Deconstructing data-driven journalism

Reflexivity between the datafied society and the datafication of news work

Colin Porlezza, City, University of London

Abstract

The datafication of society is characterized by data abundance, the increasingly dominant position of algorithms that influence the lives of millions, and a secular belief in the beneficent power of quantitative data promising a new social order. Within this wider transformation process of society, the trend of datafication has been embraced, with some resistance (Lewis & Waters, 2017), within the journalistic field as well, leading to new forms of data journalism. These changes in journalism offer new opportunities to analyse the datafication of society, relying on the same means – data and algorithms – that distinguish the datafied society itself. This entails a reflexivity between the instruments that characterize a datafied society, and their implementation in journalism in order to observe them. The proposed paper offers a framework to critically analyse the reciprocal relationship between journalism and the datafied society by deconstructing the notion of datafication into four specific functions of newswork: a) the *observation* of datafication-related issues like dataism; b) the *investigation* of data-surveillance; c) the *generation* of new data-networks by journalists; and d) *unblackboxing* algorithms in order to foster algorithmic accountability. Journalism is therefore not only a reflection of the broader datafication-related transformation in society, but the central means to critically showcase its problems – albeit not being immune to challenges of transparency on its own.

La datafication della società è caratterizzata dall'abbondanza dei dati, dalla posizione sempre più dominante degli algoritmi che influenzano la vita di milioni di persone e da una fede indiscriminata nel potere benefico dei dati quantitativi. All'interno di questo più ampio processo di trasformazione della società, la tendenza della datafication è stata accolta, con qualche resistenza (Lewis & Waters, 2017), anche all'interno del campo giornalistico, portando a nuove forme di data journalism. Questi cambiamenti nel giornalismo offrono nuove opportunità per analizzare la datafication nella società, facendo affidamento sugli stessi mezzi – dati e algoritmi – che contraddistinguono la stessa società. Ciò comporta una riflessività tra gli strumenti che caratterizzano una società dataficata e la loro implementazione nel giornalismo proprio per osservarli. L'articolo qui proposto offre un quadro teorico per analizzare criticamente il rapporto reciproco tra giornalismo e società informatizzata decostruendo la nozione di datafication in quattro funzioni specifiche: a) l'osservazione di questioni relative alla datafication come il dataism; b) l'indagine sulla sorveglianza dei dati; c) la generazione di nuove reti di dati da parte di giornalisti; e d) unblackboxing algoritmi al fine di favorire una responsabilità algoritmica. Il giornalismo non è quindi solo un riflesso della più ampia

trasformazione legata alla datafication nella società, ma è il mezzo principale per mettere in luce in modo critico i suoi problemi, anche se non è immune alle sfide della trasparenza.

Introduction

The datafication of society is characterized by data abundance, a platform-capitalism grounded on the use of big data (Ramge & Mayer-Schönberger, 2017), the reliance on data as a means for surveillance or secular belief (van Dijck, 2014), and the increasingly dominant position of algorithms as cultural artefacts that announce a new social order (Uricchio, 2017). Not only are algorithms influencing the public sphere and our own news consumption by deciding what we see on our social media streams, they might even play an increasingly important role when it comes to elections – the Cambridge Analytica scandal in relation to the US presidential election, where the obscure analytics company used millions of private Facebook user accounts to target electoral messages, offers an initial taste of what is to come and how data are in fact becoming a new currency or, literally, the oil of the twenty-first century (Kennedy, 2018; Spitz, 2017).

The trend of datafication has been embraced, albeit with some resistance (Lewis & Waters, 2017), within the journalistic field as well. However, this hesitancy is not only limited to journalists: since the revelations of Edward Snowden in 2013 we are now able to understand that there are private and public data industry at work, collecting every single day a plethora of data about citizens for national security reasons, elections, marketing or plain and simple surveillance (Greenwald, 2014). Datafication, in the public debate, has thus often been criticized as a threat to democratic institutions because the data collecting actors are acting in a grey and loosely regulated area. Particularly when it comes to news and information, there is no lack of critical voices as the enduring discussions on filter bubbles and echo chambers show. Even within the scientific community of media scholars the academic debate on data and the dysfunctional effects of social media apparently undermining the possibilities of constructive dialogue and exchange of opinion endure, with mixed and often incoherent findings (Vaccari et al., 2016; Bruns, 2017).

Although there are also visions of how datafication can actually contribute to democratic processes particularly in relation to political accountability, citizen empowerment and activism (Baack, 2018a), most journalists for a long time remained hesitant or outright hostile to adapt to the Web (Ryfe, 2012). Only slowly, as C.W. Anderson (2015) showed, (U.S.) journalists have adopted a more quantitative and data oriented perspective, but the efforts continued to be unevenly scattered throughout the journalistic field. Often, journalists even failed in trying to understand the opportunities of data infrastructures in their everyday news work (Ananny, 2018): despite the possibilities of citizen participation and the opportunities to broaden the range of sources, research has shown that journalists concentrated on elite sources, which made the disruption process of traditional journalism even worse (Wahl-Jorgensen et al., 2016).

Even if a core of irreducible traditionalists still oppose resistance to the "hazards" of digital technology, a range of new actors – entrepreneurial

journalists, digital journalism start-ups and so called interlopers at the periphery of the journalistic field such as WikiLeaks (Eldridge, 2018) – embraced the disruption, creating new forms of journalism, news and storytelling by taking advantage of the raising data collections (Usher, 2016). This process led to new forms of data journalism, entailing not only new journalistic epistemologies (Splendore, 2016), but offering also new opportunities (Porlezza, 2016), in turn, to analyze the datafication of society. As a result, these new and at the same time alternative approaches to journalism rely on the same instruments and means (data and algorithms) that distinguish the datafied society itself. New knowledge is brought into the journalistic environment and, after having been incorporated in the production routines, is used to analyse wider social changes. In other words: it facilitates reflexivity between the instruments that characterize a datafied society, and their implementation in journalism in order to observe them.

The paper offers a theoretical framework to critically analyze the reciprocal relationship between the datafied society and the datafication of journalism by deconstructing the notion of datafication into four specific functions of newswork: a) the *observation* of datafication-related issues and their discussion in the public sphere; b) the *investigation* of data-surveillance; c) the *generation* of new data-networks; and d) *unblackboxing* algorithms in order to foster algorithmic accountability. Journalism is therefore not only a reflection of the broader datafication-related transformation in society (Loosen, 2018), but the central means to critically showcase its specific problems with the very same instruments datafication entails – albeit confronted with challenges of transparency on its own.

In the next paragraph the study offers a rapid literature overview of how data and computation have changed newswork. The following chapters then outline the four areas of reflexivity between the datafication of society and journalism. Finally, by critically evaluating the contributions of data to both journalism and society, particularly as the former observes the latter, the paper discusses how the proposed conceptual lens can improve our understanding of datafication in contemporary societies and modern journalism.

The datafication of journalism

The use of data in journalism is all but new. Its historical origins are usually related to the so called computer-assisted reporting (CAR), a term coined during the 1950s in the U.S., meaning that journalists would work with data and use computers to analyze them. More specifically, references are made to the U.S. presidential election of 1952, when the CBS television network apparently used a Remington Rand UNIVAC computer to predict the outcome of the race between Eisenhower and Stevenson. Then again, others trace it back to the 1960s, when journalist and scholar Philip Meyer, then working for The Detroit Free Press, used an IBM 360 mainframe computer, social science research methods and statistics to analyze survey questionnaires about the Detroit riots in 1967. Meyer (2002) dubbed the principles of his approach as precision journalism in his seminal book of the same name published in 1972, a term that has since

established a new trend within the field of journalism when it comes to working with large amounts of data.

However, the use of large amounts of data began much earlier in journalism history. As Klein (2014) explains, 17th century periodicals that offered news from remote cities were already data-rich reports. In the early 18th century, newspapers began publishing tables and statistics about tax liens, mortality causes and rates, public education, commodity prices as well as cargo inventories of arriving ships. Towards the end of the 18th century, specialized business publications such as The Wall Street Journal started to use more complex tables on a regular basis (Usher, 2016). But the use of data in reporting was not limited to economics or trade, it can also be observed in sports journalism: with the rising success of baseball in the U.S., reporters such as English born Henry Chadwick started in the mid 19th century to record data of baseball games, to analyze them statistically and to present them in box scores (Schwarz, 2004).

A historical approach to journalism's use of data allows not only to trace back the origins of data journalism way beyond CAR or precision journalism, it also permits to identify early patterns of reflexivity between journalism, technological changes and social movements. Lindén (2017) points out that technology has been one of the key factors in the development of media and journalism (see also McLuhan, 2011; Örnebring, 2010; Pavlik, 2000; Powers, 2012). Moreover, innovations force journalists to keep up with the changes and to constantly improve their knowledge and skills. Anderson (2015) points out that throughout the history of journalism, data has played an important role, even if the type of data repeatedly changed: while the penny press focused mores strongly on oral accounts rather than documents, precision journalism brought quantitative data back in. While the survey movement of the early 20th century influenced journalism with its data collection techniques, current data journalists operate with powerful statistical programs and algorithms. Moreover, nowadays the diffusion of technological innovations might have never been quicker (Lindén, 2017).

With rising computational power, the advent of the Web and digital journalism, all journalism has become computer-assisted (Howard, 2014). Even if there are conflicting definitions in the scientific literature on what data journalism is, it can nevertheless be seen as an evolution of CAR. But there are differences in how data is used: while computer-assisted reporting used data mainly to garnish the story, in data journalism the data represents the story itself. Appelgren and Nygren (2014) describe it therefore as an emerging form of storytelling, where traditional journalistic working methods are blended with data analysis, programming and different visualization techniques. Additionally, data journalism usually makes datasets transparent and available, while CAR would store the data in internal databases without offering access to raw data to the wider public.

The increasingly widespread use of digital technology led to the formation of a new breed of journalists, although they still represent a highly specialized niche. In the context of these changes and challenges, journalism also opened up to new forms of cooperation with actors from other fields, both within and outside the

newsroom. Particularly when it comes to the handling of data, research shows that data journalists often form communities of interlocked practices with different actors such as programmers or hackers (Lewis & Usher, 2014) or civic technologists (Baack, 2018b). These collaborations involve that new actors, some of whom have not been influenced by journalistic culture in the first place, have entered the institutionalized field of journalism, influencing its practices, routines, professional norms and culture with notions originating from other cultures. Russell (2016, p. 6) has shown how hackers and hacktivists, among others, are playing an increasingly important role in shaping contemporary journalism, "expanding what it means to be involved in the production of news and, in the process, gaining influence over how traditional news stories and genres are constructed and circulated".

The increase of structured and machine-readable data across society led to yet another transformation: the use of specific forms of computational and automated data analysis through algorithms. These recent technological developments allow newsrooms to improve on automated news production through computer algorithms. These new forms of newswork called automated, robot or algorithmic journalism (Dörr, 2016) promise to produce a literally unlimited number of articles, at a lower price and higher speed compared to their human counterparts. In addition, the readers perceive the quality and credibility of the journalistic output as equal (Wölker & Powell, 2018). At the same time, however, these new opportunities have given rise to debates both in the journalistic and scientific community on whether the automation of journalism will make some editorial tasks, routines or even journalists redundant – causing eventually an "algorithmic turn" (Napoli, 2014) in journalistic news production.

Such forms of algorithmic journalism can be understood as

> "the (semi-) automated process of natural language generation by the selection of electronic data from private or public databases (input), the assignment of relevance of preselected or non selected data characteristics, the processing and structuring of the relevant data sets to a semantic structure (throughput), and the publishing of the final text on an online or offline platform with a certain reach (output)" (Dörr, 2016, p. 702).

Bringing thus together data and computational power in journalism, this means that "the combination of algorithms, data, and knowledge from the social sciences (is) to supplement the accountability function of journalism" (Hamilton & Turner, 2009, p. 2). However, even if algorithmic journalism is still in its early phase and the technology still suffers from constraints with regard to the tasks it can and cannot perform, and even if its implementation within newsrooms is full of questions marks (Graefe, 2016), it still allows us to get first insights on how the triptych between journalism, data and algorithms might evolve in the future.

If journalists themselves are able to better understand the affordances and inner workings of opaque technologies such as algorithms, they will be in a superior position to shed light on the datafication and the related algorithmic turn in society.

Reflexive datafication: a framework for critical analysis

Time and again we are confronted with different narratives that describe major transformation processes within society. Journalism plays a central role in shaping these narratives, not only because journalism has a stake in public communication, but also because it establishes the narratives in the public sphere. Currently, many narratives focus on the issues of data, big data and algorithms and how they influence the lives of each and every one of us, ranging from extensive utopian to dystopian rhetoric (boyd and Crawford, 2012).

Loosen (2018) well explained why it is useful to look particularly at journalism as a central domain to understand the wider implications of datafication in society and what kind of influences it exerts on public communication:

> "This allows us to not only better comprehend journalism's present transformation towards a more data-based, algorithmed, metrics-driven, or even automated practice, but, to consider this transformation as a reflexive process: a process that is at the same time part of a changing media environment and is journalism's response to – as well as an act of encouraging – the datafication of society."

It becomes thus clear that journalism does not operate in a vacuum, but within a constantly changing social and technological environment it is forced to adapt to. At the same time, journalism also critically observes datafication in society given that it is a transformation process that, indeed, reaches far beyond journalism, media and the construction of public spheres. The combination of data and algorithms do impact relevant domains of society such as "access to information (Google) and the social world (Facebook), [...] finance (algorithmic trading) and governance (from predictive policing to NSA-style parsing of vast troves of data)" (Uricchio, 2017, p. 128). The widespread use of algorithms, as Uricchio continues, might well enable new cultural and social forms, albeit not in esclusively positive circumstances, as O'Neill (2017, p. 12) dazzlingly showed: many algorithms that run on our personal data are "opaque, unquestioned, and unaccountable, and they operate at a scale to sort, target, or 'optimize' millions of people." Newswork does not remain unaffected either, since we can observe the continuous integration of data and algorithmic journalism in news organizations – a proof of an increasing institutionalization of data-based computational and automated news production at an organizational level (Dörr, 2016).

It is therefore necessary to take a step back and to question what kind of impact (big) data and the so called "algorithmic turn" have on our lives and social reality. It is not just a question of whether and to what extent data and algorithms influence our lives. The questions is much broader, as Ananny (2016, p. 98) points out:

> "algorithms do not simply accelerate commerce, journalism, finance, or other domains – they are a discourse and culture of knowledge that is simultaneously social and technological, structuring how information is produced, surfaced, made sense of, seen as legitimate, and ascribed public significance."

Although data and algorithms radically transform the way information is handled, a gradual normalization of datafication as a new paradigm is taking place (van Dijck, 2014). This is the point where journalism needs to step in in both its function as a critical observer of society and, at the same time, as a social domain that adapts to a changing media environment by undergoing itself a process of datafication. Datafication not only touches journalism on various levels with regard to changing practices and epistemologies, as Splendore (2016) illustrated, but it also allows journalism to critically observe datafication in other domains promoting the public communication and raising the awareness about the many challenges of datafication.

Dataism and the observation of datafication-related issues

In a 2013 op-ed of The New York Times, journalist David Brooks coined the term "data-ism". With his expression Brooks described the blind faith in (big) data as an apparently objective and emotion-free solution to the many problems of mankind:

> "If you asked me to describe the rising philosophy of the day, I'd say it is data-ism. We now have the ability to gather huge amounts of data. This ability seems to carry with it certain cultural assumptions – that everything that can be measured should be measured; that data is a transparent and reliable lens that allows us to filter out emotionalism and ideology; that data will help us do remarkable things — like foretell the future.

Steve Lohr, one of his colleagues at the New York Times, offered a similar impression that big data "are a vehicle for a point of view, or philosophy, about how decisions will be – and perhaps should be – made in the future" (2015, p. 3). Both journalists describe the ideological grounds of datafication as a philosophy. The two examples show well how the underlying ontology has evolved from a contextual use of data limited to certain social domains to a widespread and leading principle, as social reality can be observed, tracked and transformed in structured data in real time. The behavior of citizens can now be easily accessed, understood and monitored, establishing "life mining" as a new paradigm (van Dijck, 2014).

Now that data can be collected in areas previously inaccessible to structured scrutiny, the promised lands of behavioral knowledge and factfulness seem ready to conquer. Data, or better metadata (automatically generated information about who was talking for how long to whom from where etc.), have become a currency in exchange for apparently free of charge services. Social media such as Facebook, Twitter, Youtube or Instagram are among the most important actors that collect and sell this data to third parties that want to commercialize products to specific target groups or advertise political ideologies to certain voter groups. Companies monetizing on data generate questions with regard to how such business models should be regulated in the future: "While the phenomenon is taking place in an environment of uncertainty and rapid change, current decisions will shape the future. With the increased automation of data collection and analysis – as well as algorithms that can extract and illustrate large-scale patterns in human behavior – it is necessary to ask which systems are

driving these practices and which are regulating them“ (boyd & Crawford, 2012, p. 664).

The use of large amounts of data is not limited to commercial or political activities. Datafication enthusiasts can be observed in academic research as well. Particularly when it comes to studies related to social media, researchers sometimes confound large datasets with representativeness. This kind of ideology about the power of data in explaining social reality is grounded in the belief of the objectivity of quantification. boyd and Crawford (2012) lay out how even large databases are always subject to bias and limitations that have to be taken into account: “Regardless of the size of a data, it is subject to limitation and bias. Without those biases and limitations being understood and outlined, misinterpretation is the result.“ The two authors rightly declare that size does not always matter, for it is always subject to limitations. Van Dijck (2014) argues the same way when she declares that big but messy data sets collected on different social media platforms are increasingly presented as complete representations of social interaction, superior to sampling, because the sheer volume of data compensates for its messiness.

However, the trust in the explanatory power of data rests on the false assumption that data are facts. This is not the case as Gitelman (2013, p. 7) states: “Data are not facts, they are ‘that which is given prior to argument’ given in order to provide a rhetorical basis. Data can be good or bad, better or worse, incomplete and insufficient”. Raw data is therefore subject to interpretation. Mayer-Schönberger and Cukier (2013) confirm this perspective as well when they declare that data sets include some intrinsic and hidden values actors are not able to discover. This becomes an even bigger issue when we turn to algorithmic data extraction and analysis, particularly because authority is increasingly expressed algorithmically as decisions are no longer based on human reflection (Pasquale, 2015, p. 8).

These forms of dataism need critical scrutiny, particularly as there have been many examples of false starts in data analysis, algorithms and artificial intelligence (Bostrom, 2014). Journalism plays a central role in the interrogation of the belief in the power of data – and the blind trust in algorithmic data analyses – particularly because newswork has, over the last couple of years, not only developed the instruments and the skills to analyze structured data on its own, but also because many journalists have been confronted with changing conditions in the news production on their own: the automation of newswork and the increasing use of robots writing journalistic articles force journalists to confront themselves continuously with the changing media environment and the challenges of data and algorithms, as datafication occurs reflexively both within and outside the newsroom.

In this regard, data and computational journalism are important means to not only scrutinize structured data on their own for news stories, but to critically observe and interrogate forms of dataism in society and bringing them up in the public communication – ensuring the accountability of data regulation, property as well as the proper use of data. In this sense, the Tow Center for Digital Journalism report on the art and science of data journalism concludes his list of recommendations with: “Be mindful of data-ism and bad data. Embrace

skepticism (Howard, 2014, p. 1).

The investigation of data surveillance

In 2013, the Guardian and its editor in chief, Alan Rusbridger, were threatened with legal action by the government of the United Kingdom for the possession of secret British and American surveillance data revealed by the whistleblower Edward Snowden. On July 20, 2013, a Guardian journalist and a computer expert were forced to destroy hard drives and memory sticks that contained some of the encrypted files, although additional copies of the files were stored in secure places all around the world. This is just one example that shows how the relation between the government's surveillance agencies and journalism has never been easy. This is due to the fact that there is a structural conflict between the institutional secrecy of intelligence agencies and journalism's normative principles of transparency and accountability (Allen, 2008; Thompson, 2004). However, for a long time, journalism was not interested in surveillance (Ruby, Goggin and Keane, 2017).

This situation has changed. As the Snowden leaks disclosed (still ongoing) mass surveillance programs by several governments such as the U.S. and the U.K., the relation between intelligence services and news organizations has since deteriorated. The phenomenon of mass surveillance has triggered a sharp reaction by some news organizations, which is grounded in the journalistic responsibility of acting as a critical observer of those in power, and a responsibility of informing the citizens about social issues such as intelligence services going rogue without any form of democratic checks and balances. Contrary to what one might expect, Heikkilä and Kunelius (2017, p. 267) showed that journalists' reactions to mass surveillance are not uniform and differ across countries, as they are shaped by "national settings and (...) domestic struggles of power and legitimacy".

In addition, many journalists – as the example of the Guardian showed – have been intimidated by governmental action. This strategy entails multiple consequences. On the one hand, intimidation and threats of legal prosecution can actually cause "chilling effects" among journalists, as they shy away from stories and investigations into issues related to surveillance. Moreover, the risk of being brought to court may also drain or even eliminate confidential sources willing to leak classified information to journalists, as Lashmar illustrated (2017). In his study, Lashmar also revealed how journalists expressed serious concern about the intelligence agencies tracking and surveilling journalists in order to identify and neutralize their sources. Waters (2017) showed that, in fact, putting journalists under observation make their investigations more difficult and the communications with their sources more complex.

Despite the surveillance from government authorities, many (regional) journalists still do not have, or only slowly acquire the necessary skills to protect themselves and their sources from being observed (Bradshaw, 2017). This is not only due to the limited resources regional newspapers have, but also to the missing knowledge about encryption. Thorsen (2016) showed that a great deal of the articles about encryption discuss the issue in a very superficial way, which

has serious implications both for journalistic freedom and for the public discourse about mass surveillance.

However, the surveillance of journalists in the wake of the Snowden leaks have nevertheless convinced some reporters to invest in encryption. It has now become a central strategy mostly for investigative journalists to protect themselves, their work, and to build up a safe environment (Brunton & Nissembaum, 2015, p. 62). In addition, it also offers their sources protection from tracking and retaliation (Di Salvo, 2017). In this regard, the emergence of mass surveillance required journalists to adapt to the new circumstances and to improve their knowledge and skills in relation to surveillance and how to protect themselves from snooping government agencies – a reflexive process. Additionally, journalists do not only protect themselves from surveillance, they increasingly use the same methods and applications to "surveil social network users and their content via sophisticated, professional apps that are also utilised by the police and security forces" (Thurman, 2017, p. 76).

Yet, these measures raise fundamental questions with regard to journalism's ethical orientation: transparency has always been one of journalism's ethical cornerstones, but both encryption as well as the use of surveillance programs to track for instance social media accounts clash with the norm. Eventually, journalists have to cope with the paradoxical situation that they are "subjected to forces of discipline and surveillance that might, in the end, run counter to the very goals that they seek" (Allen, 2008, p. 336).

The generation of data-networks

At the Nordic Datajournalism Conference 2016 in Helsinki, Nicolas Kayser-Bril declared in his keynote that data journalists should be more careful when it comes to the use of government data and rather collect data on their own. The main reason for his appeal was that governments and public administrations regularly produce data on the grounds of political reflections and aims. According to Kayser-Bril (2016) "being able to access bogus data is pointless. What is needed to make sense of the world around us is better data, free from government interference."

In addition to biased or incorrect data, many public administrations fail to implement open government strategies and freedom of information acts. This means that journalists (and citizens as well) are unable to access and scrutinize the administration's databases. The situation varies considerably across countries, even in Europe: while Sweden adopted its access to information law already in 1766, Italy's Council of Ministers approved a still porous Freedom of Information Act just recently in March 2016.

Both the accessibility as well as the possibilities of data analysis, i.e. the format of the data, influence how the practice of data journalism evolves in a given country. In an explorative comparative study, Porlezza et al. (2018) analyzed what kind of strategies data journalists adopt when it comes to the different levels of accessibility and the related possibilities of data analysis. The findings show that in those countries with a stable Freedom of Information Act like Switzerland or the U.K., most data journalists rely on public records for their

investigations. In countries with some restrictions or fairly young FOIA, like Italy, data journalists mostly rely on their abilities to collect and record data on their own, or on contributions originating from whistleblowers.

Given that Italian journalists, for a long time, had no right to access the databases of the public administration, they had to be inventive and develop alternative strategies of data collection such as crowdsourcing, scraping or the recombination of the few data scattered among different administrations and offices. This causes the paradoxical situation that journalists confronted with difficult data environments are usually better equipped to gather and collect data on their own, because they have hardly any other means to obtain significant data. The findings also showed that journalists in countries with a lax FOIA are more inclined to collaborate informally with specialists such as hackers, programmers and scholars, creating and fostering data-networks that involve collaborative efforts to collect and scrutinize (public) data.

Such networks like Hacks/Hackers (Lewis & Usher, 2014) not only bring together experts from different backgrounds in order to facilitate skill exchange. These networks confront journalists with different cultural backgrounds and norms that have their roots in different sub-cultures such as the open-source movement. The more journalists are using data and the closer the work together with actors from other technology and data-affine areas, they "progressively move from observation into analysis and denunciation, raising awareness and recommendations. This is the case of advocacy data journalism." (Milan & Gutiérrez, 2015, p. 125). The two authors assert that by actively combining communicative practices and the social elements of collective action, even data journalists could become data activists:

> "When advocacy journalism and investigative journalism cross paths with available data and data-analysis software, we have the sub-field of 'advocacy data journalism', which combines the traditional values of journalism with crunching data and a social change ethos" (Milan & Gutiérrez, 2015, p. 129).

In this regard, the datafication of journalism and the creation of new "data-networks" bring back the ethos of the old muck-racking investigative journalism, giving back democratic legitimacy to journalism (Sampedro, 2014). Given that journalism is directly affected by the datafication of society, it has not only changed journalism, but it has helped to revitalize a journalism in profound crisis. As data journalism becomes slowly institutionalized within major newsrooms, the main focus of journalists may also shift from "being the first ones to report to being the ones telling us what a certain development might actually mean" (Gray & Chambers, 2012, p. 4).

Particularly when it comes to data journalism and its networks, the reflexivity between the datafication of society and journalism occurs on different levels. Not only it entails structural changes with regard to the epistemology and the specific practices of journalism. Journalism itself has helped to developed different platforms or portals, thanks to networks like hacks/hackers or the Open Knowledge Foundation, where citizens and people interested in data-related issues can explore details about different (not only country-related) topics such e.g. government expenditures, transparency, migration etc.

Journalism could thus become a powerful actor in today's data-rich public sphere and act as a facilitator for data literacy by sharing its experience in accessing, analyzing and interpreting data. Data journalism in its more activist understanding could even be understood, as Milan and Gutiérrez (2015, p. 130) suggest, as a "new, advanced form of citizens' media, one that has a critical approach to big data at its core. Similar to citizens' media, pro-active data activism involves a politics of the quotidian, as it alters the everyday relationship between citizens and automatized data collection."

Unblackboxing algorithms

Algorithms have gained an important position in society as they affect the lives of millions of citizens. These algorithms can be defined as "sets of defined steps structured to process instructions/data to produce an output" (Kitchin, 2017, p. 14). Algorithms are powerful computational machines, which perform tasks that would be simply too complex or time-consuming to carry out by hand. As such, they facilitate the operations of organizations. However, in doing so, algorithms have become key actors in our society, as Steiner (2012, p. 214) affirms:

> "Algorithms already have control of your money market funds, your stocks, and your retirement accounts. They'll soon decide who you talk to on phone calls; they will control the music that reaches your radio; they will decide your chances of getting lifesaving organs transplant; and for millions of people, algorithms will make perhaps the largest decision of in their life: choosing a spouse."

Taking into account the dominant position of algorithms in our data-driven society, Diakopoulos (2013) declares therefore that they have become the new power brokers in society. In many different domains where algorithms have been introduced they have exerted radical transformation processes with regard to how these domains are organized and how they operate (Kitchin, 2017).

Particularly in relation to media algorithms play a crucial role. The influence of algorithm-based social media platforms such as Facebook and Twitter on news consumption and production is overarching. These new information intermediaries have transformed the way news is consumed, distributed and produced – as journalistic judgment has been overthrown by algorithms that are now "a key logic governing the flows of information on which we depend" (Gillespie, 2014, p. 167).

However, journalism itself is no exception to this trend: Initially limited to specific genres such as sports or financial reporting, algorithmic journalism has since seen a rapid development in different major news organizations. The Los Angeles Times, Forbes, the Associated Press, Reuters, the BBC and others have started to use automation and algorithms not only to streamline the news production, but also to "outsource" breaking news or articles based on structured data such as financial reporting. However, even if the introduction of algorithms in the newsroom can the introduction of automated journalism also created challenges and a "technological drama" within newsrooms, as journalists

became concerned about their authority and the future of journalistic labor, being afraid of additional layoffs as publishers look to save costs (Carlson, 2015).

Although algorithms carry out key functions in society and, since recently, also in journalism, their operations and inner workings remain opaque and often undisclosed (O'Neill, 2017; Pasquale, 2015). Given that it is extremely difficult to analyze thoroughly their "power structures, their biases, and influences that computational artifacts exercise in society" (Diakopoulos, 2015, p. 398), one of the main questions that arise is how can algorithms be held to account. Journalism has thus taken up on this issue and has started to drive its attention towards these new power brokers in society by the means of "algorithmic accountability reporting" (idem, p. 399), trying to critically scrutinize algorithmic power articulated in their ways to prioritize, classify, associate or filter.

One way of doing this is by reverse engineering. Diakopoulos (2013, p. 13) explains that „reverse engineering is the process of articulating the specifications of a system through a rigorous examination drawing on domain knowledge, observation, and deduction to unearth a model of how that system works." Enhancing the transparency of algorithms would be a first step towards the formulation of an ethics of algorithms (Ananny, 2016). However, reverse engineering requires specific skills that most journalists will not likely have. As most news organization just recently started to implement automated forms of journalism, the expertise on algorithms and reverse engineering will be limited. In this regard, once more, collaborative enquiries together with experts from other backgrounds such as informatics or computer sciences may help reaching the goal.

However, also in the case of datafication and the power of algorithms in society, two different levels of reflexivity can be observed. First, the increasing expertise of journalists with algorithms – also due to the implementation of algorithms within newsrooms – allows reporters to engage in algorithmic accountability reporting. This journalistic genre could be further empowered by taking advantage of the computer experts and hackers present in „data-networks".

Second, some of the issues related to the opacity of algorithms are present within newsrooms as well. Particularly in the era of data and automated journalism, where "much of the work may be obfuscated via analytic procedures, and data literacy among journalists and / or readers may hinder the valid interpretation of results" (Stark and Diakopoulos, 2016, p. 1), enhancing transparency is key. This would mean sharing the different steps of an analysis, publish any form of code, software used as well as the raw data. This would allow news organizations to formulate journalistic standards and ethical norms for such complex products as algorithms. Drawing from these initial works, journalists could then formulate (ethical) criteria that could be applied to algorithms in general in order to provide full transparency.

Making the data journalism production more transparent can help reducing skepticism and allowing readers to better understand and interpret the data. Transparency would also allow data journalists to show that they have carried out their job to the best of their knowledge, which in turn reinforces the notion of media accountability, offering readers the opportunity to criticize errors in the

reporting (idem). Such a transparency-driven approach could also be applied to algorithms outside journalism.

Conclusion

The datafication between society and journalism is characterized by reflexivity. By deconstructing the notion of data-driven journalism, the paper has illustrated four areas in which this reflexivity can be actually observed. It offers thus a critical framework where characteristics of a datafied society are also present in journalism, which in turn are used to observe and scrutinize them. The four areas are: a) the *observation* of datafication-related issues like dataism; b) the *investigation* of data-surveillance; c) the *generation* of new data-networks; and d) *unblackboxing* algorithms in order to foster algorithmic accountability.

Journalism is not just another social domain where the datafication triggered fundamental transformations. It is, at the same time, the central means to critically observe datafication and to showcase its problems. Particularly in the case of mass surveillance and algorithms, (data) journalism can play a central role in shedding light on these issues. More than anything, problematizing the transformation processes in the public communication could help further develop data literacy among citizens and enforce digital citizenship.

However, research has repeatedly shown that journalists are confronted with structural problems: in the case of surveillance, journalists are often inhibited from articulating their critiques out of national security reasons – or due to chilling effects caused by legal threats from governments. In the case of algorithms, journalists might just lack the knowledge and the skills to perform in-depth analysis such as reverse engineering. This generates limitations to the potential of acting as a watchdog and a powerful actor in the public sphere. Additionally, even journalism itself is not immune against all the issues, which can be seen, for instance, in the repeated calls for editorial transparency in relation to algorithmic journalism.

Yet, as datafication continues to invade all domains of contemporary society, we need journalists, both as a concerned actors as well as a observers, to critically monitor datafication in order to raise the awareness among citizens and to foster digital citizenship in terms of citizens enacting themselves in cyberspace (Isin and Ruppert, 2015, p. 43). In this circumstance, journalism's responsibility towards society to serve as a critical observer, to hold the powerful to account, but also to strengthen public discussion in the public sphere is vital, given that it could become a breeding ground for and mobilise new forms of civic engagement, political action and data activism (Milan & von der Velden, 2016). Inform citizens about the risks and opportunities of „engaging with digital environments, and (...) how to both protect and claim their rights in cyberspace" (Hintz et al., 2017, p. 735) becomes a central tenet for journalism, but in order to transfer an adequate knowledge about datafication processes, journalists themselves have to live up to these standards.

The present framework could well serve as guidance for future research when it comes to the reflexivity between datafication in society and journalism. Future

research should focus more strongly on specific aspects such as dataism or the implementation of algorithms within newsrooms and what kind of effects they have on the coverage of similar issues, particularly when the reporting itself involves the use of (sensitive) data. At the same time, it would also be interesting to analyse how datafication processes within and outside newsrooms impact journalists' role perception and their notion on how to report (adequately) about data-related transformation processes in society through and with data. In addition, future studies could also concentrate on how networks composed of journalists and actors from other fields – such as computer sciences – evolve and whether and how they pro-actively engage in collective action or data activism and whether their efforts effectively change the way citizens approach the topic of datafication.

References:

Allen, D. (2008), *The Trouble with Transparency: The Challenge of Doing Journalism. Ethics in a Surveillance Society*. «Journalism Studies», 9, 3, pp. 323–340.

Ananny, M. (2016), *Toward an Ethics of Algorithms. Convening, Observation, Probability, and Timeliness***,** «Science, Technology, & Human Values», 41, 1, pp. 93-117.

Ananny, M. (2018), *Networked Press Freedom*. Cambridge: MIT Press.

Anderson, C. W. (2015), *Between the Unique and the Pattern: Historical Tensions in Our Understanding of Quantitative Journalism*, «Digital Journalism», 3, 3, pp. 349-363.

Appelgren, E. and Nygren, G. (2014), *Data Journalism in Sweden: Introducing New Methods and Genres of Journalism into 'Old' Organizations*, «Digital Journalism», 2, 3, pp. 394–405.

Baack, S. (2016). *What big data leaks tell us about the future of journalism – and ist past.* Internet Policy Review, available at: https://policyreview.info/articles/news/what-big-data-leaks-tell-us-about-future-journalism-and-its-past/413 (12.03.2018)

Baack, S. (2018a), *Civic tech at mySociety: How the imagined affordances of data shape data activism*, «Krisis», 2018/1, pp. 44-56.

Baack, S. (2018b), *Practically engaged. The entanglements between data journalism and civic tech,* «Digital Journalism», 6, 6, pp. 673-692.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

boyd, d. and Crawford, C. (2012), *Critical questions for big data*, «Information, Communication & Society», 15, 5, p. 662-679.

Bruns, A. (2017), *Echo chamber? What echo chamber? Reviewing the evidence*, 6th Biennal Future of Journalism Conference, 14-15 September, Cardiff UK.

Brooks, D. (2013), *The philosophy of data*. «The New York Times», 04.02.2013, Available at: https://www.nytimes.com/2013/02/05/opinion/brooks-the-philosophy-of-data.html

Carlson, M. (2015), *The robotic reporter*. «Digital Journalism», 3, 3, pp. 416-431.

Diakopoulos, N. (2013), *Algorithmic accountability reporting: On the investigation of black boxes*. A Tow/Knight Brief. Tow Center for Digital Journalism, Columbia Journalism School. Available at: https://towcenter.org/algorithmic-accountability-2/

Diakopoulos, N. (2015). *Algorithmic accountability. Journalistic investigation of computational power structures*. «Digital Journalism», 3, 3, pp. 398-415.

Di Salvo, P. (2017), *Hacking/Journalism*, «Limn», 8, available at: https://limn.it/articles/hackingjournalism/ (02.08.2018)

Dörr, K. (2016), *Mapping the field of Algorithmic Journalism*, «Digital Journalism», 4, 6, pp. 700-722.

Eldridge, S. (2018), *Online Journalism from the Periphery*. London: Routledge.

Gitelman, L. (2013), *'Raw Data' is an Oxymoron*. Cambridge: MIT Press.

Graefe, A. (2016), *Guide to Automated Journalism*. New York: Tow Center for Digital Journalism/Columbia Journalism School. Available at: http://towcenter.org/research/guide-to-automated-journalism/

Gray, J., Bounegru, L. and Chambers, L. (2012), *The data journalism handbook*, Cambridge: O'Reilly.

Greenwald, G. (2014), *No Place to Hide: Edward Snowden, the NSA, and the US Surveillance State*. Town: Hamish Hamilton

Hamilton, J. T. and Turner, F. (2009), *Accountability through algorithm: Developing the field of computational journalism*. Paper presented at the Summwe Workshop, 27-41. Retrieved from http://web.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf

Heikkilä, H. and Kunelius, R. (2017), *Surveillance and the Structural Transformation of Privacy*, «Digital Journalism», 5, 3, pp. 262-276.

Helmond, A. (2015), *The Platformization of the Web: Making Web Data Platform Ready*. «Social Media & Society», 1, 2, published online.

Hintz, A., Dencik, L. and Wahl-Jorgensen, K. (2017), *Digital Citizenship and Surveillance Society,* «International Journal of Communication», 11, pp. 731-739.

Isin, E. and Ruppert, E. (2015), *Becoming digital citizens*. Lanham: Rowman & Littlefield.

Kayser-Bril, N. (2016). *„Free data is over". Now, we need data to be free.* Keynote at the Nordic Data Journalism Conference 2016. Available at: http://blog.nkb.fr/data-free (12.03.2018)

Kennedy, H. (2018), *Living with Data: Aligning Data Studies and Data Activism Through a Focus on Everyday Experiences of Datafication*. «Krisis», 2018/1, pp. 18-30.

Kitchin, R. (2017). *Thinking critically about and researching algorithms*. «Information, Communication & Society», 20, 1, pp. 14-29.

Klein, S. (2015, Mar 17). *Antebellum data journalism: Or, how big data busted Abe Lincoln*. Available at: https://www.propublica.org/nerds/antebellum-data-journalism-busted-abe-lincoln

Lashmar, P. (2017), „*No more sources?*", «Journalism Practice», 11, 6, pp. 665-688.

Lewis, S. C. and Usher, N. (2014), *Code, collaboration, and the future of journalism. A Case Study of the Hacks/Hackers Global Network*, «Digital Journalism», 2, 3, pp. 383-393.

Lewis, N. P., & Waters, S. (2017). *Data journalism and the challenge of shoe-leather epistemologies*. «Digital Journalism», online first, https://doi.org/10.1080/21670811.2017.1377093.

Lindén, C.-G. (2017), *Algorithms for journalism: the future of news work*, «The Journal of Media Innovation», 4, 1, pp. 60-76.

Loosen, W. (2018), *Four forms of datafied journalism: Journalism's response to the datafication of society*. Bremen: ZeMKI Working paper No. 18.

Mayer-Schönberger, V. and Cukier, K. (2013), *Big Data. A Revolution that will transform how we live, work, and think.* London: John Murray Publishers.

Meyer, P. (2002), *Precision Journalism*, Lanham, Rowman & Littlefield.

Milan, S. and Gutiérrez, M. (2015), *Citizens' Media Meets Big Data: The Emergence of Data Activism.* «Mediaciones», 14, pp. 120-132.

Milan, S. and von der Velden, L. (2016). *The alternative epistemologies of data activism*. «Digital Cutlure and Society», 2, 2, pp. 57-74.

Napoli, P. M. (2014), *Automated media: An institutional theory perspective on algorithmic media production and consumption*. «Communication Theory», 24, 3, pp. 340-360.

O'Neill, C. (2017), *Weapons of math destruction. How big data increases inequality and threatens democracy*. New York: Crown Publishing.

Örnebring, H. (2010). *Technology and journalism-aslabour: Historical perspectives*, «Journalism», 11, 1, p. 57-74.

Pasquale, F. (2015), *The blackbox society*. Cambridge: Harvard University Press.

Pavlik, J. (2000). *The impact of technology on journalism*, «Journalism Studies», 1, 2, pp. 229-237.

Porlezza, C. (2016). *Dall'open journalism all'open government? Il ruolo del data journalism nella trasparenza e nella partecipazione*. «Problemi dell'informazione», 1, pp. 167-194.

Porlezza, C., Splendore, S., Pikone, I. and Bradshaw, P. (2018), *Closed Data! European Data Journalists' Strategies to Access and Analyze Data*,

Presentation at the Nordic Datajournalism Conference, Södertörn University, Stockholm, 15.03.2018.

Powers, M. (2012). *In forms that are familiar and yet to be invented. American journalism and the discourse of technologically specific work*, «Journal of Communication Enquiry», 36, 1, pp. 24-43.

Ramge, T., & Mayer-Schönberger, V. (2017). *Das Digital. Markt, Wertschöpfung und Gerechtigkeit im Datenkapitalismus*. Berlin: Ullstein.

Russell, A. (2016), *Journalism as activism. Recoding media power*. Cambridge: Polity.

Ryfe, D. (2012), *Can journalism survive? An inside look at American newsrooms*. London: Polity.

Sampedro, V. (2014), *El cuarto poder en red. Por un periodismo (de código) libre*. Barcelona: Icaria.

Schwarz, A. (2004), *The numbers game. Baseball's lifelong fascination with statistics*. New York: Thomas Dunne Books.

Spitz, M. (2017). *Daten – das Öl des 21. Jahrhunderts*. Hamburg: Hoffmann und Campe

Splendore, S. (2016). *Closed data! Il giornalismo italiano alla ricerca di dati*. «Problemi dell'Informazione», 1, pp. 195-214.

Stark, J. and Diakopoulos, N. (2016). *Towards editorial transparency in computational journalism*. Computation + Journalism Symposium, Palo Alto, 30.10.2016. Available at: https://pdfs.semanticscholar.org/9cc2/0edfbac46e953fb307f91da7df94543fc973.pdf

Steiner, C. (2012), *Automate this: How algorithms took over our markets, our jobs, and the world*. New York, NY: Portfolio.

Thorsen, E. (2017), *Cryptic Journalism*, «Digital Journalism», 5, 3, pp. 299-317.

Thurman, N. (2017), *Social Media, Surveillance, and News Work. On the apps promising journalists a "crystal ball"*, «Digital Journalism», 6, 1, pp. 76-97.

Uricchio, W. (2017). *Data, culture, and the ambivalence of algorithms*. In M. T. Schäfer, & K. van Es (eds.), *The datafied society* (pp. 125-137). Amsterdam: Amsterdam University Press.

Usher, N. (2016), *Interactive Journalism. Hackers, Data and Code.* Urbana: University of Illinois Press.

Vaccarai, C., Valeriani, A., Barberà, P., Jost, J., Nagler, J. and Tucker, J. (2016), Of echo chambers and contrarian clubs: Exposure to political disagreement among German and Italian Users of Twitter. «Social Media + Society», 2, 3, online first.

van Dijck, J. (2014). *Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology*. «Surveillance & Society», 12, 2, pp. 197-208.

Wahl-Jorgensen, K., Williams, A., Sambrook, R., Harris, J., Garcia-Blanco, I. and Dencik, L. (2016), *The future of journalism. Risks, threats and opportunities*. «Digital Journalism», 4, 7, pp. 809-815.

Waters, S. (2017), *The effects of mass surveillance on journalists' relations with confidential sources*. «Digital Journalism», published online.

Wölker, A. and Powell, T. (2018), *Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism*, «Journalism», online first.