Data Journalism: impact of technologies on data journalism

Rashmi Thapaliya

COMM 5363

Abstract

Data journalism is trending in journalism these days. Data journalism opens up the new

possibilities in the traditional way of doing journalism (Bradshaw, 2012). Technological

advancements have shaped the way journalism is practiced and operated (Powers, 2012).

Looking into the relationship between technology and data journalism through the perspectives

of technological determinism and sociomateriality can yield interesting findings in the data

journalism. This research used interviews with data journalists as a method to answer the

questions related to data journalism and technological advancements. The result of the interviews

showed that technologies do have impact on data journalism but at the same time it is not all

about the technology. The key to any news story is its content without which the news story is

not complete. Future researchers can look at different aspects of technology that influence data

journalism and also at the same time how data journalism shapes the use of technologies in the

newsrooms.

Keywords: Data journalism, technology, communication

Introduction

The traditional analog media was largely one-way in their flow of information. The audiences were passive receivers without having much contribution to the information flow through interactive feedback (Kaul, 2013). The one-way mass communication is not the limitation of journalism today. There have been many technological advancements throughout the history that have shaped the way American journalism has operated (Powers, 2012). First came the print media, then the broadcast media, and now the convergent media. With the changing media landscape, the way journalism operates has also changed. Journalists need to learn new skills that would fit the media they would be using as a channel of communication. One of the most recent developments in journalism is data journalism. When the future of journalism is being looked with doubtful eyes, the ability of journalists to respond to the availability of data can be the key to the sustainability of journalism (Bell, 2012). There are public databases in the Internet. But merely having these public databases is not enough; journalism can utilize the quality downloadable and searchable databases as a tool for gathering and disseminating news (Schudson, 2010). This means the journalists need to learn yet another skill that would enable them to find the story behind the numbers through data crunching. While many are skeptic that the journalists would be able to learn such skills in the near future, it would be interesting to watch the way data journalism operate and would operate in the newsrooms. Information and communication technology (ICT) researchers have found that technological innovations have changed the way people and society operate and even the critical writers have accepted that technological advancements have shifted if not overthrown the way we think and live (Geib, Jackob & Quiring, 2013). The purpose of this research is to find out how the communication technologies have influenced the way data journalism is done.

Literature Review

What is data journalism?

      In the digital world where almost everything can be described with data, data journalism is not only about the journalism that deals with some numbers in a spreadsheet (Bradshaw, 2012). Data Journalism is different from traditional journalism because of "the new possibilities that open up when you combine the traditional 'nose for news' and ability to tell a compelling story, with the sheer scale and range of digital information now available" (Bradshaw, 2012, p. 2). Therefore, data journalism adds a new dimension to the traditional way of finding, gathering and reporting news through data as facts. In the digital age stories get reported simultaneously as they happen so it has become more important to report the developments of a story than just breaking a story (Lorenz, 2012). "Using data transforms something abstract into something everyone can understand and relate to" (Lorenz, 2012, p. 4). Data Journalism often uses infographs and data visualizations to tell complex stories in simple form. According to Data Journalism Handbook (2012) visualization is the workhorse of data journalism. Visualization can be used for both the reporting and publishing phase of journalism (Gray, Bounegru & Chambers, 2012).

      Data Journalism is an emerging term in journalism these days. There is no one specific definition of Data Journalism. Different data journalists have different perceptions about data journalism. Data journalists talk about what data journalism means to them in Data Journalism Handbook (2012).

            Data journalism serves two important purposes for news organizations: finding unique stories (not from news wires) and execute your watchdog function.

Especially in times of financial peril, these important goals for newspapers to

achieve (Jerry Vermanen, p. 7)

Another data journalist from Financial Times writes about data journalism:

I think it's important to stress the "journalism" or reporting aspect of 'data

journalism.' The exercise should not be about just analyzing data or visualizing

data for the sake of it, but to use it as a tool to get closer to the truth of what is

going on in the world. I see the ability to be able to analyze and interpret data as

an essential part of today's journalists' toolkit, rather than a separate discipline.

Ultimately, it is all about good reporting, and telling stories in the most

appropriate way. (Cyntia O Murch, Financial Times, p. 9).

News organizations perceive data journalism as a part of a transition from being rather

passive news-and-information sites to more interactive news-and-information platforms. On

these platforms, readers can not only consume journalism, but interact with data visualisations

and raw data sets. (Sirkkunen, 2011). Data journalism is a more recent term for computer-

assisted reporting. A key element of database journalism is the concept of algorithm in

information. Diakopoulos (2012) explained the idea of algorithm as follows:

Computing is about information, about describing and transforming it, but

 also about acquiring, representing, structuring, storing, accessing, managing, processing,

communicating, and presenting it. And computing is about algorithms: their theory,

feasibility, analysis, structure, expression, and implementation (Diakopoulos, 2012, p. 3).

It is not easy to find one particular definition of data journalism. Based on what the

journalists and scholars above have to say, data journalism can be defined as using new

technologies and tools in journalism to find and report truth. Data journalists use statistics and

datasets that make the reporting process more scientific than having to rely on the quotes of an individual or group to report a story.

History and concerns of Data Journalism

Data Journalism is the term that is popular these days to refer to the journalism that uses data, computer or scientific methods to report news stories. Computer assisted reporting (CAR) was the term that was used to refer to Data Journalism earlier. According to Cox (2000), the computer assisted reporting began with the 1952 presidential election when Dwight Eisenhower and Adlai Stevenson were running for the office. It was the first time when the computer was used to predict the outcome of an election. Philip Meyer is one of the innovators of computer-assisted reporting. In the Detroit riots in 1967 he found through a survey that people who had attended college were as likely to participate in riots as were high school dropouts. It was an important finding because the finding did not match the assumed hypothesis; this finding was based on fact and not on the assumption. Meyer wrote the book *Precision Journalism* in 1973. Multiple editions of the book have been published since then. Meyer (2002) argues that with the ever-growing complexities of the world, a journalist should be able to do more than one task. "A journalist has to be data manager, a data processor, and a data analyst" (Meyer, 2002, p. 1). Meyer (2002) advocated for the scientific approach to journalism. He stated that scientific approach would help journalists compose the message that the audience not only receive but also understand (Meyer, 2002).

Another key event in the history of CAR was the story by Burnham for The New York Times. Burnham found through the analysis of court records and arrest reports that in New York City, a black person was eight times more likely than a white person to be murdered (Cox, 2000, p. 9). By the 1980s microcomputers started to become more popular among people. The

reporters took the initiative by purchasing their own computers, which was followed by organizations. Another development of CAR came from the investigative report of Elliot Jaspin at *The Providence Journal* where he found by comparing the database of school bus driver with database of traffic violations and court records that some of the bus drivers were traffic violators and some were drug dealers (Cox, 2000, p. 10). Another important CAR story in the 1980s was a series of stories produced by Bill Dedman, "The Color of Money" that exposed the racist policies in lending of Atlanta-area financial institutions. This story won Dedman the Puilitzer Prize. By the 1990s computer assisted reporting became much more common. Cox (2009) writes that in 1990, the newsrooms started using SAS, a software package, to analyze large data files. Another important technological advancement in the newsrooms was the Internet. In 1998, all the local properties were evaluated and each homeowner received the value of their house. The *Observer* created a searchable database on the Internet where the homeowners could find the price of other houses in their neighborhoods and compare with their own. It took time for everyone in the newsroom to be familiar with the use of the computer and the Internet for writing and researching. Other newspapers like The Washington Post, The *Detroit News* and others also started adopting computers and the Internet. The question is whether it was technology that drove reporting or the reporting of those pioneers that drove technology (Cox, 2000, pp. 14-17). The computational journalism research has focused on two issues: first understanding the dynamics of contemporary news practices and second designing digital tools that can supplement, routinize, or algorithmically expand the scope of these traditional practices (Anderson, 2011).

One of the earliest empirical research on the use of databases for reporting was conducted by Jacobson (1989). In the telephone survey he conducted, his main purpose was to examine the

extent to which the commercial news organizations were using the electronic database for their

news stories. He conducted a telephone survey of 247 stations in the top 100 markets, and his

goal was to find the answers to three intriguing questions: how commonly the databases were

used for news gathering, how it helped the news-gathering process, and what were, if any,

problems or dangers associated with using data bases for reporting. The results of the study

showed that the majority of the journalists had a positive opinion towards using data bases for

reporting in stories. The result showed that 87.7% of interviewees believed that the usage of

databases helped to improve the news coverage, 65.6% of interviewees believed that the use of

databases improved their stations' competitiveness against the other local news, and 21.5%

believed that the databases were helpful in improving their competitiveness against the national

news. Jacobson (1989) writes that there had been some concerns about the problems associated

with the database reporting. Some journalists were afraid that the databases might not be

accurate which can result to the faulty stories. They were also concerned that the use of

databases may be discouraging the originality in reporting and avoiding the "leg work."

However, in his study, he found out that most of the journalists were very positive about the

coverage of news stories using databases that led to "improved accuracy, increased detail, and

improved perspective" (p. 207). This study conducted by Jacobson tells us a lot about the

perspective of using databases as a reporting tool when it first developed. The majority of

journalists who were advocating for it suggest that the use of databases came as good news to

reporters. While looking at the research today, we can see how the field of database reporting has

grown and is continuing to grow. Now the journalists from different fields, not just the television

stations, are using databases as a source.

Jacobson and Ullman (1989) also conducted another research on using databases for reporting. The research surveyed the reporters and the librarians who helped the reporters with the database search. The study showed that in most of the cases both the librarians and journalists had positive opinions about such database usage while reporting. The survey also showed that the journalists had "little worry over possibilities that use of databases might lead to a weakening of coverage" (p. 24). While some journalists at the time were concerned that the use of databases may lead to loss of originality and credibility of news, the study showed that it was a minimal concern among the journalists and librarians. The opinions concerning the risk of inaccuracy in reporting showed that "reporters and editors generally expressed the notion that any potential dangers of inaccuracy or dependence on databases were guarded against by good reporting habits as traditionally taught" (p. 24). The study was conducted more than two decades ago, when the use of databases in the process of news reporting was in its infancy. Today, the use of databases for in-depth reporting has been considered as the only way to save the investigative reporting and journalism's watchdog tradition when the traditional news organizations are rapidly changing. The newsrooms are taking new shapes these days with "journalists, computer scientists, and developers working together to seek large government data sets in order to detect favoritism, incompetence, and corruption through the visualization of data" (Houston, 2010, p. 51**).** This collaboration of the journalists with the computer scientists and developers is outlining the future of investigative journalism. The aim of such collaboration is to provide accurate, original, reliable, and socially useful information to the public (Flew et al., 2011).

Journalists have to be good at reporting and writing news stories, editing, taking photographs, laying out pages, taking videos and editing videos, and using other technologies.

New skillset that journalists need to have if they want to survive in the shrinking newsrooms is the skill of gathering and analyzing the data. And there are quite a few journalists who are using the data in their stories these days. The question is whether or not the data they are using are clean and trustworthy. Messner and Garrison (2007) show their concern on the use of dirty or unverified data by journalists. In data journalism the term dirty data is used for the data that are unorganized, messy or unverified. Organizing the data means cleaning up the dirty data. Messner and Garrison (2007) quote Houston (1996) who listed some major problems with databases. These include: incompetent data, non-standardized data, data entry errors, wrong codes, and misunderstood codes.

Such data need to be revised and checked for accuracy before use. They use the example of the story on a dam failure near Kilauea, Hawaii in March 2006. The reporters used the outdated data that contained the information of the dam for the stories. This is one of the prominent problems of data base reporting. The journalists who lack the skills of data analysis may use the dirty data in their stories without taking time and effort to clean them up.

Another problem in database reporting is that the readers may not be able to comprehend the stories if the data are not well explained. A survey comparing the stories that used and did not use databases showed that the readers, "liked the Computer-Assisted Reporting (CAR) less than the other two types, and rated those as lower in quality than the other versions. Such stories were also rated harder to read than the anecdotal versions" (Mayo & Glen, 2000, p. 68). The news stories should not be hard to comprehend. People will not take time trying to read and comprehend such stories. Journalists writing such stories should be able to explain the difficult terms and patterns that can be found in numbers. They should first be able to understand the meaning behind the numbers before they can explain it to readers. Tankard and Lasorsa (2000)

have suggested some strategies that students and professionals in Computer-Assisted Reporting

can use while working with data. One of their suggestions is that the reporters should be able to

compare between two or more time periods that would help the readers and audience to look at

the development of any event from the past or the present.

Schafraad, Wester, and Scheepers (2006) provide seven guidelines to help the journalists

for electronic data collection. They stress on the importance of keywords that can be helpful in

collecting data. "A theoretically and empirically informed keyword list is an important

instrument for the construction of a valid, reliable, and reproductive corpus in all content analysis

and research themes" (p.457). It is very important for the reporters to start their search for

databases with the right keywords related to the stories they are doing research on. These

keywords can help lead the reporters towards the right path for data collection. Cohen, Hamilton,

and Turner (2011) suggest that journalists can work with the computer scientists who can help

them with cleaning the dirty data. Working with the computer scientists can help them get

started. They can utilize the experience of working with the computer scientists to understand the

way data can be analyzed and manipulated to dig out the hidden stories behind them.

A good way to teach the skills related to computer assisted reporting and data journalism

is the introduction of such skills in the journalism schools. A study found that the comprehensive

programs provide more data analytical courses than mixed adopter programs and lagging

programs (Yarnall, Johnson, Rhinne, and Ranney, 2008).

> Comprehensive programs offer hands-on, coherent curricular that develop data analytic
>
> skills over time through multiple departmental courses; they tend to be smaller and more
>
> selective. Mixed adopter programs make data analytical learning opportunities available,
>
> but the student must largely seek them out; most often, the student will be learning

outside the department, a situation that raises questions about how well students learn to

apply such skills to journalistic work. Lagging programs offer few data analytic learning

opportunities and are marked by internal disagreement over whether to emphasize such

skills and therefore risk alienating "math-phobic" journalism students. (Yarnell, et. al., p.

159)

The future of journalism depends on the future journalists. The students who are gaining

journalism knowledge and skills should be given an opportunity to learn some data analysis. This

skill can be very helpful to them while doing investigative reporting or reporting other stories. To

save the watchdog role of journalism, the students should be given the knowledge of

computing/data journalism.

The journalists can also collaborate with the public through the system based on 'a cloud

for the crowd' that utilizes the databases and the collaboration of human expertise as a tool for

the most useful investigative journalism (Cohen, Li, Yang, & Yu, 2011). Cohen, Hamilton and

Turner (2011) write about the role of computational journalism in investigative and public affairs

reporting. Cohen et. al. (2011) illustrate the use of scientific methods such as sampling,

correlation, and database technologies by journalists to find out news that were otherwise hidden.

The article also suggests that most of the journalists may not have the computational knowledge

to dig out the information from an unorganized or raw database. It is, therefore, necessary for the

journalists to work with computer scientists to be able to learn to find and analyze data. People

have blamed the technology for the destruction of journalism. But one should also keep in mind

that the seeds of journalism's rebirth are within the technology (Anderson, 2011). Data

journalism may be how journalism will be done in the future.

Journalism and Technology

Technologies have always influenced journalism. This influence of technologies on journalism is like a flux now when technological advancements and change in journalism has become so rapid (Spyridou, et. al, 2013). Spyridou et. al. (2013) found that although technology may not influence the newsroom activities in a predictable way, but the new technological infrastructures do affect the editorial practice and roles (p. 91). Journalists had a positive view about the influence of technologies on the daily journalism function such as speed, efficiency and research. However, journalists did not view technological advancements as changing the deeper norms and practices of journalism such as transparency, accountability and collaborative journalism (p. 71). Technology-focused approach in journalism applies the ethos, ideas and practices of technological communities to the traditional ideas and framework of journalism (Lewis & Usher, 2013).

Lewis and Usher (2013) write that Houston (2010) identified three important aspects of technological innovation and journalism production: database reporting, spreadsheets, and online reporting. Lewis and Usher (2013) emphasize on the importance of open source. The concepts of open data or open source are about the freedom of data from all kinds of restrictions on copyrights, patents or other controls (Sirkkunen, 2011). Sirkkunen (2011) quote Schelong & Stepanets (2011) who present eight principles of open data. The principles are completeness, primacy, timeliness, accessibility, machine readability, absence of discrimination, absence of propriety and absence of license requirements. With tools and open source technology journalists can better access the knowledge of the community. However, open source journalism does not mean documenting and exposing all the information gathered through conversation or interview. But open-source journalism helps readers to understand the process that journalists involve to

gather and report the information they have (Lewis and Usher, 2013, p. 613). Lewis and Usher

(2013) suggest that "open source offers the opportunity for technologists and journalists together

to think about traditional journalism as espousing new values-not diminishing long-standing

virtues but instead providing a news framework that makes journalism more relevant in a

participatory, digital culture" (p. 615).

With the open data and tools available online the process of analyzing, visualizing and

publishing large amounts of data has become easier (Sirkkunen, 2011). Journalism tools such as

Google Refine, Many Eyes from IBM, Google Earth, Time Flow are used to organize data in

newsrooms (Cohen, Hamilton & Turner, 2011). Messner and Garrison (2007) write that DeFleur

(1997) found some most common errors in government data which are: not all values are

recorded for each variable, not all values are complete for all years, variables do not always have

the same meaning, some values of a variable have more than one meaning and the use of

aggregate data, and the result make no sense (p. 91). The use of these dirty data can skew the

conclusion and result in the false conclusion of a story. This brings the validity of the data and

news story in question. Therefore, it is important for the news writers to have some skill of

cleaning up the dirty data before it is actually used.

After gathering and analyzing data, data journalists also have the responsibility of

presenting the data to the audience. Data visualization is a popular way of presenting a data-

driven stories these days. Data visualization helps the audience to get a clear idea of what the

story is trying to say presenting the raw data in an artistic visual form (Gray et. al., 2012).

According to Data Journalism Handbook (2012) data visualization can be used to show change

over time, to compare values, to show connections and to trace flows. Data journalists use

different tools such as Google spreadsheet charts, Google fusion tables, color brewer and others for infographics or data visualizations (Rogers, 2012).

During reporting phase visualizations can be used for identifying themes of the story, identifying outliers or any error the data may have, and finding a hole in a story. While publishing, visualization helps to display a point of a story in a more compelling way, remove technically jargon information from prose, and make the reporting process more transparent through interactive exploration of the data (Gray et. al., 2012, p. 186). Schudson (2010) compares the database in the digital age to the narrative in the modern age of novel and cinema. Schudson (2010) cites Walter Lippmann (1922), a journalist and freelancer, who criticized through his article in Liberty and the News, "that American Journalism was failing to serve the needs of modern democracy-and that it would continue to fail without the help from forces beyond itself" (p. 100). Lippmann then pointed out two reasons: first most of the journalists were the untrained amateurs and second, the conventional tools of journalism were not enough to report the complex world. The article uses an example of articles published in one issue of The New York Times which included three news sources: academic research, an internal government audit, and a nonprofit advocacy group (p. 102). Schudson(2010) states that journalism relies on these three news sources to provide information, and journalists will not be able to replace them because they lack the expertise of the professionals in these fields. The author is trying to implement the belief of Lippmann, who envisioned about a century ago, that the journalism needs the help of other disciplines and academics as well to maintain its credibility. Schudson (2010) further mentions that there is a good possibility of having more newspapers with less staffs in the coming days. He is also positive that the less staffed newspapers would maintain the quality and efficiency of their news materials by using the databases as a tool for reporting. He

writes that, "a database is not journalism, but, increasingly, sophisticated journalism depends on

quality downloadable, searchable database" (p. 107). In other words, journalism is not fully

dependent on databases, but they definitely help in increasing the quality of reporting. The article

makes the important point that journalism cannot survive on itself, and it does need the help of

academics, researchers, databases and skillsets from other disciplines to carry on its duties.

Technological Determinism

Economist and sociologist Thorstein Velben coined the term "Technological

Determinism". The Canadian theorist Marshal McLuhan was the most famous supporter of the

Technological Determinism. Technological determinism is the most popular perspective of the

relationship between technology and society though it has been prone to various criticism by

scholars (Chandler, 2000). According to Technological Determinism, the technological

developments in the media, technical field and others are seen as the fundamental factor for the

development of society. "Technological determinists interpret technology in general and

communications technologies in particular as the basis of society in the past, present and even

the future. They say that technologies such as writing or print or television or the computer

changed society" (Chandler, 2000, p. 1).

Orlikowski (1991) writes that there are three different streams of technology research.

The early work on technology defined technology as an objective or external force that had a

fundamental impact on the organization and its structure. The second type of research on

technology argued that technology is a result of human and social interpretations and

interventions. More recent or the third work on technology is focused on "soft" determinism that

recognizes the external impact of technology but at the same time also takes into consideration

the impacts of human role and organizational context (p. 2). The technological imperative model

falls in the category of the first research type discussed. According to this model technology is seen as having a unidirectional causal influences over humans and organizations (Orlikowski cites Giddens, 1984:207). This research ignores the influence of humans in developing, appropriating, and changing technology creating an incomplete account of technology and its interaction with organization (Orlikowski, 1991, p. 3). The strategic choice model sees technology as a product of current human action, design, and appropriation. The third is the model of technology as a trigger of structural change. Orlikowski (1991) writes that this model was suggested by Barley (1986, 1990) where he "posits a role for technology, not as material cause, but as a material trigger, occasioning certain dynamics that lead to anticipated and unanticipated structuring consequences such as increased decentralization that was found in his study" (p. 6). Orlikowski (1991) proposes the structuration model of technology to reconstruct the concept of technology and to propose a model for investigating the relationship between technology and organizations. "Structuration is posited as a social process that involves the reciprocal interaction of human actors and structural features of organizations. The theory of structuration recognizes that human actions are enabled and constrained by structures, yet these structures are the result of previous actions" (p. 8). Therefore, this model looks at the mutual relationship between humans and technology implementing that humans and technology affect each other. Humans are the creators of technologies and technologies are used by humans to accomplish some actions. Components of the Structurational Model of Technology are: 1. Human agents such as technology designers, users, and decision-makers; 2. Technology and 3. Institutional properties of organizations such as structural arrangements, business strategies, ideology, culture, professional norms and others (p. 16). Orlikowski (1991) proposes structuration model of technology and organization that stresses the socio-historical context and

the dual nature of technology which is both objective reality and socially constructed (p.33). Orlikowski (1991) has used field study as a research method which shows that "there are strong tendencies within institutionalized practices that constrain and facilitate certain developments and deployments of technology" (p.34). Technological determinism keeps technology in the center stage as the moving factor for human interaction and development of society (Leonardi, 2009). Leonardi (2009) writes that in the last two decades there have emerged two constructive research programs that argue about the invalidity of technological determinism. The social constructivism program has made an effort to deal with the notion of "impact" by focusing on the "cause" of the impact and argued the factor of the cause is in fact, the people involved in developing the technology (p. 282). The social construction of technology approach argues that the development of technologies is determined by the social and cultural context of the people who are involved in developing the technology (p. 283). Both of these approaches therefore, take the technology from the center of the stage and replace with something else: developers of the technology themselves or the social and cultural attributes those developers carry with them. Leonardi (2009) writes the mutually constitutive approach "suggests that technology-organizing and development use may themselves be unwarranted dichotomies" (p. 299). The mutually constructive approach drops the "impact" metaphor and replaces it with change which is an ongoing process of gradual adjustment.

A mutually constitutive approach that examines the interdependent relationship between technologies and organizational change both diachronically and synchronically frees us from the temptation to construct artificial starting and stopping points, such as the implementation line, in the change lifecycle (p.301). Leonardi (2009) therefore, suggests that the

technology and organization should both be considered mutually than separate entities with one holding more power than other.

A literature review on data journalism, technological impact on journalism and technological determinism theory gives an overview of the development and growth of data journalism and technological impact on journalism. Using the sociomateriality framework this paper aims to answer the following question:

RQ1: What are the impacts of technologies in the development and practice of

data journalism in the newsroom?

Method

This research used interviews with the data journalists working in the newspapers and new media in the USA and UK as a method of study. Five journalists working in the newspapers such as *The Guardian*, *The Los Angeles Times*, *The Chicago Tribune* and in the broadcast station *National Public Radio (NPR)* were interviewed. A qualitative approach was chosen because qualitative methods are beneficial for discovery, exploration of new area, and for developing hypotheses and it lets the researcher to emphasize on specific case and its context (Miles, Huberman & Saldana, 2014). The interview was chosen as a method because the purpose of interviewing is "an interest in understanding the lived experience of other people and the meaning they make of that experience" (Seidman, 2013, p. 9). Here, the interest is to know the opinion of data journalists about the impact technology has had on their work. Interview would be the best method to know the opinions of the data journalists. Recruitment was carried out by sending emails to the data journalists working in the news media in the USA and UK. The journalists who agreed to participate in the study were interviewed via phone or skype, a free voice over protocol service (VOIP).

Result

After interviewing the data journalists and going through the transcriptions some of the most common themes from the interviews were extracted.

Time, convenience and cost

The data journalists agreed that the technology has made working with data convenient. They agreed that with the advancement in technology and development of software for data analysis, data journalists now can accomplish their work in much shorter amount of time. They do not have to sit in front of computer for hours trying to figure out the codes because the computer programmers have already done that for them. One of the journalists working with NPR said, "Technology makes things faster. We use GIS software, statistics software, databases, we write software, build things on java script, tools that are useful like open refine to clean data, which make the process easier and faster." In terms of convenience one of the journalists working for The Los Angeles Times gave an example of back in the 1980s when the government data were stored in tapes in heavy suitcases. So the journalist who wanted to use those data for analysis had to get the permission to get access to the suitcase and make sense of the data in the tapes in a limited amount of time. But now things have changed. A lot of government data is available online. It is much easier to get access to the data and use it for analysis. One of the data journalists working for The Chicago Tribune said that the cost has really gone down due to the technological advancements in newsrooms, "Lot of what we do today was technically possible 40 years ago. But computers were incredibly expensive and very few people had the skillsets to program computers. Now that the computers are much cheaper and accessible there are more people who have these skillsets." The data journalist emphasized the fact that there are more

people with programming skills in the field of journalism and art now. But 40 years ago most of those people with programming skills were in the scientific community. Analyzing what the data journalists had to say time, convenience and cost are some of the primary impacts of technology on data journalism.

Tools

Another recurring theme of the interviews with the data journalists was the use of tools by data journalists such as excel, Google refine, SPSS, R statistics package, Tableau Public and others. One of the journalists working for the Los Angeles Times said that having tools available helps the journalists to focus on the content of the story they are working on. "If not for these tools, one will have to sit for hours trying to figure out codes. Now journalists can use point and click interface to analyze the data. Technology has made really labor intensive work easier for journalists." The journalists brought up the point of working with the programmers who know how to build tools but they may not be like journalists who can talk to actual people for stories. Therefore there is a mutual relationship where the journalists learn some basic skills of working with numbers, statistics, excel, SPSS (a point and click interface) that is the result of technological advancements. One of the journalists working with *The Guardian* said that when there is a story that uses a huge amount of data using these tools helps the journalists analyze the data within a day, which would not have been possible before.

Story is the key

While the data journalists agreed that the technological advancement has helped a lot in making it convenient for them to work with data and produce data-driven stories, they also stressed the importance of the story in data journalism. The data journalists emphasized that the

story is the core of any type of journalism. If the story is not good then using any kind of data or cool infographic would not be helpful. One of the journalists working for the Los Angeles Times pointed out that there are some new data journalists who sit in the corner and try to make the sleekest data visualization. But they do not want to talk to real people. They just want to interact with their computer. But this is not what journalists are supposed to do. Journalist has to be able to go out and talk to people to add human element to the story. Therefore, traditional journalism skills of interaction with the source and writing are equally important. All the journalists interviewed stressed the importance of the content of the story. For them content of the story helps to signifies whether the story needs to be supported by data or not.

Discussion

To answer the research question technological advancements have made the practice of data journalism easier and faster for the journalists. With the development of robust computer software and tools the journalists are able to work on robust data driven stories in a relatively short amount of time with more accuracy and reliability with less cost. The data journalists also emphasized the skillsets that journalists need to have in order to work with numbers. The data journalists who were interviewed agreed to the theme that a journalist does not need to be a programmer to make sense of data and work with data. A journalist needs to be willing to learn and have some basic knowledge of working with basic numbers and statistics like percentage, ratio, mean, median and others to be able to make sense of the data and work with numbers.

In the perspective of technological determinism and sociomateriality, interviews with these data journalists show that the data journalists agree that technology does impact their work. One of the journalists even said that data journalism exists due to technology. This statement is close to the technological determinism perspective that keeps technology in the center of the

stage. But the data journalists were also saying that being able to code or program and analyze

the data is not enough. A journalist should be able to talk to people to give human perspective to

the story. This is closer to sociomateriality approach that takes technology from the center of the

stage and replaces it with something else like developers of the technologies or social and

cultural attributes related to them.

Limitation

This study is based on the interviews with 5 data journalists. This is not a big enough

sample size to make any generalization about how the technological advancements impact the

data journalism. Further research on how the technology impacts the practice of data journalism

should interview or survey more journalists to get a more robust idea on what the data journalists

think is the impact of technology on their work and how data journalists adopt the new

technologies. This study can be used as pilot study on the impact of technologies on the practice

of data journalism.

Conclusion

Data journalism is a trending journalism practice these days. Data journalists work with

numbers and statistics to find the patterns and stories hidden behind those numbers. Working

with the large amount of data can be very time consuming. The technological advancement,

especially the developments of easily accessible software and tools have made it much easier for

data journalists to work with data. It is essential for a data journalist to be able to work with

numbers and large datasets and have the skills required for data analysis. However, the most

important skill that a data journalist needs to have is to be able to understand the story and ask

the important questions (Sirkkunen, 2011). As this study has showed, journalists agree that data

journalists need to be able to learn new skills to work with numbers but not at the expense of

traditional skills of a journalist to be able to talk to people and know what the content of the story

is.

Pavlik (2000) writes that there are there are at least four ways in which technological

change influences journalism: how journalists do their job, news content, structuration of

newsroom, and the relationships news organizations have with each other, journalists and other

publics. Since data journalism is an aspect of journalism, technology does influence its practice.

After the online journalism data journalism can be the next big thing in journalism. "Data-driven

journalism improves the way journalism can contribute to democracy-especially at a time when a

growing number of data sets are released by governments" (Parasie & Dagiral, 2012, p. 854). As

growing number of news organizations such as *The New York Times, The Las Angeles Times,*

The Chicago Tribune, The Guardian, Pro Publica are incorporating data desk in their newsroom

more research in the practice of data journalism would be helpful in developing models and

theories that are unique to the practice of data journalism.

References

Al Anderson, H. (1977). An empirical investigation of what social responsibility theory means.

Journalism Quarterly, 54(1), 33-39.

Bell, E. (2012). Journalism by numbers. Columbia Journalism Review, 51(3), 48-48.

Chandle, D. (2000). Technological or Media Determinism. Retrieved March 27, 2014 from

www.aber.ac.ok/media/DocumentsItecdet/tecdet.html.

Cohen, S., Hamilton, J.T., & Turner, F. (2011). Computational Journalism. *Communications OF

THE ACM, 54*(10), 66-71. doi: 10.1145/2001269.2001288.

Cohen, S., Li, C., Yang, J., & Yu, C. (2011). Computational Journalism: A call to Arms to

Cox, M. (2000). The development of computer-assisted reporting. Chapel Hill: University of

North Carolina. Database Researchers. CIDR, 148-151.

Diakopoulous, N. (2012). Cultivating the landscape of innovation in computational

Flew, T., Spurgeon, C., Daniel, A., & Swift, A. (2011). The promise of computational

journalism. *Journalism Practice, 6*(2), 157-171.

Geib, S., Jackob, N., & Quiring, O. (2013). The impact of communicating digital technologies:

How information and communication technology journalists conceptualize their influence

on the audience and the industry. *New Media and Society, 15*(7), 1058-1076.

Gray, J., Bounegru, L. & Chambers, L. (2012). The Data Journalism Handbook. Sebastopol, CA:

O'Reilly Media.

Houston, B. (2010). The future of investigative journalism. *American Academy of Arts &

Sciences, 139*(2), 45-56. doi: 10.1162/daed.2010.139.2.45.

Jacobson, T.L. (1989). Use of commercial data bases for television reporting. *Journal of

Broadcasting and Electronic Media, 33*(2), 203-208.

Jacobson, T.L., & Ullman, J. (1989). Commercial databases and reporting: Opinions of

newspaper journalists and librarians. *Newspaper Research Journals, 10*(2), 15-25.

journalism. Retrieved from http://www.gplacencia.com/sites/default/files/87863948-

Cultivating- the-Landscape-of-Innovation-in-Computational-Journalism.pdf.

Kaul, V. (2013). Journalism in the Age of Digital Technology. Online Journal of Communication

and Media Technologies, 3(1), 125-143.

Lewis, S.C., & Usher, N. (2013). Open source and journalism: toward new frameworks for

imagining news innovation. *Media Culture and Society, 35*(5), 602-619. Doi:

10.1177/0163443713485494.

Mayo, J., & Leshner, G. (2000). Assessing the credibility of computer-assisted reporting.

Newspaper Research Journal, *21*(4), 68.

Messner, M., & Garrison, Bruce. (2007). Journalism's 'dirty data' below researcher's radar.

Newspaper Research Journal, 28(4), 88-100.

Miles, M., Huberman, A.M., & Saldana, J. (2014). *Qualitative data analysis: A method

sourcebook*. California: Sage Publications.

Orlikowski, W. (1991). The Duality of Technology: Rethinking the Concept of Technology in

Organizations. *Organization science, 3*(3), 1-40.

Parasie, 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.

Pavlik, J. (2000). The Impact of Technology on Journalism. *Journalism Studies, 1*(2), 229-237.

Powers, M. (2012). In forms that are familiar and yet-to-be invented: American journalism and

the discourse of technologically specific work. *Journal of Comunication Inquiry, 36*(1),

24-43. Doi:10.11771019685991426009.

Rogers, S. (2012, March 28). Data visualization DIY: our top tools. *The Guardian*. Retrieved

from http://www.theguardian.com/news/datablog/2012/mar/28/data- visualisation-tools-

free

Rouse, M. (2012, Feb). Definition infographics. WhatIs.com. Retrieved March 25. 2014, from

http://whatis.techtarget.com/definition/infographics

Schafraad P., Wester, F., & Scheepers P. (2006). Using 'new' data sources for 'old' newspaper

research: Developing guidelines for data collection. *The European Journal of

Communications Research, 31*(4), 455-467. doi: 10.1515/COMMUN.2006.029.

Schellong, A. & Stepanets, E. (2011). Unchartered Waters. The State of Open Data in Europe.

CSC. Public Sector Study Series 01/2011

Seidman, I. (2013). *Interviewing as Qualitative Research*. New York: Teacher's

College Press.

Sirkkunen, E. (2011). Trends in Data Journalism. Next Media. Retrieved on January 18, 2014

from http://virtual.vtt.fi/virtual/nextmedia/Deliverables-

2011/D3.2.1.2.B_Hyperlocal_Trends_In%20Data_Journalism.pdf

Schudson, M. (2010). Political observatories, databases & news in the emerging ecology of

public information. *Daedalus, 139*(2), 100-109.

Schudson,M. (2010). Political observatories, databases & news in the emerging ecology of

public information. *Daedalus, 139*(2), 100-109.

Spyridou, L., Matsiola, M., Veglis, A., Kalliris, G., & Dimoulas, C. (2013). Journalism in a state

    of flux Journalists as agents of technology innovation and emerging news practices. *Sage*

    Journals, 75(1), 76-98. Doi:10.1177/1748048512461763.

Tankaard, Jr. J.W., & Lasorsa, D.L. (2000). Teaching strategies for analysis in computer-assisted

    reporting. *Journalism and Mass Communication Educator, 55*(3), 14-26.

What is Data Journalism (n.d.). Retrieved March 26, 2014 from http://visuals.ly/what-data-

    journalism.

Yarnall, L., Johnson, J.T,. & Ranney, A. (2008). How post-secondary journalism educators teach

    advanced CAR data analysis skills in the digital age. *Mass Communication Educator,*

    63(2). 146-164.