

The Principles, Features and Techniques of Data Journalism

Olga Kalatzi¹, Charalampos Bratsas², Andreas Veglis³

¹Media Informatics Lab, School of Journalism & Mass Communication, Aristotle University of Thessaloniki, Greece

²Aristotle University of Thessaloniki and Open Knowledge Foundation (OKF)-Greek Chapter, Greece

³Media Informatics Lab, School of Journalism & Mass Communication, Aristotle University of Thessaloniki and Open Knowledge Foundation (OKF) – Greek Chapter, Greece

Correspondence: Charalampos Bratsas, Aristotle University of Thessaloniki and Open Knowledge Foundation (OKF)-Greek Chapter, Greece.

Received: April 13, 2018

Accepted: August 30, 2018

Online Published: September 17, 2018

doi:10.11114/smc.v6i2.3208

URL: <https://doi.org/10.11114/smc.v6i2.3208>

Abstract

Digital and computational technology is steadily developing and continually bringing changes in the field of journalism, which faces a major crisis, as people's trust in the media continues to decrease. This paper studies the subject of data journalism which is increasing in popularity and is considered to be at the forefront of these changes. This kind of journalism may be a way to re-establish and strengthen journalism's value, as well as to reassure its sustainability. Datasets, tools, policies on Freedom of Information and transparency and professionals become constantly available for data journalism to flourish. However, there are still many challenges along with skepticism and confusion around its role and value in the field. In the near future, data journalism seems to gain more trust by the news organization and the public, as both start to comprehend its potentials.

Keywords: data-driven journalism, definition, process, value, data, the open data movement

1. Introduction

Recently, there has been a growing interest in data journalism by several researchers, professional journalists, programmers and graphic designers as well. Studies around data journalism identify its importance, its challenges and its future perspectives, but also there are books and reports with practical information regarding its process (Appelgren & Nygren, 2014) (Fink & Anderson, 2015) (Mair & Keeble, 2014). However, very few researches cover all aspects of the topic.

Data journalism has been developed and spread in a quite fast way the last few years not only by the biggest international news media and news agencies, but also by local news organizations. However, using data to find a story and create visualizations is not actually something new. There are quite a few examples from the past two centuries where big stories came out of data. In 1821 the Guardian, which was called Manchester Guardian back then, published for the first time a table with data about the number of pupils attending school in two cities of England as well as the average annual spending. The data proved to be useful information providing a clear picture of the state of the education system in the country. In 1858, Florence Nightingale with her report Mortality of the British Army gave step to data visualization using charts, diagrams and tables of data. The most famous one is the Coxcomb which showed that most soldiers' deaths were due to treatable diseases than bullets. (Rogers, 2013) (Veglis & Bratsas, 2017) (Felle, Mair & Radcliffe, 2015) (Baack, 2011). But after the turn of the century certain parameters have contributed to data journalism in coming into the center stage: the availability of data, of applications that can manage data, of applications that can visualize data (Veglis & Bratsas, 2017).

The purpose of this paper is to examine all facets of data journalism from a theoretical approach by combining the available bibliography to further explore data journalism's potentials. The rest of the paper is organized in seven main sections. Section 2 focuses on the definition of data journalism and its value to journalism and the public. The relation between journalism and data is explored in section 3. Specifically, there is an analysis on the definition of data and the Open Data Movement, followed by an examination of the influence of data to journalism. The process of practicing data journalism is discussed in section 4. In the next section, the value of data journalism is examined. Section 6 explores the challenges that data journalism is facing. Finally, the future of data journalism is discussed in section 7, and the last section of the paper is the conclusion.

2. Defining Data Journalism

Today the term data journalism still remains confusing even among journalists. Some argue that it cannot be defined as journalism, while others claim it could be the beginning for a new era in journalism (Rogers, 2013). There are different definitions and terms to describe data journalism, for instance data journalism, data-driven journalism, database journalism, computational journalism or data visualization. Simon Rogers, data editor at the *Guardian*, defines data journalism as having “a wide range of styles from visualization to long-form articles. The key thing they have in common is that they are based on numbers and statistics- and that they should aim to get a story from that” (Felle et al. 2015). Lorenz at the event “Data-driven journalism: What is there to learn?” in 2010 describes data journalism “as a workflow, where data is the basis of the analysis, visualization and -most importantly- storytelling”. Additionally, Hamilton and Turner use the term computational journalism to refer to the kind of journalism that incorporates algorithms, data and knowledge from the social sciences in order to make journalism more accountable (Hamilton & Turner, 2009).

Combining all the elements that are included in data journalism there is a final definition adopted by Veglis and Bratsas (2017): Data journalism can be defined as the process of extracting useful information from data, writing articles based on the information and embedding visualizations (interacting in some cases) in the articles that help the readers understand the significance of the story or allow them to pinpoint data that relate to them.

Many claimed that data journalism has its roots in Computer Assisted Reporting (CAR) (Felle et al. 2015). But, actually it is different from CAR, as it covers the whole work done between journalism and computing, while giving respect to CAR as it is grounded on better foundations (Felle et al. 2015) (Gray, Bounegru & Chambers, 2012).

The most important advantage though regarding the increasing rise of data journalism is the exploitation of the power of the internet. Nowadays journalists can find data online and they can use a vast majority of tools and programs to clean them, analyze them and create the right visualizations. Moreover, the spread of the web means that there is room for open journalism. People who read the stories can also make comments and get involved in the story. Data journalism becomes part of this, by exposing the data and encouraging people to search it and find something “new from the news” (Rogers, 2013). Another aspect of open journalism is crowd sourcing, where people are asked to contribute to the process of collecting data along with the journalists, when the latter cannot find the data in any other way (Rogers, 2013).

Some consider data journalism to be only numbers and graphics that requires mathematical skills as well as coding skills, which are out of a journalist’s expertise. Others oppose to the influence of the internet, as they only believe in the traditional journalism and media outlets. However, data journalism is as much a kind of journalism as all the others. It follows the same fundamentals of finding a good story. The only difference is that the story is mainly based on information that comes from data and not the traditional sources, such as contacts, interviews or press releases. This does not mean that there is no reporting as data journalists still need to contact sources and ask the right questions to add more contexts and make the data more understandable. Data journalism is not only about infographics and charts (Felle et al. 2015) (Rogers, 2013).

3. Data and Journalism

3.1 Definition of Data

It is apparent that this is the information age and every day more data becomes available to the public. Especially the past five years big datasets have been opened not only by governments but also by companies and organizations. The shift to digitalization has made possible for all this information to be stored online (Felle et al. 2015) (Krum, 2014). Being exposed to such a great amount of data does not necessarily imply that the term itself is clear and understandable by all the bodies that generate it (governmental departments, public bodies, organizations, big companies, NGOs), nor the public or even the journalists that use it. At first, there are two different approaches on the grammatical use of the word. Some tend to use the word’s plural form following the grammar of Latin (datum= singular, data= plural), while others, usually in a non-scientific context; adopt a more modern approach of the language using the singular form. Both ways are acceptable but there is a shift towards using the singular form as a mass noun in the journalistic world, for instance the *Wall Street Journal*, the *Guardian* and the *National Geographic* (Rogers, 2013).

As for the definition of the word, Alberto Cairo in his book “The functional art” explains the term placing it in the flow of the process from telling the reality to people. Based on that, data is the first level of encoding as a researcher analyses and observes the unstructured information (fact-reality). Then the communicator exploits the data to create an influential story (using text or graphics) for the public by transforming it into structured information (Cairo, 2013). Another approach of defining “data” is the one used by Lesage and Hackett in their research article “Between objectivity and openness-The mediality of data journalism” referencing the definition given by the *Oxford English Dictionary*. The dictionary offers two aspects of the term: 1) Related items of (chiefly numerical) information considered collectively, typically obtained by scientific work and used for reference, analysis, or calculations. 2) Computing. Quantities, characters, or symbols on

which operations are performed by a computer, considered collectively. Also (in non-technical contexts): information in digital form. According to Lesage and Hackett both renditions of data are being used by journalists, as there are many examples of scientific researches being the source of a story, as well as there are stories that are based on computational analysis of a certain data (Lesage et al. 2014).

According to Gavin Freeguard, senior researcher at the Institute for the UK government, there are three ways to perceive data. The first one is data as data, which is the raw data, the numbers published in a spreadsheet or in another format that ensures that the data can be reused. The second one is data as information, which implies turning the raw data into something meaningful, mainly through visualizations that can highlight the important points of a dataset. The last one is data as evidence, meaning that the data can be the proof of the state of performance of those in power and a tool to put some pressure for improvements by the citizens themselves (Felle et al. 2015). Finally, Randy Krum in “Cool Infographics” demystifies the term “data” by emphasizing on some of its kinds: “The data could be huge spreadsheets of numbers, quantitative research data, business processes, demographics, financial results, map locations, or web statistics.”

All four definitions show a distinctive aspect of the term, from the semantic analysis of the word, to its place in the communication process and examples of its types. Thus, combined all together make a complete description of data as it is used in the field of communication and social sciences.

3.2 The Open Data Movement

The vast availability of data in the internet began in the first place with the decision of Barack Obama to launch the site data.gov in 2009 where governmental datasets were published. After USA other countries adopted this initiative, for example Australia, UK and New Zealand. There was also a local interest in developing a site for publishing local administration’s data, such as Paris, London, New York, Vancouver, Catalonia. Although the data available is a new potential source to the journalists’ stories, they are not the target group for the governments (Hamilton et al. 2009) (Stoneman, 2015). In reality, as Simon Rogers points out governments started to publish their data and continue to put an effort on this initiative in order to increase their transparency towards the citizens. To support and enhance this argument, Gavin Freeguard from the UK government, underlines that the reasons for the UK government to open its data were government accountability to citizens, improving public services and catalyzing economic and social growth (Felle et al. 2015).

The open data movement though was not an entirely governmental initiative, instead coders, developers, news organizations, NGOs and data enthusiasts started to demand, promote and create datasets (Rogers, 2013). There are many examples of different projects aiming in opening data and publishing it in an accessible and understandable form on behalf of the public. OpenCorporates.com, for instance, currently offers an open database of more than 110 million companies from all over the world with all the relative information about them that allows anyone to search and process it (Felle et al. 2015). Another representative example, is the non-profit organization Open Knowledge International (OKI) which has a huge impact in the area of open data with many projects through the years. One of its recent projects is OpenBudgets.eu, which will be an open-source software along with a SAAS (Software-As-A-Service) platform with datasets on budgets and public spending. The project aims to strengthen fiscal transparency and create a useful tool not only for journalists, but also for all different stakeholders and the public, providing even visualization tools and comparative analysis tools (Philippides, Karampatakis, Ioannidis, Mynarz, Svátek, and Bratsas, 2016) (Koupidis, Bratsas, Karampatakis, Martzopoulou, and Antoniou, 2016). The chapter of Open Knowledge Greece is also working on the project, as well as in other initiatives of OKI.

3.3 Principles of Open Data and Relationship With Journalism

The whole movement put the term of open data in perspective. There are various definitions of open data, as for instance the one that Simon Rogers offers in his book “Facts are sacred” where he defines open data as free information available to anyone in a machine-readable format, such as CSV, that permits people, and mainly the analysts, to work with it (Rogers, 2013). The common ground of all the definitions of open data is the principles of data’s openness. According to Schellong and Stepanets there are eight principles to identify the state of openness of the data. To begin with, there is the principle of completeness which refers to the availability of data without any privacy restriction. Secondly, it is primacy which means that the data is as granular as possible and there are no alterations. Then follows timeliness, data, in order to keep its value, needs to be published in a short period of time. The next principle, accessibility, considers data to be available to the maximum number of people, while the fifth principle, machine readability, is related to the data’s structure that allows further processing and analysis. The last three principles, absence of discrimination, absence of propriety and absence of license requirements, come in accordance with the first principle, as they determine that data is accessible to all, with no requirement of registration, as well as no one in particular in control of it and it is not under any copyright or other restriction, respectively (Aitamurto, Sirkkunen, and Lehtonen, 2011). Examining closely these principles it comes as a conclusion that, although open data can be an important step towards a more transparent society, it is quite difficult to add value and stand on its own. This is because data cannot create value or change by itself. Regarding the principles

related to availability, while there is a number of datasets available, mainly from official institutions and governmental bodies, it is a very common procedure for journalists to make use of the Freedom of Information Act in order to gain access. Additionally, there have been reported cases in which journalists had to negotiate and file many FOI requests before they were given access to data, primarily in the public sector. Furthermore a great number of datasets published are not complete or without any mistakes. They need to be further processed, analyzed and even accompanied with additional information and context in order to become relative and useful for the public. (Rogers, 2013) An other issue related to timeliness and machine readability is that datasets are not published soon enough for journalists to use them nor are they available in an appropriate format (Borges-Rey, 2017) (De Maeyer et al. 2015) (Stalph, 2017) (Loosen et al. 2017).

Although open data can be an important step towards a more transparent society, it is quite difficult to add value and stand on its own. This is because data cannot create value or change by itself. One potential reason for this could be that a great number of datasets published are not complete or without any mistakes. They need to be further processed, analyzed and even accompanied with additional information and context in order to become relative and useful for the public. (Rogers, 2013, p. 273-274) At this point, it is clear the bidirectional relationship between open data and journalism. On one hand, journalism can fulfill the need of understanding, interpreting the data and demonstrating its importance for the public interest, while on the other hand data opens a new perspective for journalism to find stories that can personally affect people, to explain complicated issues, as for example in the financial or economical sector, and to become more accurate and trustworthy (Rogers, 2013) (Gray et al. 2012).

Despite the fact that both sides can cooperate in promoting transparency and accountability, while serving the public, their relationship is not so far characterized by mutual respect but rather with suspicion. Journalists claim that the data is not published by the governments and the other public bodies on time in order to be newsworthy and also, they question the reliability of the data. At the same time, public bodies do not trust the media because, according them, all stories focus on bad news distorting the facts and not showing the whole picture or the positive results (Felle et al. 2015) (Stoneman, 2015).

Data journalism has started to find its place in the news organizations and newsrooms, but still many journalists and even the audience cannot comprehend its role and value. Based on researches that focus on data journalism in news media it is confirmed that there is a growing interest in encompassing data journalism into newsrooms. However, it is also highlighted that in many cases journalists and editors do not share the same view and understanding regarding the definition of data journalism, the role of data journalist, as well as the contribution and the principles of data journalism. At this point many questions around data journalism arise. Is data journalism closer to practices and norms of traditional journalism when dealing with data or is it a new approach closer to the computers' logic? Can and should journalists obtain computational and graphic skills or they should cooperate with experts in the field? What is considered excellence in data journalism? (Hermida & Young, 2016) (Lewis & Westlund, 2015) (De Maeyer et al. 2015) (Engebretsen et al. 2018) (Young et al. 2018) The stories though that came from the WikiLeaks releases on Afghanistan, Iraq and the US embassy cables brought data journalism into spectrum and highlighted its meaning (Rogers, 2013) (Felle et al. 2015) (Baack, 2011).

4. The Value of Data Journalism

The value of data journalism can be compared in the first place with the value of traditional journalism, as it is defined by Rosen to be the solution of the "problem of scale". The need to be informed of events that take place all over the world in a real time has been satisfied by the journalists, who inform the public. Additionally, the "problem of scale" includes, apart from distance, other issues as for example complicated information and facts concerning politics and economy, that are important for the citizens to know but they are difficult to understand. Once again, the journalist has the role of publishing all this information in a simple way (Lesage & Hackett, 2014).

Data available in the internet may produce similar confusion and disorientation. People are not educated to work with data and comprehend it or at least select what could be important from the large amounts of datasets that are available today. Consequently, data journalists could take the role of the "intermediaries" between the data and the audience by turning complex datasets into easy to read and meaningful stories for the audience (Gray et al. 2012) (Lesage & Hackett, 2014). According to Simon Rogers it is a new role for journalists and it acts as a bridge and guide between those in power who have the data (and are unable to understand and communicated it) and the public who desperately want to understand the data and access it but need help.

The whole process of finding ways to work with data and be able to present them with a story that can relate to the public adds value to journalism in multiple aspects, one of which is the adoption of a new role for journalists as "sense-makers" (European Journalism Centre, 2010). Especially today, that there is an ongoing social change towards information abundance, people get confused and cannot process all the available data. Another aspect is that journalists could develop new social roles by providing data that has been carefully selected and cleaned to the public, moving from a "deadline

driven practice of news reporting into a more blended practice of reporting and social organizing” (Hamilton & Turner, 2009). It can create a new environment where the audience can collaborate with the journalist and interact with each other by exploiting the possibilities that the digital technology has to offer. Thus, data journalism could also represent a new facet of journalism as it provides the readers with interactive visualizations that let them explore and get engaged with the data (Aitamurto et al. 2011).

Apart from creating new roles for journalists, data journalism can enhance and support the watchdog role of journalism. Many journalists perceive it as a form of accountability journalism where the investigation aims to hold government account on behalf of the public interest (Felle et al. 2015). It could represent a solution for the current situation in the field. Traditional media are not anymore trusted and they cannot compete with the rapid development and popularity of social media. But, data journalism takes advantage of the current technology in order to find and publish datasets and reports in a way that the public can control those in power (Hamilton et al. 2009). The literature underlines that the emergence of social media and the general trend towards small mobile screens pose a challenge for traditional journalism. As a way to become more innovative and cope with the changes in gathering, filtering and presenting data and the whole datafication of society news organisations intent to incorporate data journalism in their practices (Engebretsen et al. 2018). It is well established though that data is not something new for journalism. The difference now can be traced at the digitation of information that requires a shift to a more technical and computational approach from journalists. Thus, data- whether “big” in the sense of being too complex for traditional database management software, or simply “big” in its potentially transformative import- has taken on particular relevance for news. “The open question in 2014”, one report noted, “is not whether data, computers and algorithms can be used by journalists in the public interest, but rather how, when, where, why and by whom” (Howard, 2014) (Lewis & Westlund, 2015) (Stalph, 2017). However, the term accountability also refers to the journalists’ responsibility to the public. So, in this way, accountability means more openness from the journalists in publishing the raw data they used for their story (Lesage & Hackett, 2014).

In this sense, data journalism can be seen as a means to enhance transparency in the journalistic process, as the audience is able to check the datasets as they were, before being refined or processed to have the story. Additionally, another aspect of transparency in the data journalism process is the visualization of data. Publishing a visual representation of the raw data, especially if it is interactive, gives the public the opportunity to explore and get engaged with the information they need (Aitamurto et al. 2011) (Lesage & Hackett, 2014). Transparency potentially means for journalism that it may gain back the public’s trust that has been in jeopardy the last few years.

The impact of data journalism may further enhance the role of serving the public interest mainly by discovering data stories that reveal mistakes, corruption and problems/malfunctions in the system. The stories, on their turn, can put pressure on the governments and all those in power in order to take action on behalf of the public’s best interest (Aitamurto et al. 2011). Because of data journalism the role of journalism starts to be re-defined in the sense that the data stories not only focus on uncovering and underlying a specific problem, but also on resolving it (Felle et al. 2015).

As it is mentioned in the book “Facts are sacred” data journalism does not necessarily need to have an investigative role to play in all circumstances in order to expose governments. However, sometimes it is more meaningful for journalists to use data as an analytical tool on facts that are a bit more complicated. That way data journalism serves the public in making it see the world a little clearer (Rogers, 2013) (Stoneman, 2015).

5. The Challenges of Data Journalism

As it is already mentioned, data journalism has started to gain field into many newsrooms that want to keep up with the changes in the technology and in the society, itself, as well as to bring new value to journalism. Though, it is still quite far from being established in its entity as there are many challenges to face.

To begin with, one of the main problems is the data published by the governments, other governmental bodies and organizations. Many of the datasets are not published soon enough for the journalists to use them for a news story. For instance, when a journalist starts an investigation the data available usually is up to two years old, which has no value in a journalistic sense (Felle et al. 2015) (Stoneman, 2015). This is also the case in US, where journalists have to deal with bureaucracy, obfuscation, or years of drawn-out wrangling over Freedom of Information Act request, fees and redactions; even though it is considered to have a strong Freedom of Information and governmental bodies tend to publish their data (Howard, 2014). In a local level, though, it is even harder to find open data so journalists have to search on their own to gather the information they need or use Freedom of Information to obtain it from the local administrations (Rogers, 2013).

As Jonathan Stoneman mentions in his paper “Does Open Data needs Journalism?” (2015) the situation in UK is the same: “Relatively few data are updated in real time, and even fewer datasets cover the mainstream of story content – money, vote, and the data which the government uses in its decision making. For everything else, journalists in the UK still tend to ask for the data they need, using Freedom of Information”.

Additionally, there are still countries that do not have a very strong Freedom of Information Act that make it even more difficult for data journalists to gain access to data, as for example Spain, Italy, Argentina and Africa. (Howard, 2014) Generally, sometimes it can be challenging for journalists to access governmental data and their relationship with governmental bodies is quite tensed (Howard, 2014).

Another important issue related to the open data is the fact that big part of the data requires cleaning or further adjustments by the journalists (Howard, 2014). Data refinement is a process that can take a lot of time to be done correctly, which is a disadvantage in the journalistic field. Thus, another obstacle for data journalism is the deadlines and the pressure of time, when it comes to news. Even more, because most of the organizations have low budgets they do not invest in data journalism project, especially the local news media that have even fewer resources (Howard, 2014) (Felle et al. 2015). This creates a big gap between small local media and big news organizations, that can afford to develop further data journalism (Felle et al. 2015). It is indicated that there have not been any advertising revenues from data journalism yet and very often these stories do not have sufficient success as front pages or a significant impact to journalism (Stoneman, 2015) (Felle et al. 2015).

Regarding the journalists themselves, even though there is an increasing demand for people who have computational and analytical skills because of the great availability of data, still the sector lacks of these abilities. Lots of journalism schools and universities have not yet fully integrated data journalism in their curriculum. (Howard, 2014) Apart from that, there is a general skepticism about data journalism from supporters of traditional journalism and its value for journalism, so it gets quite challenging for newsrooms to sustain an integrated data team that requires both programmers and journalists (Felle et al. 2015). The literature suggests that there are observed conflicts between the traditional principles and practices of journalism and the technological development. There is also a confusion regarding the hierarchy of priorities in journalism practices. More specifically, will technology and data culture impose their principles in journalism or will journalistic values prevail and thus will journalists use data to enhance the journalistic process? Another issue is interactivity, as it can be used as a tool that allows the readers engage with the story but on the other hand takes away the role of traditional journalist as a storyteller. Based on the literature it is indicated that many times interactive elements are used to attract the readers and have an explanatory function (De Mayeyer et al. 2015) (Hermida & Young, 2016). Moreover, it can also be a challenge, for journalists, to find their place in the new environment, where much of the information is open and available for everyone to find it. Thus, open data movement does not focus on journalists as end users of the data (Stoneman, 2015). Additionally, data journalists have to face the mistrust of the audience towards media that has increased the last few years and they have to work harder in order to gain it again (Felle et al. 2015). According to the report "Bias, Bullshit and Lies: Audience Perspectives on Low Trust in the Media" published by the Reuters Institute for the Study of Journalism at the University of Oxford despite the fact that people tend to get informed more often from social media, they seem to hold responsible the professional news media for low trust. Data journalism can alter people's perception of journalism, as people can participate in the journalistic process, for instance with practices such as crowdsourcing, which has the double function of offering readers the role of active participants in media discourse and, at the same time, enabling them to monitor their own positions in the visualised landscape of public opinion (Engebretsen et al. 2018) (Young et al. 2018) (Coddington, 2015).

Data journalism, among all these, has also to face the criticism that it can be consider an elite kind of journalism, despite the fact that it depends in a more open and democratized environment, as it refers only to people with internet access and basic knowledge of data analysis and visualizations (Felle et al. 2015). Finally, there is still a technical obstacle for data journalism with the current Content Management Systems that are, as Mirko Lorenz claims in an interview for the European Journalism Centre in 2010, page-orientated and they help you put together an article. But they greatly ignore the data in the text.

6. The Future of Data Journalism.

Despite the difficulties data journalism has to face, its future seems to be quite promising as more institutions, governments and citizens are getting interested in open data. In the near future it is expected that news media will make data journalism their primer choice for reporting and journalists will be working in data analysis and visualization to create news stories. Additionally, more tools will emerge that will make the data analysis process easier for the journalists, while at the same time data journalism will become a cheap and efficient sector for the news organizations. More players will enter the field apart from news organizations, like for example NGOs and independent publishers, that will probably offer data services but not in a journalistic sense. Generally, it is very likely for these bodies to exist symbiotically with news organizations, by offering their services to journalists and the latter providing data-driven stories (Aitamurto et al. 2011).

There are a number of predictions regarding the future of data journalism, based on three reports (The report by Duke Reporters' Lab at the DeWitt Wallace Center for Media and Democracy in the Stanford School of Public Policy, The "Core Skills for the Future of Journalism" report by Poynter.com and the report on innovation at the New York Times).

Firstly, news media will take full advantage of data and make it an asset, by becoming providers of datasets and APIs with payment. Secondly, it is predicted that new tools will be available that will make the process of analyzing and presenting the data way easier than it is now, making it possible even for journalists who do not have the acquired skills to create data-driven stories. These tools will be offered by big companies, such as Google and Amazon, or hackers, university students etc. (Howard, 2014) (Felle et al. 2015)

Apart from tools, it is predicted that news applications for smart phones, tablets and laptops will emerge on order to support multimedia news stories, as well as well as data-driven projects. Additionally, the devices maybe also used to collect data instead of only displaying it. Another assumption for the future is the growth of robo-journalism that is machines, like drones, wearable computers or sensors will be part of the newsgathering process. However, they are not going to replace people, as journalists still will be the ones to write a story out of the data that has a meaning to people's lives. It is also likely that people will begin to ask for more transparent work by the journalists, as well as there will probably be a dispute for opening more data as a journalists' demand, but governments and companies will be more reluctant on publishing controversial information. People's interest in data stories will also lead to a more personalized content depending on their preferences, searches and clicks (Felle et al. 2015) (Howard, 2014).

Apart from predictions, the report offers some future suggestions that are important for data journalism to develop even further and become a standard section in the newsrooms. Initially, it is very likely that wrong data or misinterpreted one will be published, as bad regular news stories come out in the news often enough. Thus, journalists will need to be very careful and critical with the datasets they work, while at the same time they have to be very accurate with the numbers and graphics they choose to use. Furthermore, as journalists get more involved with data stories they will have to protect any kind of sensitive information and newsrooms need to obtain better security systems to secure their datasets. To be able to process data, it could be wise for journalists to have additional education in social sciences and statistics, to better understand how data analysis works and be able to relate it to people's needs. Regarding the closed data from governments and companies that is predicted to happen, news organization could collaborate with universities and libraries to find information, store data and even get important feedback. Finally, newsrooms need to become more diverse and incorporate in their teams people from all groups eliminating gender, regional and any other kind of inequalities. Having a diverse will definitely lead to better journalism, as different points of view will be represented when creating news stories (Felle et al. 2015) (Howard, 2014).

Focusing more on the Open Data Movement, journalism in general should become more active in promoting it and embrace the value and potentials of open data. Data stories, for instances could be about gaps in datasets or in data scarcity. Additionally, it is imperative for journalists to collaborate with governmental bodies and, particularly, those responsible for providing data in order to help them improve the data published and force them for more transparency (Felle et al. 2015) (Stoneman, 2015).

As for the developing countries, there are three suggestions to improve data journalism. Firstly, news organizations can invest in fellowships, in order to bring people in the team with additional data analysis or visualization skills and with a different background. Secondly, it could be quite useful to organize data journalism boot camps and hackathons that will give the opportunity to professional journalists get in touch and be educated in data journalism techniques. Thirdly, cross-border global data investigations could be a great source for interesting and meaningful stories, while at the same time it is a chance to create cross border collaborations in exchanging data (Felle et al. 2015). To conclude, there are many potentials for journalism in data- driven journalism, but it requires a number of structural, technical and social changes, mainly by news media and governments.

7. Conclusions

This paper has discussed various issues concerning data journalism. Specifically, it analytically defined data journalism and discussed, in a detail manner, its relationship with data and the Open Data Movement. The value and the challenges data journalism is facing were also examined. Finally, the study collected and presented, in a comprehensive manner, predictions about the future of data journalism.

It is generally believed that data journalism is one of the pillars of modern journalism. As our society has already moved in the big data era, data journalism has the ability to play a significant role in utilizing the available data sets. This ability, that cannot be easily exercised by citizen journalists, along with the problem of fake news, is an opportunity for journalists to gain back their credibility and continue to act as the gatekeepers of information.

However there is still some confusion and skepticism around data journalism regarding its role and value. On one side many journalists hesitate to work with data, statistics and programming. On the other side news organizations, due to low budgets, do not take often the initiative to adopt new techniques, especially data journalism that has no business model to support its success and has not brought significant revenues to news media until now. The situation becomes even more difficult for local media that have fewer resources and less open data from local administrations to work with.

Nevertheless, local communities could become an important potential for Open Data Movement to expand and develop further. It is clear that there is a continuous improvement regarding open data, as more governments, institutions and organizations start to make their data available, and more NGOs and journalists push for further openness. But still there are problems to be tackled with the data available, concerning the format of the data, its lack of updating data and the type of data published. On the contrary, the tools, applications and software on data analysis and data visualization keep increasing and becoming easier for journalists or anyone else. Thus, most of them do not require any coding knowledge and offer a free version, as well as more advanced paid versions.

At the same time, journalists need to adapt to this new evolving environment and find their place, as openness and digital technology has contributed to the development of citizens' journalism and also other organizations have started to work with data and publish visualizations. Journalists could exploit all the possibilities data can offer and could be providers of datasets and data analysis. While using their journalistic status they could focus on creating stories with data that will serve the public interest. Instead of seeing digital technology and the changes it causes as threat, they could turn it into an opportunity. There is actually a belief that data journalism could be the answer to the crisis in journalism and could gain back the audience's trust. Additionally, it could strengthen journalism's watchdog role and force those in power to be more accountable to the public.

The role of education is also very important for data journalism to develop further and become established in journalism. There has been a tendency in journalism schools to include data analysis and data visualization courses during the last few years, or offer a master degree in data journalism. But usually the courses are introductory and they are not yet incorporated in all schools' curriculum. Thus, universities and journalism schools could step up and renew their study programs incorporating data journalism.

To conclude, recently the interest and research in data journalism is continuously increasing. More research is required though in order to offer more empirical data and lead to a better understanding of the evolution of data journalism and its effect on traditional journalism.

References

- Aitamurto T., Sirkkunen, E., & Lehtonen, P. (2011). Trends in data journalism. *Next media*. Retrieved from: http://virtual.vtt.fi/virtual/nextmedia/deliverables-2011/D3.2.1.2.B_hyperlocal_trends_In%20data_journalism.pdf
- Appelgren, E., & Nygren, G. (2014). Data Journalism in Sweden: Introducing new methods and genres of journalism into "old" organizations. *Digital Journalism*, 2(3), 394-405. <https://doi.org/10.1080/21670811.2014.884344>
- Baack, S. (2011). A new style of news reporting: Wikileaks and data-driven journalism. In B. Rambatan, & J. Johanssen (Eds.), *Cyborg Subjects: Discourses on Digital Culture* (pp. 113-122). http://www.academia.edu/8156415/A_new_Style_of_News_Reporting_Wikileaks_and_Data-driven_Journalism
- Borges-Rey, E. (2017). Towards an epistemology of data journalism in the devolved nations of the United Kingdom: Changes and continuities in materiality, performativity and reflexivity. Retrieved from: <https://doi.org/10.1177/1464884917693864>
- Cairo, A. (2013). *The functional art: An introduction to information graphics and visualization*. Berkeley, CA: New Riders.
- Coddington, M. (2015). Clarifying Journalism's Quantitative Turn. *Digital Journalism*, 3(3), 331-348. <https://doi.org/10.1080/21670811.2014.976400>
- De Maeyer, J., Libert, M., Domingo, D., Heinderyckx, F., & Le Cam, F. (2015). Waiting for Data Journalism. *Digital Journalism*, 3(3), 432-446. <https://doi.org/10.1080/21670811.2014.976415>
- Engelbrechtsen, M., Kennedy, H., & Weber, W. (2018). Data Visualization in Scandinavian Newsrooms. Emerging Trends in Journalistic Visualization Practices. *Nordicom Review*, 1-16. <https://doi.org/10.2478/nor-2018-0007>
- European Journalism Centre. (2010). *Data-driven journalism: What is there to learn?* Retrieved from: <https://www.slideshare.net/duarteromero/datadriven-journalism-what-is-there-to-learn-12056926>
- Felle, T., Mair, J., & Radcliffe, D. (2015). *Data Journalism: Inside the global future*. Suffolk, England: Abramis Academic Publishing.
- Fink, K., & Anderson, C. W. (2015). Data Journalism in the United States. *Journalism Studies*, 16(4), 467-481. <https://doi.org/10.1080/1461670X.2014.939852>
- Gray, J., Bounegru, L., & Chambers, L. (2012). *The Data Journalism Handbook*. California, CA: O'Reilly Media, Inc.
- Hamilton, J., & Turner, F. (2009). *Accountability through algorithm: Developing the field of computational journalism*. Retrieved from:

- <https://web.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf>
- Hermida, A., & Young, M. L. (2016). Finding the Data Unicorn. *Digital Journalism*, 5(2), 159-176. <https://doi.org/10.1080/21670811.2016.1162663>
- Howard, A. B. (2014). *The Art and Science of data-driven Journalism*. Tow Center for Digital Journalism. Retrieved from: <https://towcenter.org/wp-content/uploads/2014/05/Tow-Center-Data-Driven-Journalism.pdf>
- Koupidis, K., Bratsas, C., Karampatakis, S., Martzopoulou, A., & Antoniou, I. (2016). Fiscal Knowledge discovery in Municipalities of Athens and Thessaloniki via Linked Open Data. *11th International Workshop on Semantic and Social Media Adaptation and Personalization, SMAP 2016*, Thessaloniki, Greece, October 20-21, 171-176. <https://doi.org/10.1109/SMAP.2016.7753405>
- Krum, R. (2014). *Cool infographics*. Indianapolis, IN: John Wiley & Sons, Inc.
- Lesage, F., & Hackett, R. A. (2014). Between Objectivity and Openness—The Mediality of Data for Journalism. *Media and Communication*, 2(2), 42-54. <https://doi.org/10.17645/mac.v2i2.128>
- Lewis S. C., & Westlund, O. (2015). Big Data and Journalism. *Digital Journalism*, 3(3), 447-466. <https://doi.org/10.1080/21670811.2014.976418>
- Loosen, W., Reimer, J., & De Silva-Schmidt, F. (2017). Data-driven reporting: An on-going (r)evolution? An analysis of projects nominated for the Data Journalism Awards 2013–2016. Retrieved from: <https://doi.org/10.1177/1464884917735691>
- Mair, J., & Keeble, R. L. (2014). *Data Journalism: Mapping the future*. Suffolk, England: Abramis Academic Publishing.
- Philippides, M., Karampatakis, S., Ioannidis, L., Mynarz, J., Svátek, V., Bratsas, C. (2016). Towards Budget Comparative Analysis: the need for Fiscal Codelists as Linked Data. *International Conference on Semantic Systems-SEMANTiCS2016, Leipzig, Germany, September 12-15*. Retrieved from: <http://ceur-ws.org/Vol-1695/paper32.pdf>
- Rogers, S. (2013). *Facts are sacred*. London, England: Faber and Faber limited.
- Stalph, F. (2017). Classifying Data Journalism. *Journalism Practice*. <https://doi.org/10.1080/17512786.2017.1386583>
- Stoneman, J. (2015). Does Open Data need Journalism? *University of Oxford, Reuters Institute for the Study of Journalism*. Retrieved from: <https://reutersinstitute.politics.ox.ac.uk/sites/default/files/research/files/Stoneman%2520-%2520Does%2520Open%2520Data%2520need%2520Journalism.pdf>
- Veglis, A., & Bratsas, C. (2017). Reporters in the age of Data Journalism, *Journal of Applied Journalism & Media Studies*, 6, 225-244. https://doi.org/10.1386/ajms.6.2.225_1
- Young, M. L., Hermida, A., & Fulda, J. (2018). What Makes for Great Data Journalism? *Journalism Practice*, 12(1), 115-135. <https://doi.org/10.1080/17512786.2016.1270171>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the [Creative Commons Attribution license](#) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.