UNIVERSITY OF GOTHENBURG

MASTER PROGRAM IN INVESTIGATIVE JOURNALISM

THE EFFECTS OF OPEN SOURCE'S DUALITY ON DATA JOURNALISM

SADETTIN DEMIREL

SUPERVISOR: JENNY WIIK, PHD

MASTER'S THESIS

GOTHENBURG, JUNE 2019

TABLE OF CONTENTS

ABSTRACT

The purpose of this study is to find out how open source community influences the data journalism practices, and to what extent it contributes to solving the challenges that hinder the integration of data-driven journalism into newsrooms. The study approached the subject by drawing from the previous work of Lewis and Usher (2013) which proposes applying the duality of open source which constitutes of open-source culture (values, principles, ethics) and open-source materiality (software, coding libraries, etc.) on journalism to reconsider possible innovations. For the first time this study concerns with the effect of open source's duality only on data journalism and its problems. To find out the effects of open source community's binary structure on data journalism and its potential contributions to tackle with the constraints of data journalism a qualitative approach was preferred and in-depth interviews with 10 data journalists and data journalism scholars were conducted. The main findings reflect that majority of participants seem to be motivated to follow along with transparent and open way of practicing data journalism while the content of datasets, the license of data sources, newsrooms' policies and fear of having competitors, lack of time to working and documenting the data analysis are viewed as barriers for this way of practices. Similarly, although many of the participants acknowledge that duality of open source community has a great deal of influence in the emergence of data journalism, the half of participants do not see open source mindset as a necessity to do data journalism but only a plus. More importantly, it is mentioned that sharing datasets and codes for open and transparent data journalism may not be aligned with some of the journalistic norms and rules. The main arguments of this idea are the closed source culture of journalism, journalistic autonomy, journalistic credibility, journalism's business mindset, ethical and legal issues of publishing datasets while other half argues that these tensions will disappear as the news reporting digitalizes, the needs of audience changes or that overlook these tensions and saw open source as a driven force to improve data journalism further. Lastly, the majority agreed on the benefits of open source's duality to integrate data journalism practices into newsrooms and solve related challenges facing this integration.

1.INTRODUCTION

This study concerns with how open source movement contributes or influence data journalism practices, and to what extent it plays a productive role in the integration of data journalism into newsrooms. The proposed study scrutinized the subject above with the dual structure of open source movement which stands for open source culture (open source values, principles, ethics) and open source materiality (software, coding libraries, etc.). While the materiality of open source is the software and coding languages that further help working with data, the ethos or moral side of open source is set of values and principles such as transparency, experimentation (tinkering), iteration and participation (Lewis and Usher, 2013, p. 607). Drawing from the materiality and principles of open source, the study further tries to find out the effects of open source on data journalism practices and its potential contributions & solutions for the integration of data journalism into the newsrooms, despite professional journalistic values and norms may clash with open source principles

Data journalism is one of the rising trends that strives to use quantification and data analysis for the journalistic purposes in the news industry. This quantitative turn of journalism also was not overlooked by scholars and academics. More and more scholars from the US, UK, Canada, European countries have studied the new phenomenon's integration and its implications on the current media landscape. However, there is a lack of scholarly work that how open source movement implicate the data journalism activities with its software and cultural aspects. Since, data journalism is represented as a democratization of data, tools, and methods and it is mainly explained as more open, a transparent and collaborative journalistic practice compared the past journalistic practices, questioning open source movement's role in data journalism were addressed to clarify this subject and fill this gap. Along with the effect of open source's duality, as a possible innovation its potential contributions to overcome data journalistic challenges were included to reveal to what extent open source helps the integration of data journalism into traditional newsrooms. More importantly, to date, until this study, influences of open source community on particularly data journalism are not addressed with scholarly works. Although Lewis and Usher (2013) proposed to apply open source model on journalism to reconsider possible innovations, based on their preposition this study narrows the scope down, went further and try to explicate the implications of open source software that have been used by data journalists and open source principles (transparency, openness, collaboration etc.) that have been accepted as unofficial ethical framework to do data journalism on data journalism practices.

Moreover, in the first chapters of this thesis to understand the open source community its cultural and material aspects, the definitions of open source and its relation and decoupling with free software, historical development of open source ideals and open source software discussed. How the movement or community is born with the technological breakthrough of computing technologies and how it is divided into two similar community free software and open source because of its clash of values influencers such as Richard Stallman and Eric Raymond possess, were covered with details. Later it emphasized that this divide in the community paves the way for the application of open source mindset into different fields and facilitates to flourish new movements and communities like Wikipedia, Open Access and Open Data which has a partial role in the rise of data journalism. After that these new modulations of open source thinking and binary structure of open source community were explained with the prepositions of Lewis and Usher (2013) which suggest applying open source principles in journalism to boost innovation.

Furthermore, since the study sets out a course to find out the influence of open source community and its potential contributions on data journalism practices, definitions, and typology of data journalism, how it is emerged and became a journalistic trend along with its common problems and challenges were rigorously pointed out with the relevant literature: academic articles, case studies, and media reports. In this section, the roots of data journalism and its relationship with open source and open data community were discussed. Finally, to answer the research questions below, findings were evaluated and discussed.

2.RESEARCH QUESTIONS

The study asks how the dual aspect of open source (Lewis and Usher, 2013, p.606) affects data journalism workflow and to what extent open source mindset facilitates solving data journalism related challenges facing the newsrooms?

- How open source movement influence data journalism with its material existence (software, coding libraries) and cultural values and principles?
- Does applying open source mindset (both its materiality and cultural principles) in data journalism facilitate to deal with its challenges of working with data for journalistic purposes.

3.METHODOLOGY

A qualitative approach was preferred to study the current subject. The qualitative approach is chosen because of two main reasons. The first reason is that the qualitative approach gives the researcher the ability to probe deeper into the subject and provide a deeper understanding of the issue (Silverman, 2006; Flick, Kardoff and Steinke, 2004; Tracy, 2013). The second one is that the qualitative method offers a more flexible workflow for the researcher and the subject that will be studied (Silverman, 2006; Tracy 2013). Given the subject that I scrutinized in this study is consist of a variety of field (open source, data journalism) the need for deeper findings and exploration regarding influence of open source community in data journalism make qualitative approach the right choice for this study. For data collection part 10 semi-structured interviews with data journalists (data editors, developers/programmers) and data journalism instructors, etc. were carried out through skype calls and in person. The reason why I preferred semi-structured interview as a data collection techniques is because semi-structured interview allows researchers to ask follow up questions by remaining loyal to a specific structure and delve into the more details of the subject as the conversation goes on (Brennen, 2012). Thanks to semi-structured interview technique, I have asked not only specific questions in the interview guide (please see Appendix) but also follow up questions that may be triggered with their answers to reach through insights regarding the subject.

Here is the list of participants who were interviewed:

- P1, Participant 1 former journalist, data journalism instructor from Sweden
- P2, Participant 2 data journalist from Sweden
- P3, Participant 3 former data journalist, data journalism instructor from Sweden
- P4, Participant 4 data journalist from Sweden
- P5, Participant 5 data journalist from Qatar
- P6, Participant 6 data journalist and instructor from Europe
- P7, Participant 7 data journalist, editor and instructor from Kyrgyzstan
- P8, Participant 8 data journalist from Sweden
- P9, Participant 9 former data journalist, data journalism instructor from Turkey
- P10, Participant 10 data journalist, news editor from Turkey

The many of the respondents (6) for the study were picked among data journalists or data journalism lecturers from European countries (mainly Sweden, France) because the researcher

was in Sweden during the time of this study and reaching out to Swedish journalists was easier than the data journalists from other countries. Also, data journalists or lecturers from Turkey (2), Qatar, Kyrgyzstan were included in study to balance the views. Since data journalism emerges in US and Europe and it is predominantly practiced in these regions, the ratio of respondents' distribution regarding their countries and media ecosystem is reasonable and suitable. Also, data journalism is still a niche journalistic practice and the practice is not completely different from one country to another. Therefore, main views of current 10 respondents would not entirely change if another 10 respondents were interviewed instead.

The respondents answer 15 question in the interview (see the appendix). These specific questions divided to serve the two main research questions. While 12 questions shed lights to the open source influence in data journalism, 3 of them focused on the open source's role in integration of data journalism. The detailed interview questions began with the questioning respondents' awareness regarding open source movement and data journalism. Later, they concentrated on open source influence and possible tensions and clashes between open source principles and journalistic values. Finally, last 3 questions scrutinized the open source solutions for data journalistic challenges. The 2 interviews were conducted face to face and 8 respondents answered questions via Skype calls. The interviews lasted between 40 minutes to 60 minutes.

Before the interview respondents were guaranteed full anonymity but all the participants allow their name to be used if it is needed. The name of respondents coded as participant 1,2,3 due to the ethical reasons. Prior to the interviews being carried out, a description of the study and interview questions were sent to respondents. During the interviews the researcher gave respondents enough time to answer the questions and demanded further elaborations when it is needed.

When it comes to the validity and reliability of this scientific work, given the subject of the study, which is open source's influence in data journalism, is novel field and it had not been covered with previous scholarship thoroughly, a qualitative approach had been applied to yield deeper insights regarding the open source's role and influences. Due to the same reasons semi-structured interview, a flexible data collection technique had been conducted to explore detailed findings without losing the focus of the study. The downside of the preferred method and technique is lack of comparability at some part of the work since it included a flexible approach to ask respondents not only predefined questions. However, the advantages of the semi-structured interviews outweigh its disadvantages as it yields comprehensive explorations and

novel findings. Also, although the chosen method may lack enough comparable aspect compared to the quantitative approach, the study has a clear focus and original research questions with a supportive interview guide that contains relevant questions to scrutinize the subject of the study thoroughly. Another point that strengthens and somewhat weaken the validity of this work is that 10 respondents that were involved in the study. The downside is that distribution of the respondents concentrated on mainly European journalists. This may be improved by adding more diverse respondents from different parts of the world but as I mentioned earlier in this chapter the distribution of the respondents regarding their location or journalism traditions is relevant and reasonable because data journalism and open source are more dominant and emergent in US and European newsrooms. Besides there are 4 respondents that represented the other parts of the world to balance the views. The advantages of the chosen respondents for the study is that either they practice the data journalism in the newsrooms or teach it to new generations of journalism. Hence their perspectives and opinions regarding the open-source and data journalism came not only from the field but also from schools, curriculums. This made the findings and outputs from the respondent

4.OPEN SOURCE MOVEMENT

4.1. What is open source?

Open source is a social movement (Raymond, 2018; Kelty, 2008) and often used with the term of free software too, but there are tiny or practical differences between two terms because of how open source or free software movement developed through time with the different approaches and actors such as Stallman's (2019) and Raymond's views (2018).

When it comes to the definition, open source software is the center of the movement. Open Source Initiative (n.d.) defines open source software as "software that can be freely accessed, used, changed, and shared (in modified or unmodified form) by anyone" by basing this description to Open Source Definition[1]. According to this definition open source software should have freely distributed source code, the software should be open to modifications and free to be used by many people and groups in many areas and fields. More importantly, redistributed software should grant the same rights and licenses as original software did and should not limit the use of other software. As a prominent influencer in labeling open source Raymond (2018) points out that "open source rejects secrecy and centralized control of creative

---

[1] Open Source Definition by Open Source Iniative (OSI)

work in favor of decentralization, transparency, and open sharing of information". Also, he explains what he means by saying source that is "human-readable source code of computer programs that run on computers but cannot be easily understood or modified by people". Another important point that is made by Raymond is that despite open source was flourished with help of its predecessor Free Software Movement, with efforts of Stallman, it contains less ideological point than the term of Free Software

Compared to open source, free software is an older term which was coined by Stallman and his generation of hackers and contributors (Stallman, 2019). Stallman (2019) simply describes that "it's free software, it gives you freedom". Often free software is perceived as the meaning of price, but Stallman clarifies this notion of being free software as:

> When we call software "free," we mean that it respects the users' essential freedoms: the freedom to run it, to study and change it, and to redistribute copies with or without changes. This is a matter of freedom, not price, so think of "free speech," not "free beer." (Stallman, 2019.)

The rift between free software and open source is also explained by Stallman himself. He stresses that the phrases of free software and open source seem to be same meaning, but free software is a movement that advocates for freedom of computing for the sake of users, whereas open source only utilizes the practical benefits of free software but not its principles. Similarly, Kelty (2008, p.2) argues that free software is a collaborative set of practices to create, distribute software, share its source code and make it available with free licenses.

Moreover, when it comes to the descriptions of open source and free software, these definitions seem to be akin in terms of the attributions and ascriptions that they were assigned. For instance, open source and free software use the same terms to explicate what they are: free distribution, sharing source code, freely accessible, collaborative practice, transparency, openness, freedom. While the differences between these two terms may still be discussed further, there are also reconciliation efforts that bind this two approaches into one by proposing Free and Open Source Software (FOSS) or adding a Spanish and French word "libre", which means free as Free Libre and Open Source Software (FLOSS) (Stallman, 2016). Whilst this turf war may go on between the Free Software and Open Source Community, the term open source will be used to cover software side and philosophical side as the part of Free and Open Source Software for the purpose of this study. Open source is already part of FOSS (Free and Open Source Software) which stands for "collective noun for all software with available source code, adaptable by all, under the limitation that the adaptations should be made available to others" (Wynants and Cornelis, 2005, p.15), it is also a "philosophy of sharing" (Lewis and Usher, 2013, p.603) based

on universal right to access information and knowledge (Coleman, 2013; Kelty, 2008). However, in order to delve into deeper understandings regarding the software aspect of the movement or community, historical developments of free and open source software are going to be discussed in the next section.

4.2. Historical Background

In the 1960s early computer science practices were indicated as the roots of open source movement and ideals (Raymond, 2018). According to Raymond (2018), early computer users hack, modify and share the software as the way needed except without any license or movements. The development of the UNIX operating system in 1970 by Bell Laboratories is reflected as the cornerstone of this movement (Kelty, 2008). In fact, Kelty (2008, p.141) praises this as not only technical development but also the generation set of norms and principles to share source code with the different background of people which are from academics to commercial communities. AT&T company which owns the Bell Laboratories as its subsidiary decided to commercialize the UNIX operating system to use microcomputers with licenses that prevent its modifications and redistributions (Raymond, 2018). Before this, one of the prominent influencer of Free Software and Open Source and known as the father of the Free Software Movement, Richard Stallman, resigned from his job MIT and found the Free Software Foundation to protect the copyright of developers and contributors against malpractices of commercial firms (Raymond, 2018, MIT Sloan, 2000). Also, he started his GNU project to develop a UNIX compatible operating system during 1980s. In 1991 the efforts to develop a system that runs UNIX Operating System in personal computers were completed with achievements of Linus Torvalds and contributors, hackers and programmers (MIT Sloan, 2000; Raymond, 2019) Later the system is called Linux and it is beginning to be improved further with help of contributors. Linux is accepted as a first internet centered operating system and currently is used for Android systems too. While technological innovations are booming new services and coding libraries such as Sendmail, Python, Perl, Apache, the approaches to free software are also splintered with Raymond's article of Bazaar and Cathedral development model of software (Open Source Initiative, 2018, MIT Sloan, 2000, Raymond, 2018). According to Kelty (2008), some groups do not share the same ideas with Stallman's free software ideals and view the commercial market as an opportunity for the movement. Raymond's article in 1997 is accepted as the cornerstone of open source software development with the commercial market and industry (MIT Sloan, 2000; Raymond, 2018). Despite Stallman's free software views regarding market economics, in 1998 a new term, open source,

is started using to describe the commercial-oriented version free software movement with the intersection of business and media communities (MIT Sloan, 2000, Raymond, 2018). The splintered views further go on with the foundation of Open Source Initative to advocate open source software in 1998 (MIT Sloan, 2000). According to Kelty (2008, p.99), while Free Software groups saw open source advocates as collaborators of non-free and proprietary software firms, the latter community views the perceive the former as dogmatic groups. Kelty (2008, p. 109) also describes the motivations of Raymond against the Free Software approach of the Stallman as:

> Raymond was determined to reject the philosophy of liberty that Stallman and the Free Software Foundation represented, but not in order to create a political movement of his own. Rather, Raymond (and the others at the Freeware Summit) sought to cash in on the rising tide of the Internet economy by turning the creation of Free Software into something that made more sense to investors, venture capitalists, and the stock-buying public. To Raymond, Stallman and the Free Software Foundation represented not freedom or liberty, but a kind of dogmatic, impossible communism

Hence, it can be said that the open source movement saw the free software movement as a barrier for the application of free software into the commercial sector. Kelty (2008) agrees that the term open source was coined to utilize the free software development model into market economics and commercial sector.

However even though the movement is split into two communities in terms of their principles an ideals, Kelty (2008, p.112) claims that the fundamentals practices of the movements are not changed at all, "sharing source code, conceiving openness, writing licenses, coordinating projects" went on and on despite this philosophical differences were divided the movement into two factions. Furthermore, that interpretation of free software as an open source model of building software with market forces and commercialization paves the way for the application of free software values, principles and ideals on new fields and practices. According to Kelty (2008, p.3), although utilization of free software and open source mindset into different activities may not be approved inside the free software movement, these new projects such as Open access, Creative Commons and Wikipedia project are the new instances that emerge out of Free Software Movement and they share the same principles and values and connected through similar practices. Aside from these Wikipedia and Open Data movement can also be counted as the fruits of Free and Open Source Movement, and especially Open Data Movement has been viewed as a significant part of the data journalism practices despite opinions on this subject may vary.

4.3. Inspirations of Open source

Kelty (2008, p.98) identifies the key practices of open source movement with the five aspects: sharing source code, defining openness, writing copy-right licenses, coordinating collaborations, and forming a movement. Baack (2015, p.2) interprets that these components of open source are flexible enough to adapt and modify into different fields and areas. Thus, Kelty (2008, p.246) views the new application of open source values and principles to various domains such as Wikipedia, Open Access, Open Data, etc. as the different modulations of open source.

There are several instances of open source practices that vary over many fields. Some of them are Wikipedia, Open Access and Open Data. Wikipedia defines itself as a free encyclopedia and as much as its motto, Wikipedia's one of five pillars[2], which is that Wikipedia is free content that anyone can use, edit, and distribute, also demonstrates the effect of inspiration it got from Free and Open Source Software Movement. Another application of the open source mindset is Open Access[3] in the academic sphere. Open access is described as making academic article and studies open and accessible to anyone who is interested in (What is open access? n.d.). The simple cause of open access is to bring down the pay or subscription walls to the knowledge for who are eager to further develop academic studies and knowledge production.

In addition, Open Data might be the broadest application of open source ideals after Wikipedia. The Open Data Handbook (n.d.) describes that " open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike". As can be illustrated in the definition, availability, and access, re-use and redistributing with universal participation are seem to be the main pillars of open data notion (Open Data Handbook, n.d.). Although the term of open data was also born with early efforts to opening up research data in the 1940s, it has become a buzzword with the Transparency and Open Government Memorandum which is signed by Obama in 2009 (Bode, 2013; Peter, 2013). The memorandum is about the Freedom of Information Act and opening the government data in a machine-readable format for the sake of transparency and accountability (Peter, 2013). Later, data.gov site was a launch to publish the open government data and this policy decision is repeated and applied by many governments in Europe too (Peter, 2013). This may reflect that the open data movement differs from free software and open source software movement since

---

[2] The five pillars of Wikipedia
[3] What is Open Access

it has been adopted by the governmental institutions, but Baack (2015, p.2) argues that "open data activists apply practices and values from open source culture to the creation and use of data" and he also points out that this connects the open data movement to the different interpretation and application of open source movement such as Open Access, Wikipedia, Creative Commons, etc. In the light of the five components, Kelty (2008) proposed and modulations of open source what would be the implications of applying open source mindset into journalism and given the adaptations of open source logic by the many fields and professions in which principles and values can be applied in the context of news reporting?

4.4. Duality of open source and its application to journalism

As can be illustrated in the definitions and historical periods of free and open source software movement, FOSS is consisting of two different components which are both software development and community's philosophical values and ideals that further shed lights community's practices.

This duality of open source is pointed out as it is both an "operating system and social system" (Kelty, 2008, p.57). Drawing from Kelty's argument, Lewis and Usher (2013, p.606) propose that there is a dual emphasis, binary structure, which is consist of "materiality (architectural frame-work) and ethos of open source (moral and cultural context)". They later stress that this duality is significant to understand the application of open source concept or mindset into journalism and examine open source values such as transparency, experimentation (tinkering), iteration and participation, "which connect with and depart from journalism", to reconsider innovation in news-making. (Lewis and Usher, 2013, p.602).

Moreover, the work of Lewis and Usher (2013) is one of the crucial scholarly works that examine the application of open source into news reporting. Although there are a variety of studies that approach the technological influence on journalism, according to them their work differentiates from the previous one which are concentrated on how journalism can be improved through journalism-focused approach to the technology and instead they suggest focusing the journalism as a process and finding news solutions via a technology-focused approach (Lewis and Usher, 2013). They mentioned four principles or values of open source that are related or can be applicable with news-making too: transparency, participation, experimentation, and iteration (Lewis and Usher, 2013, p.602).

Transparency is pointed out as a prominent principle of journalism (Karlsson, 2010; Singer, 2007), in fact, it is argued as new objectivity too. (Appelgren and Salaverria, 2018). Singer

(2007, p.83) explains this connection with a direct relationship between transparency, accountability, and social responsibility. According to Karlsson (2010, p. 537), there are two types of transparency that can be applied to news reporting: "disclosure transparency and participatory transparency". While the first one refers to the sharing methods and sources through links to show how the information is acquired, the latter is about enabling users to influence news workflow (Karlsson, 2010, p.538). Lewis and Usher (2013, p.607) accept the idea that journalism tends to be transparent, but open source would further improve these efforts and provides broader range of transparent workflow for the news making, because open source working practices require transparency and openness of software and codes to further improve with the collective efforts and intelligence.

Participation is another value that is identified by Lewis and Usher (2013). Participation aspect of users is generally coupled with the concept of convergence culture, user-generated content, media participation, digital and participatory journalism or citizen journalism, etc. Also, Jenkins (2006, p.3) mentions the active participation of audiences or consumers and argues that users obtain an active role in the ever-growing digital media with their collective intelligence. When it comes to the participation practices of open source, distributed participatory workflow of open source software development by "giving enough eyeballs to all bugs" in the code is assumed to be led to better software and code (Raymond, 2018). In this sense, Lewis and Usher (2013, p.609) highlight that the principle of participation may play in an important role to make users and readers more active in the news workflow, but what they suggest seems to be more like assigning the audience monitorial and supervisor parts that further contribute to the process rather than commenting the publication or the news articles.

The other open source principles that are suggested to apply on journalism innovations are the experimentation and iteration. Main components of these values are the basis for considering news reporting as a process rather than a product (Lewis and Usher, 2013, p. 608). As the software development requires experimentation and iteration through learning by failing and playing with the product to further improve, journalism and its products as stories can be better through toying, failing and further experimenting and iterating (Lewis and Usher, 2013, p.608). On the other hand, although open source community and its moral values can be interpreted to hold value of collaboration, I think collaboration must be addressed properly to understand the importance of open source community in data journalism, because I, personally, found this one of the main missing principles of Lewis and Usher (2013) proposition of open source values. Of course, the collaboration between news organizations happened before the digital age (P6, 2017) but what I refer here is the collaboration between certain actors inside and outside of

newsrooms. In fact, collaboration among people who have different mindsets and skillsets can be seen in journalistic coverage of data leaks from Wikileaks to Panama Papers. Data journalism or data-driven journalism terms have been also used to describe the collaborative and quantitative way of reporting overwhelmingly after the Wikileaks. Despite prior to Wikileaks different actors from open source and software community started to be employed by newsrooms, the need for collaboration between journalists, programmers, developers, etc. intensified with the amount of information and data overload the current skillsets of average reporters.

In summary, Lewis and Usher (2013) propose these open source values and principles that connect with or relate with acts of journalism to reconsider the innovations of news reporting with open source mindset. More importantly, these values are stressed as the ethical frameworks of open source that is applicable for both materials (architectural, software) and cultural (moral, values, principles) side of open source. In the light of these, this study tries to build on the works of Lewis and Usher (2013) and find out role and effects of open source on data journalism by probing into both software and cultural side of the implications. Also, it asks to what extent open source mindset might be a solution for the integration of data journalism into newsrooms. For these reasons, the detailed descriptions of data journalism, how it has emerged as a new popular trend in the existing media landscape and its current challenges and barriers are going to be discussed in the next chapters.

5.QUANTITATIVE TURN OF JOURNALISM: DATA JOURNALISM

5.1.What is data journalism?

Data journalism has different definitions were pointed out by journalists and scholars (Stray, 2011; Gray, Chambers and Bounegru, 2012; Howard, 2014; P6, 2015). Most of the definitions related to the importance of the news story but there are also definitions that emphasize the important steps in this type of news reporting. Bradshaw (2012, p.2) defines data journalism by saying doing journalism with data does not mean much, and he describes that data journalism is the converging the traditional way of news reporting with the ability to covering the story with the large range of data available today. Similarly, Rogers states that "data journalism is still about telling story best way possible, and it is just journalism" (Rogers, 2011). While Stray (2011) differently defines that "data journalism is obtaining, reporting on, curating and publishing data in the public interest", whereas P6 (2015) questions the specifications of data in news reporting by arguing that doing journalism with data does not define data journalism and also, he insists that a particular feature of data journalism is to do news reporting with structured data. A holistic definition comes from Aron Pilhofer:

> 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 (Gray, Chambers and Bounegru, 2012, p. 6)

Similar to the Stray's (2011) definition Howard (2014, p. 4) describe data journalism as "…gathering, cleaning, organizing, analyzing, visualizing, and publishing data to support the creation of acts of journalism". But he also describes the term of data journalism simply as" the application of data science to journalism, where data science is defined as the study of the extraction of knowledge from data" (Howard, 2014, p. 4). While Howard (2014, p.5) acknowledges the computer science aspect of data journalism by mentioning data science, he also proposes four components of data journalism: (1) using data as a source to gather and

verify, (2) interrogating data with statistical knowledge, (3) visualizing the findings to present significant insights, (4) publishing the data and methods behind the data journalistic work. Although the last one is not fully approved by Howard (2014, p.5), it is highly relevant with open source values such as transparency and openness. In fact, Howard (2014) later on connects the dots between open source developers' practices of publishing data, code and methods behind the data-driven journalism. According to him, being transparent in data journalism workflow enables people to hold account and check the use of methods and analysis behind the publication and also this way of working, publishing datasets, codes resembles with open source developers' using GitHub to push and pull the updates of software (Howard, 2014, p.19). Also, he doesn't hesitate to point out that transparent practices of data journalism particularly opening up datasets may clash with journalism's own principles to protect the sources, for this reason not all data should be shared (Howard, 2014, p.19). In this context it can be clearly seen that open source practices have influenced data journalism whether the publishing datasets and opening up the codes and methods behind the news articles come from journalism's own ethical frameworks or this way of practices may put sources in jeopardy and contradict with the one of the main norms of news reporting. In my opinion, I disagree with the second scenario that publishing datasets may put the sources at risk, because how the datasets will be published is still at the hands of journalists or programmers who successfully anonymize the datasets to avoid any malpractices or exploitations of publishing data.

A comprehensive study to cover the definitions of data journalism and come up with a broad and inclusive descriptions of data journalism craft is done by Royal and Blasingame (2015). According to them, in order to explain what the term of data journalism is, various aspects such as "process, product, convergence of fields, traditional, outside influence and skills" have been used frequently (Royal and Blasingame, 2015, p.40). They also found that the most frequently used component in definitions is process (Royal and Blasingame, 2015, p.40) and they later combine the common dimensions to devise their own definition:

> Data journalism is a process by which analysis and presentation of data are employed to better inform and engage the public. Its roots are in the fields of computer-assisted and investigative reporting, but data journalism products may add engagement through customization and user contribution made possible by web development and programming techniques. (Royal and Blasingame, 2015, p.41).

Along with Philhofer's (Gray, Chambers and Bounegru, 2012) and Howard's (2014) definitions, this definition successfully delineates the flexibility of data journalism with its roots and features. Also, it can be argued that open source has a role even in definitions of this new

emergent way of doing journalism from using process to programming, opening data sources and code, user contribution, application of data science and customization etc. As a result, these are the definitions of data journalism that were pointed out by journalists who practice data-driven news reporting and scholars who study the current trend.

Despite there are different and various definitions for data journalism, the trend has already found itself a place in journalism literature through news media reports (Howards, 2014; Rogers, Schwabish, and Bowers, 2017), case studies (Parasie and P9irel, 2013; Appelgren and Nygren, 2014; De Maeyer et al., 2015; Karlsen and Stavelin, 2014; Fink and Anderson, 2015; Borges-Rey, 2016) and academic and research papers (Lewis and Westlund, 2014; Coddington, 2015; Figl, 2017; Young, Hermida and Fulda, 2017) in previous years. Of course, these developments come with the news media's inclinations towards data-driven news reporting, for example, Guardian DataBlog[4], New York Times's UpShot[5], Five Thirty-Eight[6], ProPublica[7], LA Times Data Desk[8], De Zeit[9], Washington Post Wonkblog[10] (Rogers, Schwabish, and Bowers, 2017, pp. 22-23). In addition, according to latest Google News Lab's report on data journalism, "data journalism has become more mainstream than any other time in the history" (Rogers, Schwabish and Bowers, 2017, p.3). Also, the report supports this argument with the finding which is that 42% of journalists use data regularly to tell stories, and more than half of the newsrooms have a dedicated data reporter on staff (Rogers, Schwabish and Bowers, 2017, p.3).

5.2.Typology of data journalism

Aside from the definitions that tried to clarify what is data journalism is, Coddington (2015) proposes a comprehensive framework to examine the proliferated quantitative reporting practices from computer-assisted reporting to data journalism and computational journalism. He stresses that there emergent trends of journalism are the effect of interaction between journalists and programmers, and also as the programmers and developers are beginning to be employed in the newsrooms, so journalism is influenced by the technical and cultural aspects

---

[4] https://www.theguardian.com/data

[5] https://www.nytimes.com/section/upshot

[6] http://fivethirtyeight.com/

[7] https://www.propublica.org/

[8] http://www.latimes.com/local/datadesk/

[9] http://www.zeit.de/datenjournalismus

[10] https://www.washingtonpost.com/news/wonk/?utm_term=.722eb39f09fd

of open source (Coddington, 2015, p.332). In his typology to analyze and explain these seemingly same but refers the different set of practices and crafts of journalism, Coddington (2015, p.337) introduces four parameters: professional orientation, openness, epistemology and vision of the public. The first element is about how these journalistic practices approach participation in the news production process: one side is prone to be broad, open and networked participation and other refers to the professional competency (Coddington, 2015, p.338). In this context data journalism represents the overlapping practices between traditional professional practices & norms and new ways telling story with data (Coddington, 2015, p.339), because in this axis while computer-assisted reporting views data as an evidentiary tool to make the case for investigative reporting with a traditional journalistic expertise, computational journalism is prone to be based on purely distributed participation and use of technical abilities. Thus it can be said that data journalism is a hybrid form of reporting that put professional values and newly earned principles and ideals to work despite as Lewis (2012), Lewis and Usher (2013), and Howard (2014) argue these principles may clash or might not be suitable with journalistic norms as can be seen in the case transparency and opening up the data sets. The second parameter is openness which compares the mentioned practices with their being transparent and opaque (Coddington, 2015, p.340). Transparency is seen as one of the main components of journalism (Karlsson, 2010). Coddington (2015, p.340) indicates that the significance of having transparent workflow is appreciated by the data journalism, despite transparency is sometimes viewed as a threat for professional journalistic autonomy (Lewis and Usher, 2013). More importantly Coddington acknowledges (2015, p.340) that transparency in the sense of putting data sets out in the open for everyone's access does not hurt the aims of the story, instead "it simply adds to it". Hence being transparent and open is valued in data journalism as important as in open source community. Apart from networked participation and transparent open workflow of data journalism, Coddington (2015, pp. 341-343) strives to explain quantitative turn of journalism practices with their epistemology which stands for how news and information are obtained and analyzed, and their vision of public which refers how these journalistic practices see the public: as an active or passive representation audience. In the epistemology framework Coddington (2015, p.341) compares the targeted sampling and big data methods of knowledge production so that computer-assisted reporting stands for more of a social science perspective of journalism, whereas data journalism utilizes exploratory analysis and correlations & associations in the ever-growing size of big data and information. The last but not the least parameter is about how these quantitative news reporting practices approach the public. According to Coddington (2015, p.343), computer-assisted reporting is bound to the

traditional journalism with a strong focus on using computers for investigative journalism, thus it sees the public in the passive role and its task to illuminate and inform public by holding power accountable. This is a way of seeing the world in the ideology of journalism and attributes journalism and its practitioners' authority, prestige and autonomy over any kind of criticism (Deuze, 2005, p.445). On the other hand data journalism consider the public in an active role, for this reason, it retains a transparent and open workflow and strives to publish datasets, methods, and codes behind the practice for the public's monitorial role: accessing data and insights, examining works and reaching further improved understanding regarding the issues (Coddington, 2015, p.343). Given the typology of data journalism, it is clear to indicate that data-driven journalism has overlapping aspects that bind professional journalistic values and open source mindset together. In fact, Coddington (2015, p.344) summarizes data journalism as a journalistic practice that "have arisen from the intersection of professional journalism with open-source culture".

In summary Coddington's (2015) typology and some definitions of data journalism retains similar practices in terms of principles and workflow, but data journalism is not the first systematic example of the use of quantification or data inside news making. In order to answer the questions of how this new trend emerged and picked up by newsrooms and journalists are going to be discussed in detail. This will facilitate to assess the role of open source on data journalistic practices from beginning to now.

5.3. Developments of data journalism

Despite fact that data journalism is seen as an novel form of news reporting with a mixture of data analyses and visualizations (Appelgren and Nygren, 2014, p.394; Gray, Bounegru, and Chambers, 2012, p. 17), reporting and working with data have a historical background (Bounegru, 2012, p. 18) and also it may have since the first journalistic practices (Howard, 2014, p.1). For instance, Guardian former data editor Simon Rogers states that the first example of data journalism at Guardian was in 1821 and it is a data table that shows a number of students who were in school and the costs per school in Manchester (Bounegru, 2012, p.19). Also, Howard (2014, p.9) stresses the same argument that statistics began to be used in news reporting as Guardian did in the 19th century and financial stock market data also had been published by Dow Jones & Company in Wall Street Journal before the financial market data went digital. Moreover, Liliana Bounegru (2012, p.18) points out that most akin example of data journalism came from computer-assisted reporting (CAR) that was the first systematic way to collect and

analyze data with computers in order to improve the news coverage. While the initial example of CAR was in 1952 to predict presidential election' result (Cox, 2000, p.6), in the 1970s Meyer's precision journalism term was used to describe this way of news reporting, which is based on the use of social science methods in journalism (Coddington, 2015, p.334). By the end of the 20th century, CAR was used in mainly investigative reporting projects because microcomputers and computing technologies facilitate more powerful scrutiny and monitoring on official statistics, thus computer, as new technology, facilitated comprehensive investigative journalism practices (Howard, 2014, p. 9).

On the other hand, Howard (2014, p.12) confirms that the term of computer-assisted journalism is outdated in the 21st century because computers are not an innovation in the newsroom anymore, and ability to use computers in newsrooms are the necessity rather than a sophisticated skill that the minority of reporters possess in the past. Also, he acknowledges that web technologies that enable journalists to employ data journalism processes (scraping, cleaning, visualizing etc.) on raw data for journalistic purposes, growing open data movement and collaboration in the newsroom are the distinguishable features of data journalism compare to the computer-assisted reporting (Howard, 2014, p.14). Similarly, while Bounegru (2012, p. 21) indicates there is a "continuity and change debate" about calling journalism that employs computer techniques on news reporting, she distinguishes data journalism from computer-assisted reporting by arguing data journalism focuses on data itself in all journalism processes rather than just uses it to improve investigative reporting as computer-assisted journalism did. Even though there are two communities argue about the labelling this new quest of journalism, Bounegru (2012, pp. 21-22), who tries to reconcile the turf war upon the terminological debate, states that it is not useful to argue about what is data journalism or not, instead assuming data journalism part of the previous journalism trends would be more suitable solution because the most media practices have also historical bonds with former one. Also, she asserts that even though data journalism is new kinds of journalism that use cutting-edge web technologies, it may benefit from the critical approach and experiences of computer assisting reporting (Bounegru, 2012, p.22).

Furthermore, the discussion about labeling quantitative turn of the journalism may continue in academic papers while it seems that the term of data journalism was started to use widely in US and European newsrooms aftermath of WikiLeaks and its coverage by Guardian, New York Times and Der Spiegel (Bounegru, 2012, p.17). Nowadays data journalism is a mainstream term and popular practice and even if it is very widespread in the US and Europe (Heravi, 2017), this trend also has effects on small newsrooms in other parts of the world, La Nacion in

Argentina (Mazotte, 2017), Swiss daily Neue Zürcher Zeitung, the US South Florida Sun-Sentinel, or the Ukrainian Texty (Splendore et al. 2015, p. 139). Also, data journalism practice has its own awards in journalism industry which is also evidence of the pervasiveness of data journalism that can be seen in submissions from different countries to Data Journalism Awards[11] that is arranged by Global Editors Network. There are also networks such as European Data Journalism Networks[12], Data Journalism DEN[13] that support and promote data journalism and data-driven reporters.

Aside from the number of news media that aims to integrate data journalism practices into their workflow, there are studies that reflect benefits of data journalism in terms of both quantity and quality for news media nowadays. According to the global data journalism survey (n = 181 respondents from 43 countries) that were conducted between 3rd December 2016 and 1oth May 2017, 46 % percent of the participants said they have data-driven teams or blogs or desks in their news media (Heravi, 2017). Also, while 65 % of the respondents answered that data journalism facilitates them to produce more news stories, 91 % of them notes data journalism improves the news reporting in their media organizations and along with this 70 % of participants told that they cannot do their work without using as a source (Heravi, 2017).

5.4.Challenges of data journalism facing the newsrooms

As always as the technology progress in any fields of study so the new challenges and barriers arise to keep up with the new developments. Since exploring binary structure of open source's potential contributions or solutions for data journalism challenges is another purpose of this study, finding out these challenges in the literature is matter.

Several studies and media reports in the US and Europe focused on the challenges that constrained newsrooms regarding integration of data journalism and data visualization. Also, these studies delved into not only external challenges but also internal constraints in newsrooms. Hence as the new trends of reporting are applied into journalism, so further challenges and constraints regarding integration issues into journalistic practices arise. Following studies in different countries concentrates on these barriers and challenges. Newsrooms in Sweden points out mainly two constraints: the first is the lack of time to cover news with data, which contains the steps of collecting, analyzing and visualizing data, another

---

[11] https://www.datajournalismawards.org/past-winners/
[12] https://www.europeandatajournalism.eu/eng
[13] https://datajournalismden.org/

is the absence of staff who has required skills to work with data efficiently (Appelgren and Nygren, 2014, p.404). Similarly, the study in United States reflects that newsrooms face four main challenges that may impact their way of works which are lack of time to make sense of raw data in 24 hours news cycle, lack of visualization and data mining tools, lack of competent reporters, and lack of legal supports when reporters work on official data (Fink and Anderson, 2015, p.475). Also, the study focused on Norwegian newsrooms notes similar challenges which are the accessibility of data regarding practical and legal aspects and visualizing analyzed data into understandable formats which readers can make sense of it easily (Karlsen and Stavelin, 2014, p. 45). Moreover, Google News Lab Data Journalism in 2017 report shed lights to the akin challenges and barriers that deter newsrooms to do efficient data reporting (Rogers, Schwabish, and Bowers, 2017, pp. 16-21). According to the report newsrooms in US and Europe suffers from time pressure when they cover stories with data, inconsistent data structures and formats which may cage the data into non-machine-readable formats like scanned pdf, deficiency of required skills, limitations of existing visualization tools and most importantly unclear return of investment (Rogers, Schwabish, and Bowers, 2017, pp. 16-21). Recently I have studied the subject of use of data visualization in Turkish media landscape which approaches the craft of data visualization as a process for the data journalism workflow (Demirel, 2018). During the working with data the challenges and constrains newsrooms and journalist facing are similar to previous studies' findings (Demirel, 2018, p.55).

On the other hand, there are studies that highlight the current challenges of data journalism and data visualization by classifying them into the internal and external constraints and also relating to the organizational and managerial structure of the newsroom (Smit, Haan, and Buijs, 2014, p. 351; De Maeyer et al., 2015, pp. 441-443). In the first study Smit, Haan and Buijs (2014, p. 351) indicate "challenges relating to skills at an individual level, mindset at a group level and management at an organizational level". While the first challenge relates to the diverse data visualization skills that reporters or designer must have to cover stories with data, the second challenges is about mentality of data team which may block collaboration among the departments and lastly the third constraint is based on evolving concept of news in newsroom rather than hiring people from different field by managerial decisions (Smit, Haan, and Buijs, 2014, p. 351). Similarly, the second study mainly assesses the situation in Belgian newsrooms, also distinguishes constraints into three categories, which are "obstacles within the news organization, obstacles outside the news organization, and obstacles that emerge at a more individual level" (De Maeyer et al., 2015, p. 441). So that internal obstacles are lack of time, financial resources and traditional workflow of news reporting whereas external constraint is

that even though Belgium has an open data policy, accessing and using public data are not an easy task (De Maeyer et al., 2015, pp. 441-442). Also, the third challenge in Belgian newsrooms is the journalists' avoidance of numeracy skills which is "a mental block at play rather than real technical impediments" (De Maeyer et al., 2015, p. 443).

6.FINDINGS & DISCUSSION

As can be mentioned in the methodology section qualitative approach was preferred as research and semi-structured interviews were conducted for data collection regarding the research questions. In order to answer the questions, 10 people who have experiences in data journalism as a practitioner or instructor were interviewed face to face or via conference call about 40 - 60 minutes.

- P1, Participant 1 former journalist, data journalism instructor from Sweden
- P2, Participant 2 data journalist from Sweden
- P3, Participant 3 former data journalist, data journalism instructor from Sweden
- P4, Participant 4 data journalist from Sweden
- P5, Participant 5 data journalist from Qatar
- P6, Participant 6 data journalist and instructor from Europe
- P7, Participant 7 data journalist, editor and instructor from Kyrgyzstan
- P8, Participant 8 data journalist from Sweden
- P9, Participant 9 former data journalist, data journalism instructor from Turkey
- P10, Participant 10 data journalist, news editor from Turkey

The five of the participants have experiences in teaching in universities and practicing data journalism in newsrooms or as freelancers. Other five is currently employed as a date editor or data journalist in their news outlets. Also the participants are mainly consists of European journalists, because the current trends of journalism shows that news media in Europe and North America are the first to pick up and practice the data journalistic practices and still dominate the field, even though there are new sparks regarding data savvy journalism in many countries all over the world. The findings are divided into six different titles: Defining data journalism and open source, Use of Open Source Tools and Transparency, Open Source Influence on Data Journalism, Tensions Between Data Journalism and Open Source, Open sourcing data journalism against its challenges in terms of overlapping themes and answers to the interview questions.

6.1. Defining Data Journalism and Open Source

How participants describe the open source and data journalism is significant to assess their views and stances and compare with the descriptions in the current literature regarding these two practices.

First, when it comes to the defining open source the findings reflect that 4 out of 10 participants described the term as in exact terms that are pointed out in the literature. Also, the others associate the term of open source with its values, principles that add data journalism transparent, open, collaborative and participatory workflow. P2 defines the open source as "software that is developed by the community. It can be coding language or software that developed by the community and anybody use, add on that software and develop" (P2, 2019, Personal Communication, May 6th). Similarly, P5 explains that open source as "having open source code is openly available to critique, build upon without worry of license" (P5, 2019, Personal Communication, May 9th). Also two of the descriptions emphasizes the software side and moral side of open source as "using any kind of source, code, software, content etc. with the rights of re-distribute, re-modify to create value within the legal frameworks of open licenses (P9, 2019, Personal Communication, May 9th), and

> any computer programs whose source code is available to see, which does not have to mean that the code is not copy-righted, or anyone can use it. The code is available to use it. (P6, 2019, Personal Communication, May 10th)

Since open data is one of the inspirations of open source movement, two of the participants approach the idea of open source as an "open data that is freely available and accessible for everyone without any permit" (P8, 2019, Personal Communication, May 10th). Likewise, P7 views the open source particularly as an open data and sees the notion of open source's fundamental as "sharing is caring and it is a mindset rather than a techy term" (P7, 2019, Personal Communication, May 7th). The other participants also acknowledge the significance of open source values within the context of data journalistic practices and praise the principles of transparency, openness, sharing, collaboration (P4, 2019, Personal Communication, May 7th; P1, 2019, Personal Communication, May 6th; P3, 2019, Personal Communication, May 8th ; P10, 2019, Personal Communication, May 15th). While the four out of ten definitions

explain the open source with close descriptions in the literature (Wynants and Cornelis, 2005; Kelty, 2008; Raymond, 2018; Open Source Initiative, n.d.), the others relates the open source with some of its principles and values which predominantly imply the "philosophy of sharing in open source" (Lewis and Usher, 2013; Coleman, 2013).

Second of all, the findings indicate that the participants have common perspectives regarding the definitions of data journalism. It demonstrates the similarity with the main definitions that were pointed by scholars and other practitioners (Bradshaw 2012; Rogers, 2011; Stray, 2011; Howard, 2014; P6, 2015). The common grounds between the participants' answer are "working with structured data, finding stories by gathering, cleaning, analyzing data and finding insights and answers from it and combining journalistic and data skills, etc.". It can also be said that data journalism is described through "process" dimensions as Royal and Blasingame (2015) put it. Especially the definitions of P1, P3, P9 and P5 emphasize respectively that "data journalism is not about how to present, it is about how gathering the material, data and structured way of finding something new inside the data" (P1, 2019, Personal Communication, May 6th), that is working with data to produce stories by using skills of verifying, analyzing data with the mindset of asking the right questions and simply combining journalistic and data skills (P3, 2019. Personal Communication, May 8th; P9, 2019, Personal Communication, May 9th) and that

> Data journalism is utilizing data and storytelling and reporting. It would mean extracting values from data and analyzing data in order to tell stories. I think it is another form of journalism, but it requires skilled people who can analyze and interpret larger datasets in order to get meaning for the readers and audiences (P5, 2019, Personal Communication, May 9th)

Along with the highlights of process dimension many definitions focus on finding stories within data by asking the right questions and using structured data or structuring the datasets. P2, P4, P8, P6, and P7 point out similar terms and phrases to describe data journalism as finding stories in data, using data to answer questions for the news reporting, using structured information any kind and using that data to find journalistic stories and using data as a source for journalism. The definitions in the findings are highly aligned with current or previous descriptions that were pointed out by academics and data journalists (Stray, 2011; Bradshaw, 2012; Howard, 2014; P6, 2015; Royal and Blasingame, 2015).

In summary, when it comes to explaining what open source and data journalism are, compared to the data journalism descriptions, it is likely that open source definitions differ from one participant to another. This may be explained with an open source having a broad range of

meanings, philosophies, and inspirations that are reflected in news media and journalist. In my opinion, explaining the data journalism as a process of gathering, verifying, cleaning, analyzing data and presenting information for the journalistic purposes emphasizes similarities with Lewis and Usher (2013) proposition to approaching journalism as a process rather than a product.

Furthermore, the findings regarding both descriptions also imply that awareness of data journalism is higher than open source among the participants. This is not to say that participants have no clue about open source, in contrast, many participants who cannot properly define open source use phrases of transparency, openness, collaboration, sharing, etc. that have close associations with free and open source software and its ideals

6.2. Use of Open Source Tools and Transparency

This section reflects the participants' choice of software (open sourced or closed source – proprietary) in data journalism practices and whether they follow transparent workflow by publishing datasets, methods, and codes behind their journalistic works or not.

As it mentions above chapters software represents a core of the Free and Open Source Movement. Knowing participants' preferences regarding the tools and software they use in data journalism workflow will yield their familiarity and attitude toward open source software. The findings show that six out ten participants have highly advanced skills and they apply these skills on journalism with open source tools, whilst four of them possess fundamental data-savvy skills but they used at least one open source software in their line of work. Of course, there are participants that prefer closed source tools such as Microsoft Excel or Google Spreadsheet to do quick data journalism practices.

The open source tools that are pointed out by participants vary from standalone software with a user interface, a single purpose software (Open Refine, Tabula, Open Office, Libre Office), web-based tools or extensions (data wrapper, flourish) and advanced coding libraries or software(R programming language, Python, QGIS, D3 Visualization Library, SQL, Git, React.js etc.). Aside from these, Microsoft Excel, Google Tables or Docs are the common closed source or propriety tools that are preferred by the participants. More importantly what I gather from the answers of the participants is that data journalists that are equipped with programming or coding skills know which tool is open source and which software is not while some of the journalists who have less competency and interest in coding or programming point out that they do not discriminate tools or never thought about whether the tools they use are

open source or not. Also, P2 states that she has a pragmatic viewpoint when it comes to the open source, so she sees open source and tools as the means to an end, which is particularly for journalistic practices and breaking stories (P2, 2019, Personal Communication, May 6th.). The increasing awareness of open source software among the journalists with a coding and programming background can be understood as a natural tendency since many programming library and software are prone to be open source because they are developed by the hackers, programmers or developers themselves. Similarly viewing the open source software or tools as means for the craft of telling stories with data in the journalism sphere can be explained with the journalism-focused approach to technological innovations (Lewis and Usher, 2013). This approach refers to the perceiving the technological developments as the means to improve news stories but what Lewis and Usher (2013, p.602) propose is to apply technology-focused approach to journalism innovation which is simply applying values and principles of open source community on news reporting to reconsider journalistic innovations. Therefore, what is important here is the awareness of using open source software with the embrace of open source values and principles. Lewis and Usher (2013) argument shed lights to the discussion regarding viewing journalism and data journalism as a product or process. In this situation, it is useful to ask that is data journalism fancy way to do journalism? or does it distinguish from traditional journalism with some of the values, principles and working disciplines it held?

Another importing finding is about how participants practice some of the informal rules such as publishing datasets, methods and code documentation behind their work. This section will yield plenty of insights regarding their compliance with openness and transparency in data journalistic workflow. Majority of the participants have an optimistic approach regarding having a transparent and open mindset when they practice data journalism and they argue that transparency and openness of work would further improve and strengthen the practice. However, according to P1 and P4, having this mindset is good practice for data journalism but not every dataset are not legal or suitable to publish publicly, in fact, some datasets may contain sensitive information or copyrighted by their owners (P1, 2019, Personal Communication, May 6th; P4, 2019, Personal Communication, May 7th). Also, P1 states that at some point maintaining this practice might contradict or clash with newsroom policies, in fact, these open way of practicing data journalism may not be supported by newsrooms to avoid any competitors since many newsrooms is a profit-oriented and commercial organization (P1, 2019, Personal Communication, May 6th). Also, P2 views time as a barrier to these kind of practices and stresses about the their work in SVT "we try to be transparent how we do things and we publish data but we do not do regular basis but it is not about principles but time" (P2, 2019, Personal

Communication, May 6th). Another good point regarding following these frameworks in data journalism is made by P5 (2019, Personal Communication, May 9th):

> there are always excuses when it comes to this and this is not an easy decision. This decision is not easy to do by yourself. Why do people open source their work, because they are confident about their work but you may also concern about what people think.

Despite these constraints, P5 states that they adopt this way of working two years ago and now on they are publishing methods and R codes behind their stories (P5, 2019, Personal Communication, May 9th). Similarly, other participants claim they try to live by these rules, and they are either using GitHub to host their code and datasets or sharing the datasets in a spreadsheet through hyperlinks inside the article (P8, 2019, Personal Communication, May 10th; P3, H, 2019. Personal Communication, May 8th; P7, A, 2019, Personal Communication, May 7th.). Among the participants, P6 has a skeptic view of current practices of newsrooms regarding whether they are following along with these unformal frameworks and states that not many of them abide by these rules (P6, 2019, Personal Communication, May 10th). Lastly, P9 points out that she always "apply transparency and openness when it comes to both practicing and teaching data journalism" and her main motivation of pursuing this way of working is that the shift would not happen in a day but this approach will be useful in the long term (P9, 2019, Personal Communication, May 9th).

As a result, majority of participants seem to be motivated to follow on this transparent and open way of practicing data-driven journalism which is also stated by Howard (2014) as the main component of data journalism. In contrast, the content of datasets, the license of data sources, newsrooms' policies and fear of having competitors, lack of time to working and documenting the data analysis are viewed as a constraint for transparent and open data journalism.

6.3.Open Source Influence on Data Journalism

This section contains findings regarding participants' point of views regarding the role of the open source community in the emergence of data journalism and open source software on data journalism practices. More importantly, it is about what extent open source mindset is a requirement to work with data for news stories.

Firstly, majority of the participants acknowledge that open source movement has a great deal of influence in the emergence of data journalism in terms of its material aspects (software, code) and values, principles that affect data-driven workflow. Some of the participants associate the

open data, an inspiration from open source movement, with the rising trends of quantitative journalism. P2 views open source as an important factor for the advent of data journalism:

> Open source has affected the emergence of data journalism, because of the access to the tools. Also, it made people aware that they can collaborate with code through GitHub. Compare to the past tools and abilities we do with open source tools and codes are easier and more accessible now thanks to open source movement (P2, 2019, Personal Communication, May 6th)

P1 and P3 acknowledge the impact of open source: it is like a frontline, offers new possibilities and solutions for the steps of data journalism (P1, 2019, Personal Communication, May 6th). Especially accessing and using open source software and its principles and values further help for data journalistic efforts (P3, 2019. Personal Communication, May 8th). Similarly, P4 and Velevia who agree with positive effects of open source for the development of data journalism further state that this development may not be flourished in a country that has no community that is tied to open source and data and thus the emergence of data journalism may vary from one place to another. (P4, 2019, Personal Communication, May 7th; P7, A, 2019, Personal Communication, May 7th). Contrary to fully fledge effects of open source in the advent of data journalism, P8 and P5 stresses that open source movement has a partial role and instead open data movement is the main driver for the rising tide of data journalistic work (P5, 2019, Personal Communication, May 9th; P8, 2019, Personal Communication, May 10th). P8 bases his argument the fact that using data for news stories is older than the open source movement (P8, 2019, Personal Communication, May 10). While P5 emphasizes the role of open data, he also claims that the actors or people who are from open source movement and data-driven journalism have similar mindsets.

As can be seen, some of the participants saw open source and its software features and values as an inevitable factor for the advent of data journalism, while others have similar thoughts for open data. When it is compared with the historical background of data journalism, P8's argument is correct. There are early journalistic examples that do not benefit from the open source and open data (Bounegru, 2011).

On the contrary, P6 argues that although the open source and open data further improves data journalism practices, it is possible to do quantitative reporting without them too (P6, 2019, Personal Communication, May 10th). As it already discussed that open source is one the inspiration of open source ideals that applied into data sources from public and private sector data. Hence, P9 and P10 reconcile these arguments that praise the effect of open source and open data separately by stating that when it comes the emergence of data journalism, open data

and open source further facilitate the journalists and news media in terms of accessing the data analysis, visualization tools and large size of data and encouraging transparent, open and collaborative workflow (P9, 2019, Personal Communication, May 9th; P10, M. 2019, Personal Communication, May 15th).

Aside from the cultural effects of open source community or movement which embodies the software aspects and cultural aspects inside, the software and tools that journalists use have immense contributions for data journalism practices. Many of the participants highly praise the using, accessing the open source software without any financial cost and any kind of license limitations that would worry journalists or newsrooms. Also, some of these open source tools such as Tabula, Open Refine, Workbench are developed for the direct needs of data journalism workflow. In fact, these tools may obsolete the necessity of learning code-based tools (R, Python, Javascript) to produce sophisticated data-driven stories. P1 shares her experiences in teaching regarding the quality of tools that can be an influential alternative for the programming or coding skills:

> the quality of the tools has been really increased. They make it more sophisticated. For example, we have a python course, but we are not sure should we have this? Because there are a lot of tools that do not require coding to data journalism (P1, 2019, Personal Communication, May 6).

Another important advantage of open source software or coding libraries is that these tools come with instructive learning materials like tutorials and documentation that are produced by developers, contributor or even users and P8 believes that not only software but also free learning materials (MOOCs) and the open source community's voluntary support further helps out beginners to master the tool and software (P8, 2019, Personal Communication, May 10).

On the other hand, open source software is not a pure solution for all newsrooms or journalists. P2 indicates that despite using open source tools reduce the financial cost, your organization may not adopt these tools in their workflow because they may prioritize the proprietary tools such as Microsoft Office and Adobe Apps because of its reliability and customer support in case of any errors or bugs etc. (P2, 2019, Personal Communication, May 6th).

In contrast to the opinions regarding role open source community or open source software, participants approach toward open source mindset which refers to have a transparent, open, collaborative and participatory workflow and mentality seem to differ. The five out of 10 participants think open source mindset is not a necessity or requirement but a plus, an extra effort to do better data journalism. The main argument of the participants who view the open source mindset as only a plus is that data journalism can be practiced with closed sourced tools

too (P2, 2019, Personal Communication, May 6; P8, R. 2019, Personal Communication, May 10; P1, S. 2019, Personal Communication, May 6; P4, J. 2019, Personal Communication, May 7th; P5, M. 2019, Personal Communication, May 9th). But P5 and P8 also mention the significant contribution of having an open source mindset:

> It is a plus, not a necessity. I think they can go naturally hand in hand. Value these ideas following along with data-driven workflow with time in my opinion (P8, 2019, Personal Communication, May 10)

> I would not call it as a necessity it is a very big plus. As soon as start doing data journalism and learn R and Python programming language, you are already dealing with the open source even if you are not directly interested. What I'm saying is you can do data journalism without open source tools and mindset, but you would be limiting yourself quite a lot (P5, 2019, Personal Communication, May 9)

The others believe open source mindset has a fundamental part in data journalism workflow. P7, P3, and P10 recognize open source mindset as a must to practice data-driven journalism by relating having this mindset to a transparent, open and collaborative way of doing quality journalism(P7, 2019, Personal Communication, May 7th; P3, H, 2019. Personal Communication, May 8; P10, M. 2019, Personal Communication, May 15th). Similarly, Dağ views the open source mindset, the community and its ideals as "the center of data journalism practices". She further explains that

> Data journalism practices must be open, transparent and accountable to create public value. Having open source mindset would be useful for readers' trust issues toward media and quality of data journalistic work, because this act shows the how reliable, accountable and open is the work and, but also it may encourage other newsrooms to do data journalism and at the end it could help out media ecosystem to thrive (P9, 2019, Personal Communication, May 9).

As a result, it can be evidently said the majority of participants accepts the unprecedented contributions of the open source community and technical boost and advantages open source software and tools provide. Contrary to these findings when it comes to the open source mindset nearly half of participants said that they do not see this as a necessity, because data journalism can be practiced without publishing datasets, methods and codes behind the works. In fact, some participants already mention sharing datasets and codes for open and transparent data journalism may not be aligned with some of the journalistic norms and rules. For this reason, this ambivalence toward open source mindset can be explained with fundamental tensions or clashes between open source logic and professional logic as Lewis and Usher (2013) argues.

As the new players (developers, programmers, hackers) come to journalism sphere so this tension is blurring but what are these clashes between two fields? What would be the further implications? These issues will be discussed further in the next chapter.

6.4. Tensions Between Data Journalism and Open Source

As can be said that data journalism is just a different way of doing journalism and Coddington (2015) identifies data journalism as one of the quantitative reporting practices that combine both open source culture and professional journalistic norms and values and he stresses that potential tensions or clashes in data journalism practices are inevitable. Before delving into the details regarding clashes between two sets of norms and values, it is useful to identify the similar values and norms open source and data journalism hold.

When it comes to the similarity of values between open source and data journalism, four participants identified three common values: transparency, collaboration, and participation. Sandra pointed out that "collaboration part has been adapted more and collaboration have a big role in data journalism even bigger than transparency" (P1, 2019, Personal Communication, May 6). Similarly, P2 indicates that "data journalism embraces transparency and collaboration but not all the time the participation" (P2, 2019, Personal Communication, May 6). The two participants also relate the value of participation in open source with the crowdsourcing practices in journalism. While P8 and P3 also agree with P1 and P2's opinions regarding similar values these two communities possess, P8 emphasizes the increasing number of collaborations between small-scale newsrooms and freelancers (P8, 2019, Personal Communication, May 10; P3, H, 2019. Personal Communication, May 8).

Furthermore, some of the participants state that data journalism has similar principles with open source community due to the open source influence (P4, 2019, Personal Communication, May 7th; P7, A, 2019, Personal Communication, May 7th; P9, P. 2019, Personal Communication, May 9). P4 explains the similarity of values and principles:

> Data journalism community has brought in lots of open source values in journalism. Largely people that come in data journalism are influenced by developer, community. Or actual developer, programmer come in data journalism and be part of the editorial process. And those people have the commitment to open source community and brought some of these values into the newsrooms (P4, 2019, Personal Communication, May 7th)

Similarly, Velevia claims that the mentioned values are purely impacted by the open source and open data community and before the influence of these communities, journalistic practices do not include or apply these kinds of principles (P7, 2019, Personal Communication, May 7th). P5's clarification regarding the resembling principles between data-driven journalism and open source community is about the features of actors or practitioner in these two community. According to him,

> open source and data journalism attract the same people that are already curious about data. The idea that you could build with ubuntu and extent and do a lot more with open source software. It provides more opportunities and it is enticing and encourage creativity and curiosity. Journalists should be curious and try to find out how things work. I think this kind of people have a kind of similar features (P5, 2019, Personal Communication, May 9).

In short, P5 indicates that similarity between two fields is derived from the need these two practices require and same mindset practitioners of these two fields have. But he also thinks that "(data journalism and open source community) one does not need or is not dependent fully other but there is definitely a very close relationship between two" (P5, 2019, Personal Communication, May 9). On the contrary to some of the participants' opinions regarding data journalism embracing transparency due the open source community's influence P5 claims that journalism always promotes the value of transparency and but making data open and available for scrutiny come from the community (P5, 2019, Personal Communication, May 9). Thus, he highlights that open source community amplified the existing principles of being transparent inside the journalism with the data journalism.

In the light of the opinions of participants, it can be clearly illustrated that data journalism and open source have common principles even if the application of these principles and values varies. In the context of data journalism, similar values and principles between data journalism and open source community keep can be explained by the influence of open source given the opinions of participants and resembling working practices of open source and data journalism. What about possible tensions and clashes between data-driven journalism and open source practices? As Coddington argues data journalism is a kind of mixture of professional journalistic values so it carries the features from open source and data journalism community. Also, Lewis and Usher (2013) argue that there are potential conflicts, contradictions in the application of open source culture in journalism although the adoption of open source values through the employment of programmers, developers, hackers who hold open source values and principles dear starts to impact and blur the epistemological differences in the news making gradually.

In addition, when it comes to the possible contradictions and clashes between journalistic values and open source principles, the findings suggest that some of the participants acknowledge the existence of potential conflicts whereas others view this contradiction as a minor issue of inconsistency or never accept there is tension between these parties in the first place. The main arguments of participants who see clashes between open source and professional journalistic mindset is the closed source culture of journalism and journalistic autonomy. P2 explains the possible tensions and clashes:

> we want to share but we want to do things in our terms. That has something to do with the nature of our works. We are selecting what we publish. So, in that matter, it clashes we don't want to invite everybody to do our story. Sometimes we use crowdsourcing and audience. But it is always in our terms. I think this is the difference that is our job, our job is the gatekeeper and what is the story. If we do this selection wrong, the public will punish us. I think this is the responsibility that we don't want to share (P2, 2019, Personal Communication, May 6th).

As can be seen from P2's answer, journalists want to share details regarding their work, but it should be within the conditions and terms they find suitable. More importantly, these journalistic norm or values are again succinctly explicated by her words of "…that has something to do with nature of our works" and "…this is the responsibility that we don't want to share" (P2, 2019, Personal Communication, May 6th). Therefore, journalists don't want to give up this responsibility of truth-seeking, telling stories by being a gatekeeper of the society. This can be related with the ideology of journalism profession and journalists' attributing principles and values to their profession in order to retain prestige, authority and journalistic independence and to protect themselves and their work from the critical view of others as Deuze (2005) put it. Similarly, P4, P3, and P1 emphasize the nature of journalistic practices regarding ethics of sharing data and sources that may implicate other's lives and clashes with the business mindset of journalistic work which prevents the sharing their successful recipe to avoid potential competitions in the journalistic ecosystem. So P4 says

> journalistic values in line with open source values but clashes are with the practice of journalism is more about the culture of the closed source culture. Let's say we gather some data and publish the story on it. We should also publish source data and, but your counter argument would what if someone else does the story. It is like a competition with other news media, so it contradicts with open source ideals (P4, 2019, Personal Communication, May 7th)

Along with the competition argument, P4 also indicates that being wrong would be embarrassing for the journalists if someone finds out mistakes in the journalistic work. P8's point of view clarifies the matter of credibility and prestige successfully:

> open source and journalistic values can complete each other in a certain degree, but the main difference is that journalists cannot be wrong. Because your main selling point is credibility. If you are wrong, you lose value. This might not be a huge failure in the open source community. Because it is not the same as dealing with a bug in your code (P8, 2019, Personal Communication, May 10)

Hence, P4 and P8 opinions illustrate that applying open source principles to software or code and journalism is not the same thing, in fact, this mindset can create some contradictions and conflicts with the very foundation values of news reporting. if the work requires these principles to be applied to it should be in the terms of journalism and journalists as P2 argues.

On the other hand, Velevia and P10 disagree with any kind of clashes and tensions between journalism and open source in the context of data journalism. In the first place Velevia approves of journalists approach to protect their exclusive stories and scoops from other competitors by covering or breaking the story first and calls this a "fundamental bug" but she thinks that "journalism would be socially responsible if it follows along with publishing the source data and disclosing the methodologies, codes behind the news work" (P7, 2019, Personal Communication, May 7th). Likewise, P10 firmly advocates adapting these values by indicating that publishing data, methods and codes behind the stories would encourage the other newsrooms and journalists to do data journalism and similar practices. This would contribute not only to create a market and positive influence for newsrooms but also to increase the journalistic standards (P10, M. 2019, Personal Communication, May 15th). So contrary to open source conflict with business or profit aspects of journalism as other participants claim, P10's takes on the matter suggests that open source way of data journalism might benefit the newsrooms and journalists.

The other faction of participants who are against the potential clashes and conflicts between open source and journalistic values is the participants who believe these are just minor issues and inconsistencies. P9 agree that there are contradictions between the professional norms and open source principles, but it is gradually disappearing with media landscapes get digitalized and the technological development redefines what journalism is and what readers demand from the newsrooms (P9, 2019, Personal Communication, May 9). According to the P9, new technologies form a new kind of audiences that expects transparency, accountability, being inclusive and participatory from newsrooms and journalists even though the traditional value

of journalism contradict with some of these. Therefore, this is only a minor issue (P9, 2019, Personal Communication, May 9). P5 mainly compares the data journalism and open source workflow and criticizes the current application of open source logic in data journalism because he did not often come across data journalism projects that were built upon the similar journalistic projects.

> We got the same general goals. We got the same general idea that if you make something available for other people utilizing it. In software terms, a better product is better code & software. In the journalistic context better context is a better story. Both try to improve through transparency by making it available. However, benefits in the open source ecosystem we saw more tangible returns. Open source community understands the open source ideas and continues and sustains it. But in the data journalism community, we don't really build upon other people's work (P5, 2019, Personal Communication, May 9).

Consequently, the findings suggest that data journalism and open source have common principles such as transparency, collaboration, participation even if the application of these principles and values varies. However, there are also contradictions and tensions between these two fields which are clearly pointed out by the 5 participants while other 4 saw these tensions either as an opportunity that newsrooms should adopt or as a minor issue to overcome. Main tensions are similar to the constraints in front of the transparent way of doing data journalism: journalistic autonomy (independence), ethical and legal issues with sharing datasets, business mindset of the newsrooms, fear of having competition, lack of confidence in sharing the work or the hesitation of losing credibility and prestige incase of failures or errors. On the contrary to these views two participants believe applying open source principles on data journalism makes it more responsible, ethical practice. This further encourage other newsrooms and journalists to invest in this field. Thus, it could boost the journalistic standard and be a positive influence. The other reluctant journalists accept these tensions but consider as minor issues because with the digitalization some of these tensions disappearing and are redefined by the audiences' or readers' novel needs from journalism.

This and previous sections of the findings & discussion chapter try to come up with answers for the first research question that focus on the implications of open source community on data journalism through material existence and philosophical or moral principles of open source community. The next section mainly is going to concentrates to what extent open source community can be a solution for the integration issues and potential challenges of data journalism with its binary structure.

6.5 Open Sourcing Data Journalism Against Its Challenges

This section is consisting of two components. One is about the role of open source principles in the integration of data-driven journalism. Other questions the potential contributions and practical benefits of open source software and coding libraries to overcome constraints of data journalism, which are already discussed in the challenges of data journalism section above. The findings suggest that the majority of the participants (8 out of 10) has optimistic opinions regarding open source solutions for integration problems of data journalism into newsrooms, but there are some minor differences between 8 participants' views when it comes to adapting data journalism workflow into journalism and trying to solve constraints newsrooms facing. First of all, the two participants argue that open source comes with not only advantages but also drawbacks and downsides which proprietary solutions may overcome. When it comes to adopting open source values to integrate data journalism workflow, P6 finds the utilizing open source principles as "hard to apply a single framework into the whole news ecosystem" and

> it is hard to think media as a whole because the media landscape is very different and non-homogeneous because it is hard to apply a single framework on news ecosystem. It depends on why the newsroom exist in the first place (P6, 2019, Personal Communication, May 10th)

Also, he agrees that this approach may work in some conditions, but it depends on the goals of newsrooms and new agencies because "why newsrooms that funded by oligarch wanted to be transparent. It is the same with newsrooms that sell advertising. They have zero incentive to be transparent" (P6, 2019, Personal Communication, May 10th). P4 also shares the same views regarding the role of open source values in integrating data journalism into traditional newsrooms. They also claim that open source software has certain benefits compared to the proprietary alternatives but this not to say that open source software solutions are flawless (P6, N. 2019, Personal Communication, May 10th; P4, J. 2019, Personal Communication, May 7th). According to P4,

> open source is not an obvious answer to solve the challenges of data journalism newsrooms facing. Coding libraries like Python help to work with different data formats without paying anything but this is about building your workspace in coding. Also, non-open source tools such as Excel, Adobe generally are better and user-friendly (P4, 2019, Personal Communication, May 7th).

Similarly, on the one hand P6 acknowledges the practical benefits of open source software by saying that "open source software is free to use and that helps tremendously just because you don't need to pay" P6, N. 2019, Personal Communication, May 10th). On the other hand, this does not justify that open source software is better than its proprietary counterparts all the time, in fact in order to overcome some of data journalistic challenges using non-open source could be more useful (P6, 2019, Personal Communication, May 10th).

Second of all, unlike the opinions of P6 and P4 other participants have affirmative views in terms of values and software of open source community. P1, P3, and Velevia highlight the importance of changing newsrooms' mindset to integrate the data journalism in their work (P7, 2019, Personal Communication, May 7th; P3, H, 2019. Personal Communication, May 8; P1, S. 2019, Personal Communication, May 6). P1 with specialization in Investigative Reporting also points out that this needs a good structure that journalists can be collaborative, transparent and feel secure without undermining journalistic norms (P1, 2019, Personal Communication, May 6). These three participants also believe that open source software plays productive role in overcoming relevant challenges of data journalism by bringing down financial barrier to access new tools, working with different data formats, expanding journalists toolsets with flexible features compare to the proprietary paid tools and replicating, converting, automating the data (P7, 2019, Personal Communication, May 7th; P3, 2019. Personal Communication, May 8; P1, 2019, Personal Communication, May 6). Likewise, P2 accepts the practical benefits of open source software: accessibility to a variety of tools free and helping people who are not data-savvy with publishing data, sharing methods (P2, 2019, Personal Communication, May 6th). However, when it comes to the embracing open source values to facilitate quantitative shifts in newsrooms, she praises the practical or material aspects of open source community:

> Using open source software is the means to an end. We use things that do the job best. But these do not mean I embrace the principles and values of open source. We are journalist, we break stories and for my open source is a more of pragmatic things that may be beneficial for my work (P2, 2019, Personal Communication, May 6th)

This approach might be a significant issue in this context. Do journalists care about open source ideals while they use open source software? Do they use these tools for supporting existing practices of traditional journalism and professional journalistic norms? Again, Lewis and Usher (2013) view this kind of practices and approach as a barrier to use open source in its full potentials.

In addition, P8, P10, and Dağ find useful to have open source values and principles for the integration of data journalism practices as P1, P3, and Velevia indicate, but the former trio stresses that the change in the news ecosystem would take time and the effects of open source mindset might not be the entire solutions for the integration issues (P8, 2019, Personal Communication, May 10; P9, 2019, Personal Communication, May 9; P10, 2019, Personal Communication, May 15th). It seems the problem has more dimensions and layers to address. When it comes to contributions of the open source software for the constraints of working with data, Along with P5, P8, P10, and Dağ acknowledge that using open source software would help out to solve some of these problems. First and foremost open source software is free and accessible for anyone without any license fee or financial cost, secondly it offers more flexible solutions and features to work with data compare the non-open source alternatives and thirdly open source coding libraries improve doing data journalism from scraping, cleaning, analyzing and visualization data in the same workspace without paying anything or using any paid software and this even can save time compared to using different tools for the each step. (P5, M. 2019, Personal Communication, May 9; P8, R. 2019, Personal Communication, May 10; P9, P. 2019, Personal Communication, May 9; P10, M. 2019, Personal Communication, May 15th). Of course, using code-based tools requires some efforts to master and utilize in journalistic work. P8 agrees with this but also views open source community as a great opportunity because the community produces a variety of learning materials for newcomers to use these code languages and tools (P8, 2019, Personal Communication, May 10).

7. SUMMARY & CONLUSION

In order to answer the two main research questions that concern with influence and productive role of open source community in data journalism, 10 data journalists or data journalism lecturers are interviewed. The findings were analyzed and discussed in 5 sections. First 4 sections are about the influence of open source community on data journalism and contradictions between open source ideals and journalistic norms. The last one focuses on to what extent open source can be a solution for the integration and challenges of data journalism with its moral and material aspects.

Firstly, findings demonstrate that when it comes to the explaining what open source and data journalism is, compared to the data journalism descriptions, it is likely that open source definitions differ from one participant to another. This may be explained with an open source having a broad range of meanings, philosophies, and inspirations that are reflected in news media and journalists. In my opinion, explaining the data journalism as a process of gathering, verifying, cleaning, analyzing data and presenting information for the journalistic purposes emphasizes similarities with Lewis and Usher (2013) proposition to approaching journalism as a process rather than a product. More importantly, the related findings regarding both descriptions also imply that awareness of data journalism is higher than open source among the participants. This is not to say that participants have no clue about open source, in contrast, many participants that cannot exactly define open source use phrases and words such as transparency, openness, collaboration, sharing, etc. that have close associations with free and open source software community and its ideals.

Secondly, although open source tools which participants use vary from web-based tools, standalone programs to coding libraries, what I gather from the answers of the participants is that data journalists that are equipped with programming or coding skills know more about which tool is open source and which software is not. The increasing awareness of open source software among the journalists with a coding and programming background can be understood as a natural tendency since many programming library and software are prone to be open source because they are developed by these hackers, programmer or developers themselves. Similarly viewing the open source software or tools as means for the craft of telling stories with data in the journalism sphere can be explained with the journalism-focused approach to technological

innovations which is mainly focuses on the utility of technological tools rather than it is open source or not (Lewis and Usher, 2013). Another important point is that majority of participants seem to be motivated to follow along with transparent and open way practicing data-driven journalism which is also stated by Howard (2014) as one of the main components of data journalism. In contrast, the content of datasets, the license of data sources, newsrooms' policies and fear of having competitors, lack of time to working and documenting the data analysis are viewed as a constraint for transparent and open data journalism.

Thirdly, many of the participants acknowledge that open source movement has a great deal of influence in the emergence of data journalism in terms of its material aspects (software, code) and values, principles that affect data-driven workflow. Some of the participants associate the open data, an inspiration from open source movement, with the rising trends of quantitative journalism. Contrary to these findings when it comes to the open source mindset nearly half of participants said that they do not see this as a necessity but only a plus, because data journalism can be practiced without publishing datasets, methods and codes behind the works. In fact, some participants already mention sharing datasets and codes for open and transparent data journalism may not be aligned with some of the journalistic norms and rules. For this reason, this ambivalence toward open source mindset can be explained with fundamental tensions or clashes between open source logic and professional logic as Lewis and Usher (2013) argues. When it comes to the possible contradictions and clashes between journalistic values and open source principles, the findings suggest that the five participants acknowledge the existence of potential conflicts whereas others view this contradiction as a minor issue of inconsistency or never accept there is tension between these parties in the first place. The main arguments of participants who are sure of clashes between open source and professional journalistic mindset are the closed source culture of journalism, journalistic autonomy, journalistic credibility, journalism's business mindset, ethical and legal issues in publishing datasets. The participants who view these contradictions as minor issues claim that these clashes are gradually disappearing with media landscapes getting digitalized and the technological development redefining what journalism is and what readers demand from the newsrooms because new technologies form a new kind of audiences that expects transparency, accountability, being inclusive and participatory from newsrooms and journalists even though traditional value of journalism contradict with some of these. Contrary to participants who stresses the existing contradictions and tensions, two participants do not view any controversy between open source values and data journalism instead they highlight those data journalism activities that follow along with open source principles are socially responsible and this way of practice would

encourage other newsrooms and journalists and create demands to do data journalism along with further improving the journalistic standards.

Lastly, the second research question aims to find out to what extent open source can be a solution for the constraints of data journalism with its tools (software & coding libraries) and values. Despite most of the participants agree on existing conflicts between journalistic norms and open source values, 8 out of 10 participants view the open source tools and open sourcing data journalism as a solution for the integration of data journalism. The two participants admit the advantages of open source, but they also argue that open source comes with some drawbacks and downsides. Their main argument is that a single framework of open sourcing data journalism may not work for whole news ecosystem and this differs from one newsroom to another because their editorial policies and motivations: why newsrooms who funded by oligarchs or news portal aims to sell more advertising want to be transparent (P6, 2019, Personal Communication, May 10th). Another thing is that open source software solutions are not flawless and sometimes non-open sourced tools can be more user-friendly or practical and easy to use for data journalism. Unlike the two participants, others have a more optimistic view regarding the integration issues even if they think it would take time to ease the current integration issues and to change the mindset in the newsroom and top management. Also, they praise the practical benefits of open source software to overcome some of the challenges of data journalism. According to majority of participants, open source contributes to bringing down financial barrier to access new tools without paying any cost for a license, working with different data formats, expanding journalists toolsets with flexible features compare to the proprietary paid tools and offering variety of functions scraping, cleaning, analyzing visualizing and automating the data through coding libraries and tools.

All in all, this study examined the influence and role of open source community in data journalism by drawing from the work of Lewis and Usher (2013) for the first time. In the light of qualitative insights from the semi-structured interviews with 10 data journalism professionals, it can be said that data journalism practices are influenced by the material and moral or philosophical structure of open source community. While some of the open source values such as transparency, collaboration etc. started to be adopted by data journalists and functionality of open source software are highly praised, there are certain tensions and contradictions between open source mindset and data journalism due to the connection of data journalism with the professional journalistic values (Lewis and Usher, 2013; Coddington, 2015). Journalistic autonomy, ethical and legal concerns, business mindset of newsrooms along with pragmatic perspective of journalists, fear of having competition, journalism's credibility

and prestige issues and insecure structure of open source software compared to the non-open source alternatives' lasting premium customer support were viewed as main concerns and conflicts by a few participant. On the other hand, some of the participants either approached these conflicts as minor issues that will be fixed with open sourcing data journalism or not even mentioned them. Another important finding is that majority of the participants admitted open source principles and open source software are beneficial to integrating a newsroom into data journalism workflow if the newsrooms and top management change their mindset regarding these practices and overcoming some of the common challenges with practicing data journalism: lack of financial resources, non-open data formats, lack of quality and flexible visualization and analysis tools and lack of time to crunch data.

Consequently, these findings support that Lewis and Usher's (2013) preposition to reconsider journalistic innovation can also be applied into data journalism to integrate data-driven workflow and solve not all but some of the challenges which newsrooms facing, despite there are tensions and contradictions between open source principles and professional journalistic values. Of course, this is not to say that this is flawless recipe to facilitate the smooth transition instantly or show snap effects to solve related problems. As much as this study attempted to bring together data journalists and data journalism instructor from different countries and journalism traditions, in order to expand these findings further and obtain richer outputs regarding the integration period and common challenges, follow up scholarly works that include more journalists and newsrooms from diverse and different medium of news media with the combination of quantitative and qualitative approaches will be needed.

8. SOURCES

A Brief History of the Open-Source Movement (2000). Available at:
https://web.archive.org/web/20110411022933/http://sloanreview.mit.edu/the-magazine/2000-fall/4211-3/a-brief-history-of-the-opensource-movement/# (Accessed: 20 May 2019).

Appelgren, E. and Nygren, G. (2014) 'Data Journalism in Sweden', *Digital Journalism*, 2(3), pp. 394–405. doi: 10.1080/21670811.2014.884344.

Appelgren, E. and Salaverría, R. (2018) 'The Promise of the Transparency Culture', *Journalism Practice*. doi: 10.1080/17512786.2018.1511823.

Baack, S. (2015) 'Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism', *Big Data & Society*. doi: 10.1177/2053951715594634.

Bode, B. A. (2013) *Open Data: A History*. Available at: https://www.data.gov/blog/open-data-history (Accessed: 20 May 2019).

Bounegru, L. (2012) 'Data Journalism in Perspective', in Gray, J., Bounegru, L., and Chambers, L. (eds) *The data journalism handbook: How journalists can use data to improve the news*. O'Reilly Media, Inc, pp. 17–21.

Bradshaw, P. (2012) 'What is Data Journalism?', in Gray, J., Bounegru, Liliana, and Chambers, L. (eds) *The data journalism handbook: How journalists can use data to improve the news*. O'Reilly Media, Inc, pp. 2–3.

Brennen, B. S. (2012) *Qualitative Research Methods for Media Studies : An Introduction*. London, UNITED KINGDOM: Taylor & Francis Group. Available at: http://ebookcentral.proquest.com/lib/gu/detail.action?docID=1075433.

Coddington, M. (2015) 'Clarifying Journalism's Quantitative Turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting', *Digital Journalism*. Routledge, 3(3), pp. 331–348. doi: 10.1080/21670811.2014.976400.

Coleman, E. G. (2013) *Coding freedom: The ethics and aesthetics of hacking*. Princeton University Press.

Demirel, S. (2018) *Use of data visualization in news reporting in Turkey*. Kadir Has University. Available at: https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp.

Deuze, M. (2005) 'What is journalism? Professional identity and ideology of journalists reconsidered', *Journalism*. doi: 10.1177/1464884905056815.

Figl, B. (2017) *Bigger is not always better: What we can learn about data journalism from small newsrooms*, *Reuters Institute*. Available at: http://reutersinstitute.politics.ox.ac.uk/our-research/bigger-not-always-better-what-we-can-learn-about-data-journalism-small-newsrooms (Accessed: 20 May 2019).

Fink, K. and Anderson, C. W. (2015) 'Data Journalism in the United States: Beyond the "usual suspects"', *Journalism Studies*, 16(4), pp. 467–481. doi: 10.1080/1461670X.2014.939852.

Flick, U., von Kardoff, E. and Steinke, I. (2004) *A Companion to Qualitative Research : Paradigms, Theories, Methods, Practice and Contexts*. London, UNITED KINGDOM: SAGE Publications. Available at: http://ebookcentral.proquest.com/lib/gu/detail.action?docID=354934.

Gray, J., Chambers, L. and Bounegru, L. (2012) *The data journalism handbook: How journalists can use data to improve the news*. ' O'Reilly Media, Inc.'

Heravi, B. (2017) *State of Data Journalism Globally*, *Web Blog Post*. Available at: https://medium.com/@Bahareh/state-of-data-journalism-globally-cb2f4696ad3d

(Accessed: 20 May 2019).

Howard, A. (2014) 'The Art and Science of Data-Driven Journalism', *Tow Centre for Digital Journalism - a Tow/Knight Report*. doi: 10.1017/CBO9781107415324.004.

Jenkins, H. (2006) *Convergence culture: where old and new media collide*, *New York: New York UP*. New York and London: New York University Press.

Karlsson, M. (2010) 'RITUALS OF TRANSPARENCY', *Journalism Studies*. doi: 10.1080/14616701003638400.

Kayser-Bril, N. (2015) *Data Journalism*. Available at: http://blog.nkb.fr/datajournalism (Accessed: 20 May 2019).

Kayser-Bril, N. (2017) *Collaboration in Journalism*. Available at: https://blog.nkb.fr/collaboration (Accessed: 20 May 2019).

Kayser-Bril, N., Valeeva, A. and Radchenko, I. (2016) 'Transformation of Communication Processes: Data Journalism', pp. 414–421. doi: 10.5281/zenodo.51091.

Kelty, C. M. (2008) *Two bits: The cultural significance of free software*. Duke University Press.

Kimpton, P. (2013) *Obama to Berners-Lee, Snow to Domesday: a history of open data*, *Theguardian.com*. Available at: https://www.theguardian.com/news/datablog/2013/oct/25/barack-obama-tim-berners-lee-open-data (Accessed: 20 May 2019).

Lewis, S. C. (2012) 'THE TENSION BETWEEN PROFESSIONAL CONTROL AND OPEN PARTICIPATION', *Information, Communication & Society*. doi: 10.1080/1369118X.2012.674150.

Lewis, S. C. and Usher, N. (2013) 'Open source and journalism: Toward new frameworks for imagining news innovation', *Media, Culture and Society*. doi: 10.1177/0163443713485494.

Lewis, S. C. and Westlund, O. (2015) 'Big Data and Journalism', *Digital Journalism*, 3(3), pp. 447–466. doi: 10.1080/21670811.2014.976418.

De Maeyer, J. *et al.* (2015) 'Waiting for Data Journalism', *Digital Journalism*, 3(3), pp. 432–446. doi: 10.1080/21670811.2014.976415.

Mazotte, N. (2017) *How the Argentinian daily La Nación became a data journalism powerhouse in Latin America*, *www.niemanlab.org*. Available at: http://www.niemanlab.org/2017/04/how-the-argentinian-daily-la-nacion-became-a-data-journalism-powerhouse-in-latin-america/ (Accessed: 14 November 2017).

Parasie, S. and P9iral, E. (2013) 'Data-driven journalism and the public good: "Computer-assisted-reporters" and "programmer-journalists" in Chicago', *New Media and Society*, 15(6), pp. 853–871. doi: 10.1177/1461444812463345.

Raymond, E. (2017) *Open Source | Social Movement*, *Encyclopædia Britannica, inc.* Available at: https://www.britannica.com/topic/open-source (Accessed: 20 May 2019).

Rogers, S. (2011) *Data journalism at the Guardian: what is it and how do we do it?*, *www.theguardian.com*. Available at: https://www.theguardian.com/news/datablog/2011/jul/28/data-journalism (Accessed: 14 November 2017).

Royal, C. and Blasingame, D. (2015) 'Data Journalism: An Explication', *#ISOJ*.

Silverman, D. (2006) *Interpreting Qualitative Data : Methods for Analyzing Talk, Text and Interaction*. London, UNITED KINGDOM: SAGE Publications. Available at: http://ebookcentral.proquest.com/lib/gu/detail.action?docID=585423.

Singer, J. B. (2007) 'CONTESTED AUTONOMY', *Journalism Studies*. doi: 10.1080/14616700601056866.

Smit, G., de Haan, Y. and Buijs, L. (2014) 'Visualizing News: Make it work', *Digital Journalism*, 2(3), pp. 344–354. doi: 10.1080/21670811.2014.897847.

Splendore, S. *et al.* (2016) 'Educational strategies in data journalism: A comparative study of six European countries', *Journalism: Theory, Practice & Criticism*, 17(1), pp. 138–152. doi: 10.1177/1464884915612683.

Stallman, R. (2016) *FLOSS and FOSS*, *GNU Operating System*. Available at: https://www.gnu.org/philosophy/floss-and-foss.html (Accessed: 20 May 2019).

Stallman, R. (2019) *Why Open Source misses the point of Free Software*, *GNU Operating System*. Available at: https://www.gnu.org/philosophy/open-source-misses-the-point.html (Accessed: 20 June 2019).

Stray, J. (2011) *A computational journalism reading list*, *Web Blog Post*. Available at: http://jonathanstray.com/a-computational-journalism-reading-list (Accessed: 20 May 2019).

Tracy, S. J. (2013) *Qualitative Research Methods : Collecting Evidence, Crafting Analysis, Communicating Impact*. Chicester, UNITED KINGDOM: John Wiley & Sons, Incorporated. Available at: http://ebookcentral.proquest.com/lib/gu/detail.action?docID=1120701.

What is Open Access? (no date) *opensource.com*. Available at: https://opensource.com/resources/what-open-access (Accessed: 20 May 2019).

What is Open Data? (no date) *Opendatahandbook.org*. Available at: http://opendatahandbook.org/guide/en/what-is-open-data/ (Accessed: 20 May 2019).

What is Open Source software? (no date) *opensource.org*.

Wynants, M. and Cornelis, J. (2005) *How Open is the Future?: Economic, Social & Cultural Scenarios Inspired by Free & Open-source Software*, *Citeseer*.

Young, M. L., Hermida, A. and Fulda, J. (2017) 'What Makes for Great Data Journalism?', *Journalism Practice*, 0(0), pp. 1–21. doi: 10.1080/17512786.2016.1270171.

9. APPENDIXES

1. Interview Guidelines

Main RQ 1: *How open source movement influence data journalism with its material existence (software, coding libraries) and cultural values and principles?*

- How do you define open source?
- How do you define data journalism?
- Do you think that open source and data journalism have similar values or norms?
- Does data journalism workflow require having open source mindset?
- Do you comply with some of the unformal norms such as publishing datasets, methodologies, and codes behind the data journalistic work?
- To what extent data journalism embraced open source values (in the terms of participation, collaboration, transparency)
- What do you think about the relationship between open source values (participation, transparency, experimentation, iteration (free to fail and focusing the process rather than product) and journalistic norms (autonomy, prestige, gatekeeper role and control over contents)? (Do these undermine or complete each other?)
- Can data journalism workflow be open source? (by making more transparent, participatory and collaborative)
- Do you know any open sourced software (programs, coding language) for data journalism? (Please list their names)
- Have you ever used open sourced software (programs, coding language) for data-savvy journalistic work? (Please list their names and explain your experience)
- What do you think about the effect of open source software on data journalism?
- What do you think about role of the open source movement in the emergence of data journalism?

Main RQ 2: Does applying open source mindset (both its materiality and cultural principles) in data journalism facilitate to deal with its challenges of working with data for journalistic purposes.

- Do you think embracing open source values facilitate newsrooms to adopt data journalism? How?

- What do you think about role of open source software (programs and code language) to deal challenges of working with data for journalistic purposes? (Please elaborate)
- What do you think about role of open source values in data journalism regarding its problem-solving aspects?