentary moved toward social spaces.[16] On-site, notable changes consisted mostly of cosmetic tinkerings and visual alterations. As the year turned, the counter reset to 0 (briefly). Other story pieces were moved around, some sizes changed, and the main heading (Figure 5.6, top) added motion to the odometer-like wheels. When loading *The Counted*'s primary narrative articles page, the wheels spun like a slot machine, while faces of the deceased populated the background.

For the *Post*, there was not much tinkering apparent in the early stages of 2016. This happened despite obvious errors that became apparent upon reader engagement with the data set. One algorithmic mistake occurred when attempting to share news about a particular death. that

[16] This represents a shift from Domingo et al. (2008), who saw social spaces as separate from debate and commentary spaces. At the time, social was thought of more as a distribution platform only (like Digg).

seemed impossible to ignore occurred with revisions in April 2016 to the primary data presentation page showed a dramatic overhaul that went beyond just visual. Their presentation of 2016 data showed an elevated commitment to highlighting transparency elements (Figure 5.15). One key element to transparency according to Lesage and Hackett (2015), however, was providing additional documentation. Here the *Post* was lacking despite its top-level offer to "download the data," which as this study found was made prohibitively difficult by use of Github (Figure 5.4), which is perceived as simple to a certain technologist crowd (Lewis & Usher, 2013). The question to ask became do they make it easier to access the data in usable form, or do they make it an obstacle to participation (Domingo et al., 2008; Hurrell & Walton, 2013). And there is the clearest difference, despite their looking nearly identical prior to first mouseover.

Conceptually, as people killed by police related to research, the *Post's* changes added new data points – including whether or not the person shot was fleeing from police, and whether or not video of the fatal incident (and the lead-up to it and the aftermath) was captured on police body cam. The officer body camera frame was enhanced in story options in the 2016 as one of six analysis stories in a rotating headline frame. They removed some of the readers' ability to sort, filter, and arrange the data for their own analysis – instead offering a rotating choice of six *Post* settings for data filtration – which shows the frames they were consciously putting forward. The *Post* clearly elevated the importance of transparency to the data set, they removed text, links, suggesting the effort to highlight the data set as their own proprietary creation. However, lest *Post* readers be shut out, comments were added to the primary data page. The revisions were telling because they represented iteration based on learned information. And the graphic design looked much less like *The Counted.* Some notable changes on the primary visualization page,

expressible according to the thematic elements indicative of philosophical outlook on data journalism that ultimately affected the data frame (Figure 5.15):

- Rebranded feature, now labeled "Fatal Force"
- Increased prominence of methodology link
- Increased prominence of data download link
- Data points affecting frame added: suspect fleeing, caught on police body cam
- No data presets
- Continued downplay of racial frame
- Language frame: "Police accountability" preferred over "police violence" when referring to legislation.
- Individual narratives made robojournalism-friendly
- Data set dehumanized – pictures removed along with humanoid graphic figures
- External sourcing (links to local papers and TV stations) removed
- Fixed algorithmic error by removing individual-level social share

All of these and more were visible (Figure 5.16 and Figure 5.17). Not to overvalue a single page of analysis, but these changes were more telling than others. They provided an additional set of data points, construed with the benefit of hindsight and time to consider different decisions regarding how the *Post* conducted its data journalism. *The Washington Post* was championing the statistics they were generating, and crafting their narrative. (*Washington Post*, 2015)[17].

[17] It should be noted in this Medium feature, an oral history they call it, they did give credit – with links – to KilledbyPolice.net and FatalEncounters.org. This was presumably appreciated though links from the *Post*'s Medium account are not nearly as valuable as links from the actual site.

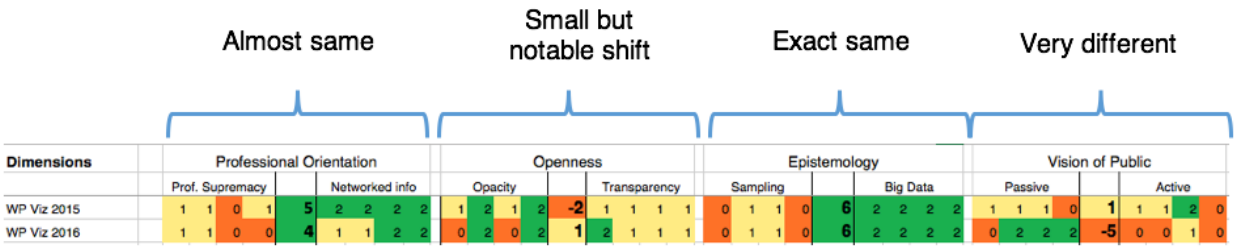


Figure 5.16. Visible shifts in *Fatal Force* redesign. *The Washington Post*'s data visualization page after April 2016 showed slight changes in professional orientation, a small but determinative move toward transparency, no change in epistemology, and definitive change in areas related to vision of the public.

In their own data analysis, text revealed a shift in language in different places that could be telling. These included changing the name from *Police Shootings* to *Fatal Force*, and a new phrase emerged in their stories about "police accountability" being the issue as opposed to "police violence." One notable occurrence that stayed the same, however, were open comments on the redesigned primary page – a comment section that seemed to be undergoing minimal moderation, as of the few comments posted, prominently positioned was an off-topic 9/11 conspiracy theory, with a live link to a YouTube video dismissive of any hijacking, exhibiting the beginnings of digital decay not unlike their Investigations Facebook page.

Summary of Findings

The findings in this study were extensive. The comparative case study methodology using content analysis proved to be an effective means of gaining understanding, and at an important time as data journalism research advances, but not on pace with data journalism practices, uses, and techniques. Most notably, this research added empirical support for Coddington's (2015) typology. In doing so, the research identified two different approaches to engagement with the audience, which were apparent in social distribution settings and in

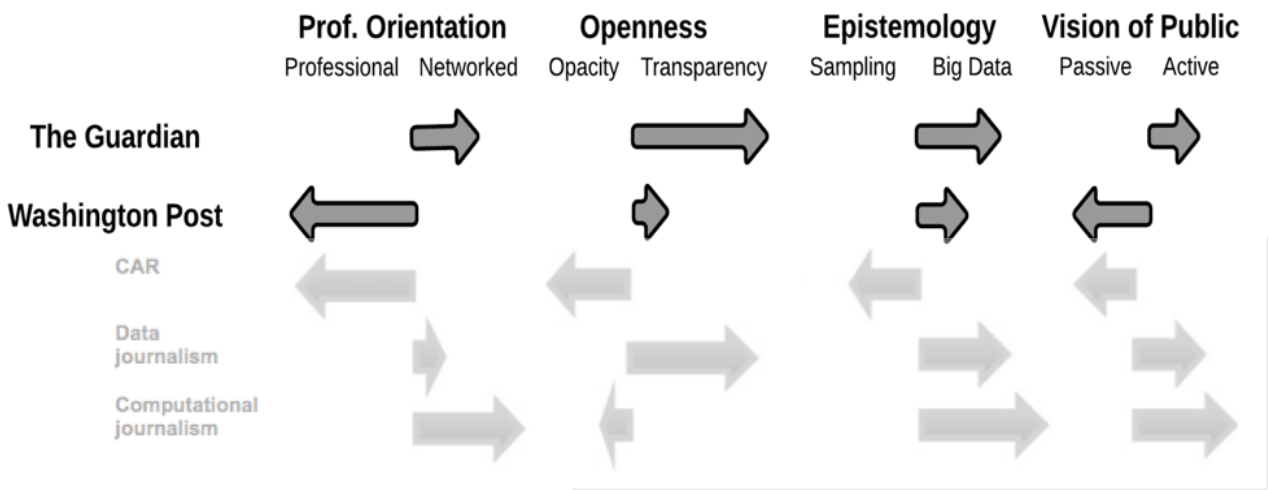


Figure 5.17. Synthesizing results with Coddington. This study provided support for Coddington's typology as a sensible storyline explaining two different approaches to data journalism.

commentary spaces. There also were differences in how they used data – revealing varying ways of practicing journalism even when going after same data and providing readers similar tools.

This study also noticed that much of scholarly discussion about engagement was not considering an epistemic reality that reader engagements occurred not just in different spaces on the Internet, but also engagement took place in different types of spaces, either public or private. Using Coddington's (2015) typology, I found support for his dimensions as determinative indicators of professional outlook across four dimensions, and found an additional dichotomous dimension to be discussed in the next chapter. These outlooks extended beyond the content themselves and into branded media spaces, where *The Counted* showed how they made social part of it, and as a result maintained active engagements in social spaces, which allowed them to facilitate the spreading of stories, *The Washington Post*, however, showed more focus on their end product in the *Post*, but not in other spaces. The only process they highlighted was a somewhat opaque representation of their reporters doing the hard and important work. They treated data as a source and without letting go of well-seasoned connections with higher authorities.

Results showed that social investment yielded social return. Results also revealed that the *Post* used structures to enable one-way engagement between the publication and its readers, and constrain the engagements with data (despite elevated appearance of transparency). *The Guardian* used structures to support two-way engagement, but that meant moving commentary and debate off their own news spaces and into their social spaces. The *Post* let readers have less constrained access to commentary spaces, which were left at the bottom of on-site pages but were not carried into social realms. As will be discussed in the next chapter, these findings answered the research questions in ways that provided guidance and insight for scholars, journalism educators, and policy makers who can better understand the character of data journalism, both as a product and a process.

CHAPTER SIX: DISCUSSION

This thesis began as exploratory research looking to understand what mechanisms in data journalism practices potentially mediated the relationships between journalists and readers when infused with data. I also wanted to know how (and why) *The Guardian* and Washington *Post* used these mechanisms to constrain or enable reader participation. My content analysis comparing the two along these engagement lines showed how both used similar tools to connect with readers on a private level – making for a personal experience when engaging with the data – but provided very different experiences when engaging with readers in public spaces.

Following news discussions helped provide context for these two cases about reader engagements that occurred. About two weeks after the *The Washington Post* launched their newly redesigned Webpages in April 2016, word leaked out from New York that the *Post* had won the Pulitzer Prize. That their efforts in data journalism would be considered was hardly a surprise. Both the *Post*'s *Investigation: Police Shootings* and *The Guardian*'s *The Counted* had been named for many prizes, several of them listed in Table 4.1. Among them, the *Post* won a Polk Award for investigative journalism, and *The Counted* won a Data Award for "Best use of Data Collection the Government Isn't Doing but Should Be" (*The Washington Post*, 2015; *The Guardian*, 2016). Often the two cases were paired together. Both were among the six finalists[18] for the Goldsmith Award, given by Harvard University's Shorenstein Center on Media, Politics, and Public Policy. But that prize went to the Associated Press, for a series of stories that would prove near impossible to beat, as their reporters working to track down unjust labor practices in

[18] Future research may want to explore the use of data among finalists for the Goldsmith Prize for Investigative Reporting. All six of the 2016 finalists (for work done in 2015) used data in some capacity, and one finalist received its accolades for a story about the abuse of data to mislead the public.

the Thailand fishery industry uncovered more than 2,000 people living as slaves, some for decades, on a remote island, peeling shrimp that found its way into supply chains of American retailers such as Red Lobster, Chicken of the Sea, PetSmart, Wal-Mart, Kroger, Safeway, and Olive Garden among others (Pulitzer Prize Board, 2016a).[19]

But the *Post* won the Pulitzer (for National Reporting). *The Guardian* did not and were not even mentioned as a finalist. The *Post* won for "its revelatory initiative in creating and using a national database to illustrate how often and why the police shoot to kill and who the victims are most likely to be" (Pulitzer Prize Board, 2016a). A look at the *Post*'s nomination letter claimed they were first. It would not have been without precedent for the Pulitzer Prize Board to name two winners for covering the same topic. As recently as 2014, co-winners were named in the public service category – *The Guardian* and *Washington Post* – for news coverage related to Edward Snowden's leaked NSA documents. But amid cries of shenanigans and controversy, it became clear that less than a snub, *The Guardian* had entered *The Counted* in different categories – not as national news, but as public service and explanatory journalism.[20] The winner for public service was the AP's story about the finding and freeing of slaves. (This story could be indicative of journalism's future – relying on old-fashioned "shoe leather" reporting with high-tech, data-based investigative tools to unearth injustice.)

Further controversy would emerge over a reported leak about the certainty of a win for the *Post*. Being that a fellow staffer was on the national reporting category's judging panel showed how transparency in and of itself does not tell the whole story, even if it does absolve

[19] Data's role in that story was in satellite surveillance and geo-tracking used to trace boats visiting the slave island and the ocean-to-table route of seafood.

[20] The Pulitzer Prize organization allows nominees to submit their entries into up to two categories, and historically have shown incidence of re-categorizing a submission when appropriate.

some from perceptions of impropriety. It took several clicks and some scrolling to find this conflict of interest. This suggests that transparency may be expected, but it is no guarantee of certain expected independence. Interactions on Twitter and in other published accounts spoke to the potential for "prize bait" in driving editorial decisions, and also helped better answer the "why?" part of RQ_2. It also addressed engagement differences on a personal vs. institutional level. Meanwhile, this journalistic prize fight also showed the limits of transparency, as none of the media gossip sites engaged in the debate made notice of a potential source of the controversial leak.

THE JURY

* Past Pulitzer Prize winner

Josh Meyer (Chair)
Lecturer and director of education and outreach of the National Security Journalism Initiative, Northwestern University Medill School of Journalism, Media, Integrated Marketing Communications, Washington, D.C.

Rita Ciolli
Editorial page editor, *Newsday*, Long Island, NY

Karen Peterson
Editor, *The News Tribune*, Tacoma, WA

Jason Szep*
U.S. national affairs editor, Thomson Reuters, Washington, D.C.

Karen Tumulty
National political correspondent, *The Washington Post*

Figure 6.1. Judges list for *The Washington Post*'s Pulitzer category. An apparent conflict of interest on the National Reporting judging panel stayed a non-issue amid buried transparency akin to fine print. (Pulitzer Prize Board, 2016b)

My interpretation of these findings was that *The Guardian* practiced data journalism one way, based on technology and open-source principles, while *The Washington Post* practiced it

another way – based on defensive attempts to maintain their autonomy and show the strength of their reporting standards and techniques. The *Post*, however, showed little clue or interest in using social forums to develop community relations, and it was possible their holding on to traditional means held back some advancements. But maybe that is not relevant, as maybe both types of journalistic expression were needed to push policy in the directions it began moving around the time these two data journalism cases emerged.

The implications for practitioners of data journalism had to be considered with social implications, as this could be what newspaper battles look like in the future. This opens an array of ethical questions that have not been fully dealt considered, and will have different practitioners falling on different sides. If there was one line that neither publication seemed willing to cross was an involvement with activism, and particularly the Black Lives Matters movement. Accusing a journalist of being involved with a cause has been considered an unforgiveable sin in the past, but technologists are more inclined to develop their access through involvement (Meyer, 2004). Such was the case for *The Guardian* when in the early 2000s when they were actively on the frontlines petitioning for increased data access in the UK. Meyer called the phase that was upon journalism was "the end of pseudo-objectivity" (p. 54) as media consumers no longer required objectivity as much as they did transparency. (But Meyer was sure to point out that in this phase this only enhanced the need for true objectivity in the handling of data. "True objectivity is based on method, not result," [p. 54] he wrote.)

The implications of the research should matter to journalists because they could be indicative of newspaper competition in the future. This potentially influences resources committed to data, who has access, and what can be protected related to data. This was particularly true with video, which has such powerful reach (Hess, 2013). At present, the *Post* is

building up a library of shooting fatalities captured on bodycam, filing a barrage of FOIA requests to successfully secures some of them. Data rights and access – who keeps it and who collects it? Under FOIA, who gets what access how? The *Post* is currently using FOIA to secure video and assemble a library. *The Guardian* is using video technologies to tell compelling stories with digital narratives – making their focus on video about how these data, video data, interact with the reader on a personal level. Both have a role. Both are data journalism.

Future Research: A New Dimension

Coddington (2015) said his framework was designed to be adapted. "This typology is only an initial attempt to classify more systematically these data-driven journalistic practices. These dimensions are hardly the only ones differentiating them, and this area of journalism remains unsettled, so new dimensions and forms of practice may emerge over the next several years" (pp. 343 – 344). This interesting contextual clarity could not be ignored. Indeed, the two cases showed differing visions of the public. But what they also potentially showed, made clear by their submissions for the Pulitzer Prize, were two differing visions of themselves as journalistic entities. If Vision of Self is to be considered a new dimension adaptable to Coddington's typology, the new binary to be tested, it would seem, is public service vs. hard news. Future research would do well to operationalize this binary, and begin comparing multiple data journalism examples within the context of public service vs. hard (or national) news.

Figure 6.2. Videos and activism with police body cams and UGC. User-generated content (top left) tells one narrative, officer-borne video (lower left) tells another, and all are connected to a social network capable of spreading certain messages in real life (right).

A difference apparent in much of the content analysis was the connection to activism. The connection to activism showed both of these cases with journalists attempting to distance themselves from seeming unobjective in their coverage of activist-related activities – and are worthy of further study. The Black Lives Matter movement was closely connected to both projects in 2015, with the *Post* stating it specifically in 2016 (Table 4.1). But since then they have shifted away from that frame. *The Guardian* was more unapologetic about their continued focus on the race frame, as they assess it as one of the realities their research unearthed (Reddit, 2015). Both publications maintained a certain distance – never specifically advocating for civil action, but always there to cover it, with *The Guardian* making that key to their story in pictures and images and social media. Hence the 2016 changes. *The Washington Post* started as an

investigation attempting to personify a racial injustice has since removed emotional elements and changed language related to its promotion of its data.

There were of course limitations to this research. Most generally, case study methodology did not make for automatically generalizable insights (Yin, 2003). The study drew from research in North America, Europe, and a little bit in Australia, non-uniformity of norms and laws and regulations make the insights drawn from the research in this thesis not geographically generalizable. Similarly, there were further nuanced differences between outlooks on data journalism stemming from the US and the UK, which merits further examination. But perhaps most notably, challenges to the Web have been well-noted in literature (McMillan, 2000; Schneider & Foot, 2004; Messner & Garrison, 2007; Sjøvaag & Stavelin, 2012) and were experienced here. As predicted by Lewis and Usher (2013), Web media productions exhibit iterative and tinkering natures – resulting in regularly irregular changes, posing challenges for both intra-coder validity and inter-coder reliability.

Case studies may not be generalizable beyond the specific context in which they are observed (Yin, 1994), but it was not a stretch to conceive of how what happens with future data related to civilian fatalities by police could be similar for other data sets of public interest. Journalistically, the competition that existed between *The Guardian* and *Post* likely represents what competition could look like in the future over different data sets built on similar raw data regardless of who collects it. If such competition exists, either now or in the future, ethical standards may need to be revisited as it was becoming apparent that old standards of ethics through accuracy (Gray et al., 2013) may no longer suffice.

In the end, this research showed support for Coddington's typology serving as a framework. It was quite effective and can easily be adapted to operationalize new elements and

discover new binary decisions that could determine fundamental outlooks on and approaches to data journalism, and indeed, journalism more generally. Lastly, making this research a success was being able to conceptualize an added dimension for the Coddington framework in Vision of Self, which could be seen as a binary choice between public service and hard news.

Future research should want to look at public policy related to data. Who gets to keep it, who has what rights to it. Current law says you cannot copyright facts, but can and should someone be able to protect a certain collection of data points? Lewis and Westlund (2015b) say such proprietary concepts must be considered in the context of business, and to be sure, future producers and consumers of data journalism may have consider for-profit vs. non-profit models. Future research may also want to look at social movements, and how data exists with them.

Conclusion

The FBI announced in December 2015 its intent overhaul its data collection system for tracking people killed by police. Changing data collection methods and how those data will be presented, how accessible they are, and to whom has not yet been made clear. But the agency claims its new public data set will be ready and available starting in 2017 (Kindy, 2015). Likewise, it also was not made clear yet on what standards the FBI will employ, and what elements they will draw from in their data collection methods. And then how the data is disseminated remains to be seen. Lawyers, journalists, policy makers, conspiracy theorists alike will all want to see how it takes shape, with the *Post* and *Guardian* both sure to be watching its emergence doggedly.

This study has presented new empirical evidence about how data journalism works, particularly with an idea of how data, and different applications of data, can begin to have an influence over the relationship between reader and journalist. It is why this research was a success. With so many people looking at this data set and how it emerges in coming years, it could become a standard for how government, the media, and the public. It is in the public interest with much to be learned from either publication (and not just *The Washington Post* because it won the Pulitzer) that both the media and the government understand that not all data are alike, but some certainly are more impactful, not all data journalism is alike. While there is much to learn from both these cases, studying just the *Washington Post*, perhaps because of its Pulitzer win, would be a disservice to data journalism. And scholars will be interested to know its determined by the vision of their audience, and their professional orientation. And media historians may find interest in the evolution of precision journalism, for example.

Different media outlets handled data in different ways, particularly as they presented it to the reader. It made a rather competitive effort to tell stories with a data set about people killed by police. *The Washington Post* won the Pulitzer Prize, celebrating the strength of legacy media, and establishing its institutional power in data journalism. *The Guardian* meanwhile earned its respect and established relationships with its growing community, while both found themselves in a position to assess their models for financial viability, and to consider to what extent they will and should serve as future stewards of data in the public interest.

APPENDIX A: URLs Analyzed for Case 1*, The Counted*

Here is a complete list of URLs demarking pages analyzed for Case 1, *The Counted*. Items noted with an asterisk (*) were the ones included for operationalization of thematic elements according to Coddington's (2015) typology. URLs were current as of April 20, 2016.

* http://www.theguardian.com/us-news/ng-interactive/2015/jun/01/the-counted-police-killings-us-database
* http://www.theguardian.com/us-news/ng-interactive/2015/jun/01/the-counted-police-killings-us-database#[deceased-name]
* http://www.theguardian.com/us-news/ng-interactive/2015/jun/01/us-police-killings-tips
* http://www.theguardian.com/us-news/ng-interactive/2015/jun/01/about-the-counted
* http://www.theguardian.com/us-news/series/counted-us-police-killings
* http://www.theguardian.com/us-news/2015/dec/31/the-counted-police-killings-2015-young-black-men
* http://www.theguardian.com/us-news/ng-interactive/2015/dec/11/kern-county-california-victims-police-killings-justice-video
* http://www.theguardian.com/us-news/2015/dec/31/the-counted-killings-by-police-editorial
* http://www.theguardian.com/us-news/2015/jun/06/congress-police-reform-mandatory-killing-database
* https://www.facebook.com/TheCounted
* https://www.facebook.com/TheCounted/videos
* https://www.twitter.com/TheCounted

http://www.theguardian.com/us-news/2015/sep/01/moving-targets-police-shootings-vehicles-the-counted
http://www.theguardian.com/us-news/2015/dec/01/the-county-kern-county-deadliest-police-killings
http://www.theguardian.com/us-news/2015/dec/10/kern-county-california-police-killings-misconduct-district-attorney
http://www.theguardian.com/us-news/2015/aug/15/filming-police-violence-walter-scott-michael-brown-shooting
http://www.theguardian.com/us-news/video/2016/apr/11/the-counted-police-killings-guardian-us-video
http://www.theguardian.com/us-news/2015/jun/03/counted-police-killing-victims-unnamed-texas-california
http://www.theguardian.com/profile/laughland-oliver
https://twitter.com/oliverlaughland
http://www.theguardian.com/profile/jon-swaine
http://www.theguardian.com/us-news/2015/dec/13/justice-department-database-police-killings-counted-statistics

APPENDIX B: URLs Analyzed for Case 2, Police Shootings

Here is a complete list of URLs demarking pages analyzed for Case 2, *Police Shootings*. Items noted with an asterisk (*) were the ones included for operationalization of thematic elements according to Coddington's (2015) typology. The webpage with two asterisks (**) was analyzed separately (p. 87). URLs were current as of April 20, 2016.

* https://www.washingtonpost.com/pb/policeshootings/
* https://www.washingtonpost.com/national/how-the-washington-post-is-examining-police-shootings-in-the-us/2015/06/29/f42c10b2-151b-11e5-9518-f9e0a8959f32_story.html
* https://washingtonpost.wufoo.com/forms/help-report-fatal-shootings-by-police-in-the-us/
* https://www.washingtonpost.com/graphics/national/police-shootings/
* http://www.washingtonpost.com/sf/national/2015/10/08/fatal-police-shootings-captured-by-body-cameras-2015/
* http://www.washingtonpost.com/sf/national/2015/10/08/police-withhold-videos-despite-vows-of-transparency/
* http://www.washingtonpost.com/sf/national/2015/08/08/black-and-unarmed/
* http://www.washingtonpost.com/sf/investigative/2015/06/30/distraught-people-deadly-results/
* http://www.washingtonpost.com/sf/investigative/2015/04/11/thousands-dead-few-prosecuted/
* http://www.washingtonpost.com/sf/investigative/2015/12/10/new-style-of-police-training-aims-to-produce-guardians-not-warriors/
* https://m.facebook.com/postinvestigations/
* https://twitter.com/wpinvestigates

** http://www.washingtonpost.com/graphics/national/police-shootings-2016/

https://medium.com/the-washington-post/inside-the-washington-post-s-police-shootings-database-an-oral-history-413121889529
http://www.washingtonpost.com/pr/wp/2015/11/13/the-washington-post-shatters-previous-traffic-record-soaring-to-66-9-million-users-in-october/
http://www.pulitzer.org/prize-winners-by-year/2016
http://www.pulitzer.org/winners/washington-post-staff

REFERENCES

Anderson, C. (2013). *Rebuilding the news: Metropolitan journalism in the digital age.* Philadelphia: Temple University Press.

Appelgren, E., & Ngyren, G. (2014). Data journalism in Sweden: Introducing new methods and genres of journalism into "old" organizations. *Digital Journalism, 2*(3), 294-305.

Ausserhofer, J., Gutounig, R., & Oppermann, M. (2015, October). News Production Workflows in Data-driven, Algorithmic Journalism: A Systematic Literature Review. Presented at the Dubrovnik Media Days, Dubrovnik. Retrieved from https://speakerdeck.com/jpr/news-production-workflows-in-data-driven-algorithmic-journalism-a-systematic-literature-review

Benford, R., & Snow, D. (2000). Framing processes and social movements: An overview and assessment. *Annual Review of Sociology, 26*, 611–639.

Berret, C., & Phillips, C. (2016). A crucial skill that most J-schools aren't teaching. Columbia Journalism Review. April 1, 2016. Retrieved from http://www.cjr.org/analysis/data.php

Bivens, R. K. (2008). The Internet, mobile phones and blogging. *Journalism Practice, 2*(1), 113-129.

Brunyee, J. (2015, July 13). How Jeff Bezos is transforming *The Washington Post* for the digital age. *MagWorld*. Retrieved from http://www.fipp.com/news/features/how-jeff-bezos-is-transforming-the-washington

Byers, D., & Gold, H. (2014). Why the Post passed on Ezra Klein. *Politico*, January 21, 2014. Retrieved from http://www.politico.com/story/2014/01/ezra-klein-leaves-washington-post-102424

Calderone, M. (2016, April 22). Why *The Washington Post* won a Pulitzer for a story *The Guardian* also owned. *Huffington Post*. Retrieved from http://www.huffingtonpost.com /entry/washington-post-guardian-pulitzer_us_571a3eb4e4b0d4d3f7230cb0

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.

Coleman, G. (2004). The political agnosticism of free and open source software and the inadvertent politics of contrast. *Anthropological Quarterly, 77*(3), 507-519.

Davis, A., & Lowery, W. (2015, October 7). FBI director calls lack of data on police shootings 'ridiculous,' 'embarrassing'. Retrieved from http://washingtonpost.com /national/fbi-director-calls-lack-of-data-on-police-shootings-ridiculous-embarrassing/2015/10/07/c0ebaf7a-6d16-11e5-b31c-d80d62b53e28_story.html

Berkhead, S. (2015). On data's role in investigative reporting: "It's all about the evidence". IJnet, 7 September.

De Maeyer, J., Libert, M., Domingo, D., Heinderyickx, F., & 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.

Department of Justice. (2016). Attorney General Lynch: Use-of-force data is vital for transparency and accountability. Press Release, Oct. 5, 2015. Retrieved from https://www.justice.gov/opa/pr/attorney-general-lynchuse-force-data-vital-transparency-and-accountability

Domingo, D., Quandt, T., Heinonen, A., Paulussen, S., Singer, J., & Vujnovic, M. (2008). Participatory journalism practices in the media and beyond: An international comparative study of initiatives in online newspapers. *Journalism Practice, 2*(3), 326-342.

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

Flynn, K. (2003, January 4). Recent shootings by officers are anomalies, experts say. *The New York Times.* Retrieved from http://www.nytimes.com/2003/01/04/nyregion/recent-shootings-by-officers-are-anomalies-experts-say.html

Franklin, B. (2014). The future of journalism: In an age of digital media and economic uncertainty. *Digital Journalism, 2*(3), 254-272.

Gearing, A. (2014). Investigative journalism in a socially networked world. *Pacific Journalism Review, 20*(1), 61-75.

Gourarie, C. (2015, October 12). Over 700 people killed by police this year. But who's counting? *Columbia Journalism Review.* Retrieved from http://www.cjr.org/author/chava-gourarie/

Gray, J., Chambers, L., & Bounegru, L. (2012, July 12). *The data journalism handbook*. O'Reilly Media, Inc.

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

Hedley, D. (2013). Social moments in solo videojournalism: Conveying subjective social reality through editorial vision and video. Digital Journalism, 1(1), 64-81.

Hermida, A. (2010). Twittering the news: The emergence of ambient journalism. *Journalism Practice, 4*(3), 297-308.

Hermida, A. (2012). Tweets and truth: Journalism as a discipline of collaborative verification. *Journalism Practice,* 6(5-6), 659-668.

Herring, S. (2010). Web content analysis: Expanding the paradigm. In J. Hunsinger, L. Kastrup, & M. Allen (Eds.), *International Handbook of Internet Research* (pp. 233-249). The Netherlands: Springer.

Hess, K. (2013). Breaking boundaries: Recasting the "local" newspaper as "geo-social" news in a digital landscape. *Digital Journalism, 1*(1), 48-63.

Howard, A. (2014). The art and science of data-driven journalism. Tow Center for Digital Journalism.

Hurrell, B., & Walton, J. (2013). How does that affect me? Making data personally relevant to your audience. In J. Mair & R. Keeble (Eds.), *Data Journalism: Mapping the Future* (pp. 19-25). Suffolk, UK: Abramis.

Johnson, J.T. (1992). The unconscious fraud of journalism education. *Quill*, June 1992, 31-34.

Kim, I., & Kuljis, J. (2010). Applying content analysis to web-based content. Proceedings of the ITI 2010 32nd Int. Conf. on Information Technology Interfaces, June 21-24, 2010, Cavtat, Croatia. 233-238.

Kindy, K. (2015, December 8). FBI to sharply expand system for tracking fatal police shootings. *The Washington Post*. Retrieved from https://www.washingtonpost.com/national/fbi-to-sharply-expand-system-for-tracking-fatal-police-shootings/2015/12/08/a60fbc16-9dd4-11e5-bce4-708fe33e3288_story.html

Kindy, K., & Elliot, K. (2015, December 26). 990 people were shot and killed by police this year: Here's what we learned. *2015 Police Shootings Investigation: The Washington Post.* Retrieved from https://www.washingtonpost.com/graphics/national/police-shootings-year-end

Klein, S. (2015, Mar 17). Antebellum data journalism: Or, how big data busted Abe Lincoln. The Pro Publica Nerd Blog. Retrieved on 1/25/16 from https://www.propublica.org/nerds/item/antebellum-data-journalism-busted-abe-lincoln

Kohut, A. (2009). But what do the polls show? How public opinion surveys came to play a major role in policymaking and politics. Pew Research Center. Retrieved from http://www.pewresearch.org/2009/10/14/but-what-do-the-polls-show/

Kovach, B, & Rosenstiel, T. (2010). Blur: How to know what's true in the age of information overload. New York: Bloomsbury.

Krippendorff, K. (1980). Validity in content analysis. In E. Mochmann (Ed.), *Computerstrategien für die kommunikationsanalyse* (pp. 69-112). Frankfurt, Germany: Campus. Retrieved from http://repository.upenn.edu/asc_papers/291

Krippendorff, K. (2013). Content analysis: An introduction to its methodology. Los Angeles: SAGE.

Krueger, K. (2014). How J-schools are tackling demand for data journalism skills. *Mediashift.org*. Retrieved from http://mediashift.org/2014/07/how-j-schools-are-tackling-the-demand-for-data-journalism-skills/

Las Vegas Review-Journal. (2015). Retrieved from http://www.reviewjournal.com/business/review-journal-expanding-newsroom-create-special-projects-team

Laughland, O., Lartey, J, & McCarthy, C. (2015). The tracking of police violence in the US may have reached a turning point. *The Guardian*, Oct. 8, 2015. Retrieved from http://www.theguardian.com/us-news/2015/oct/08/us-police-violence-tracking-database-loretta-lynch-james-comey

Lesage, F., & Hackett, R. (2013). Between objectivity and openness: The mediality of data for journalism. *Media and Communication, 1*(1), 39-50.

Lewis, S. (2012). The tension between professional control and open participation: Journalism and its boundaries. *Information, Communication & Society.*

Lewis, S., & Usher, N. (2013). Open source and journalism: Toward new frameworks for imagining news innovation. *Media, Culture & Society, 35*(5), 602-619.

Lewis, S., & Westlund, O. (2015a). Actors, actants, audiences, and activities in cross-media news work: A matrix and a research agenda. *Digital Journalism, 3*(1), 19-37.

Lewis, S., & Westlund, O. (2015b). Big data and journalism: Epistemology, expertise, economics, and ethics. *Digital Journalism, 3*(3), 447-466.

Lewis, S., Zamith, R., & Hermida, A. (2013). Content analysis in an era of big data: A hybrid approach to computational and manual methods. *Journal of Broadcasting & Electronic Media, 57*(1), 34-52.

Loosen, W., Reimer, J., & Schmidt, F. (2015). When data become news: A content analysis of data journalism pieces. *The Future of Journalism Conference: Risks, Threats, and Opportunities.* Cardiff.

Mayer-Schönberger, V., & Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. New York: Houghton Mifflin.

McCarthy, T. (2015). Police must report shootings to federal government, suggests Obama taskforce. *The Guardian*, March 2, 2015. Retrieved from http://www.theguardian.com/us-news/2015/mar/02/police-shootings-task-force-21st-century-policing-recommendations

McCombs, M., Cole, R., Stevenson, R., & Shaw, D. (1981). Precision journalism: An emerging theory and technique of news reporting. *Gazette, 27*(1), 21-34.

McMillan, S. (2000). The microscope and the moving target: The challenge of applying content analysis to the World Wide Web. *Journalism and Mass Communications Quarterly, 77*(1), 80-98.

Messner, M., & Garrison, B. (2007). Journalism's "dirty data" below researchers' radar. *Newspaper Research Journal, 28*(4), 88-100.

Meyer, P. (1973). Precision journalism: A reporter's introduction to social science methods. Bloomington, IN.: Indiana University Press.

Meyer, P. (1991). The new precision journalism. Bloomington: Indiana University Press.

Meyer, P. (2004). The next journalism's objective reporting. *Nieman Reports, 58*(4), 54.

Meyer, P. (2009). *The vanishing newspaper: Saving journalism in the information age.* Columbia, MO: University of Missouri Press.

Meyer, M. (2014). Brick by brick: After years of shrinking ambition at *The Washington Post*, Jeff Bezos has the paper thinking global domination. *Columbia Journalism Review, July 2014*. Retrieved from http://www.cjr.org/cover_story/washington_post_jeff_bezos.php

MIT Open Documentary Lab. (2015). Mapping the intersection of two cultures: Interactive documentary and digital journalism. Retrieved from http://opendoclab.mit.edu/interactivejournalism/Mapping_the_Intersection_of_Two_Cultures_Interactive_Documentary_and_Digital_Journalism.pdf

New-York Tribune. (1848, December 22). The mileage of Congress. Retrieved from https://projects.propublica.org/graphics/greeley

Onuoha, M., Pinder, J., & Schaffer, J. (2015). Guide to crowdsourcing. Tow Center for Digital Journalism.

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

Parisie, S., & Dagiral, E. (2012). Data-driven journalism and the public good: "Computer-assisted-reporters" and "programmer-journalists" in Chicago. *New Media & Society, 15*(6), 853-871.

Pinder, J. (2015). Crowdsourcing, in theory and practice. *A how-to guide to crowdsourcing*. Retrieved from http://towcenter.org/crowdsourcing-in-theory-and-practice/

Pulitzer Prize Board. (2016a). 2016 Pulitzer Prizes: Journalism. Retrieved from http://www.pulitzer.org/prize-winners-by-year/2016

Pulitzer Prize Board. (2016b). The 2016 Pulitzer Prize winner in national reporting. Retrieved from http://www.pulitzer.org/winners/washington-post-staff

Reddit. (2015). The Counted project. 100 Best AMA Highlights, 1. Retrieved from http://amahighlights.com/the-counted-project

Rogers, S. (2011, July 28). Data journalism at the Guardian: What is it and how we do it. *The Guardian* Datablog. Retrieved from http://www.theguardian.com/news/datablog/2011/jul/28/data-journalism

Rogers, S. (2012). Behind the scenes at The Guardian data blog. In J. Gray, L. Bounegru, & L. Chambers (Eds.), *The data journalism handbook* (pp. 34-37). USA: O'Reilly.

Rosegrant, S. (2011). Revealing the roots of a riot. University of Michigan Institute for Social Research. Retrieved from http://home.isr.umich.edu/sampler/revealing-the-roots-of-a-riot/

Rosenbaum, M. (2010, Sept. 1). Why Tony Blair thinks he was an idiot. BBC Open Secrets. Retrieved from http://www.bbc.co.uk/blogs/opensecrets/2010/09 /why_tony_blair_thinks_he_was_a.html

Ršumović, N. (2013, March 25). Interview: Theory and practice of data driven journalism. *Mediacentar Online.*

Schneider. S., & Foot, K. (2004). The Web as an object of study. *New Media & Society, 6*(1), 114-122.

Segel, E., & Herr, J. (2010). Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics, 16*(6), 1139-1148.

Silverman, C. (2015). Verification handbook for investigative reporting: A guide to online search and research techniques for using UGC and open source information in investigations. *European Journalism Centre*.

Sjøvaag, H., & Stavelin, G. (2012). Web media and the quantitative content analysis: Methodological challenges in measuring online news content. *Convergence: The International Journal of Research into New Media Technologies, 18*(2), 215-229.

Smith, D. (2015, December 14). Washington Post bids farewell to office where it broke Watergate. The Guardian. Retrieved on April 13, 2016, from http://theguardian.com/ world/2015/dec/14/washington-post-bids-farewell-to-office-where-it-broke-watergate

Stray, J. (2010). How The Guardian is pioneering data journalism with free tools. Nieman Journalism Lab, August 5. Retrieved from http://www.niemanlab.org/2010/08/how-the-guardian-is-pioneering-data-journalism-with-free-tools/

Swaine, J., & Laughland, O. (2015). FBI resists calls to reform voluntary reporting system for police killings. The Guardian, Sept. 28, 2015. Retrieved from http://www.theguardian.com/us-news/2015/sep/28/fbi-police-killings-voluntary-reporting-system

Tandoc, E. (2014). Journalism is twerking? How web analytics is changing the process of gatekeeping. *New Media & Society, 16*(4), 559-575.

Tufte, E. (2001). *The visual display of quantitative information.* Cheshire, CT: Graphics Press.

Viner, K. (2013, Oct. 9). *The rise of the reader: Journalism in the age of the open Web*. Speech presented as an AN Smith lecture, Melbourne, Australia. Retrieved from http://www.theguardian.com/commentisfree/2013/oct/09/the-rise-of-the-reader-katharine-viner-an-smith-lecture

Washington Post. (2015, December 16). Inside the *Washington Post*'s police shootings database: An oral history. https://medium.com/the-washington-post/inside-the-washington-post-s-police-shootings-database-an-oral-history-413121889529

WashPostPR. (2015, November 13). The Washington Post shatters previous traffic record, soaring to 66.9 million users in October. The Washington Post. Retrieved from https://www.washingtonpost.com/pr/wp/2015/11/13/the-washington-post-shatters-previous-traffic-record-soaring-to-66-9-million-users-in-october/

Westlund, O. (2013). Mobile News: A review and model of journalism in an age of mobile media. *Digital Journalism, 1*(1), 6-26.

Wormer, H. (2007). Figures, statistics and the journalist: An affair between love and fear. AStA *Advances in Statistical Analysis*, *91*(4), 391-397.

Yin, R. (1994). Case study research: Design and methods (2nd edition).

Yin, R. (2003). Case study research: Design and methods (3rd edition).

Yu, R. (2014, March 7). Booming market for data-driven journalism. *USA Today*. Retrieved from http://www.usatoday.com/story/money/business/2014/03/16/data-journalism-on-the-rise/6424671/

CURRICULUM VITAE

Dan Michalski
Email: danray.michalski@gmail.com
Phone: 702-475-3765
Twitter: @danmichalski
Skype: danmichalski

Education
University of Nevada, Las Vegas. (2016). Master of Arts - Journalism & Media Studies.
European Journalism Centre. (2014). Certificate of Data-Driven Journalism.
Northwestern University. (1994). Bachelor of Science in Journalism.
St. Mark's School of Texas. (1990).

Academic and Professional Presentations
GUD LUK: Decoding the Messages of Vanity License Plates for Gamblers
16th International Conference on Gambling and Risk Taking. Las Vegas, 2016.
Jackpot of Ideas: Finding Talent on College Campuses
iGaming North America. Las Vegas, 2016.
Reality Check: Time to Update Your Social Media Strategy?
iGaming North America. Las Vegas, 2015.
License to Ethnograph: What Vanity Plates Say about a City
Far West Popular and American Culture Associations. Las Vegas, 2014
Food Porn: Sexualized Imagery in Portrayals of Edible Mass Consumables
Far West Popular and American Culture Associations. Las Vegas, 2014
OMG, LOL, WTF? Professional Practices for Gaming in Social Media
15th International Conference on Gambling and Risk Taking. Las Vegas, 2013
Cheating: Assessing and Addressing the Danger
Digital Gaming & Lottery Policy Summit. Washington DC, 2011
The New Challenge for Online Poker: Playing the Hand That's Been Dealt
iGaming North America. Las Vegas, 2011

Book Chapters
Michalski, D. (2003). The costs of child sexual abuse litigation threaten the Catholic Church. In L. Gerdes (Ed.), *At issue: Child sexual abuse in the Catholic Church* (pp. 75-83). San Diego, Calif.: Greenhaven Press.