Narrating the Stories of Leaked Data: The Changing Role of Journalists after WikiLeaks and Snowden

Traditionally, investigative journalists had a gatekeeping role between their confidential sources of information and the public sphere. Over the last two decades and with the arrival of new media, this role has been undergoing changes. Recent cases of whistleblowing, such as WikiLeaks and Snowden, illustrate how contemporary media allow individuals to release data directly to the global audience. This raises the question of how recent leaks affect how journalists operate.

In this study we compare how *The Guardian* covered two cases of whistleblowing which are commonly referred to as WikiLeaks and Snowden. We analyse how access to leaked data is provided or facilitated on *The Guardian* website, how readers are invited to interact with these data and how journalists present their own activities. A qualitative analysis of the leading articles further shows how the stories are framed and how much prominence is given to the data and the various actors.

The results show how the roles of journalists shift from gatekeeping to data management, interpretation, contextualisation and narration. Journalists may no longer be needed to publish leaked data but they are still needed to tell the stories of leaked data.

Keywords: mass media, whistleblowing, journalism, media discourse

1. Introduction

In June 2013, a major leak by whistleblower Edward Snowden revealed that most communications over digital networks were accessible to the US secret services. The surveillance was carried out through access to large scale information technologies controlled by multinational corporations (e.g. Microsoft, Google, Facebook, Apple), which are beyond the control of localized jurisdiction. It soon became clear to the public that the US, who had accused other countries of spying and hacking their computers, had been using the central position it holds for having created the internet to its own advantage, turning this global communication infrastructure into a sort of global panopticon (Sullivan, 2014; Zuboff, 2015). This revelation sparked an unprecedented international debate about digital surveillance in contemporary societies. It also led to a range of political consequences, including tensions between the US and other countries. Not least, it raised fundamental questions concerning the roles of journalists in such emerging societal and power configurations. Traditionally, journalists used to have the

role of gatekeepers who control how much information from their sources is passed on to the general public. This role is challenged today, when individuals can leak information online without relying on journalists as intermediaries. This leads us to our research question, which is: How do contemporary forms of online whistleblowing change the role of journalists as intermediaries between data and the general public?

   Against this broad background, we present and discuss the case of how one newspaper, the British *The Guardian*, covered part of these revelations. *The Guardian* had an active role in both the WikiLeaks and the Snowden cases by being granted advance access to the leaked data prior to their public release. In addition, *The Guardian* had already taken a leading role in establishing new models of data journalism before these two whistleblowing cases, for instance by launching their data blog in 2009 (Rogers, 2013).[1] This makes *The Guardian* a perfect focus for studying how new modes of releasing leaked data go hand in hand with new roles for journalists and new forms of reporting. We will study these innovations through a qualitative analysis of the content and function of the articles that were published on *The Guardian* on the two stories.

2. Mass data: whose stories?

2.1 Open participation and media bias

   It has long been acknowledged that objectivity in news reporting is an unattainable ideal. On the macro-textual level, the selection, omission and framing of

---

[1]   The first big story on *The Guardian* data blog was launched on 31 March 2009 and dealt with MPs' expenses. See
http://www.theguardian.com/news/datablog/2009/mar/31/mps-expenses-jacquismith (last retrieved on 14 June 2016)

news events is driven by the aim to maximise the news values of a story, which can lead to bias (Galtung and Ruge, 1973); Cohen and Young (1973) even use the expression "the manufacture of news" in the title of their collected volume. In a similar vein, Bell (1991: 147) makes the point that "[j]ournalists do not write articles. They write stories." He calls journalists "professional story-tellers of our age (1991: 147) and points out the structural similarities between personal narratives and news stories (1991: 147–155). News values determine the structure and content of news stories and thus help journalists to tell their stories in a way that appeals to the audience (Bell, 1991: 155). However, if the need for audience appeal is taken too far, it can lead to misreporting and distortion of facts (Bell, 1991: 216). At a micro-textual level, various linguistic devices have been identified that can create bias in news texts (e.g. Fairclough, 1988, 1995; Floyd, 2000; Fowler, 1991; Locher and Wortham, 1994; Stenvall, 2008, 2014; Wortham and Locher, 1996). In addition, it has been pointed out how the practices of text production are closely interwoven with organizational structures and economic interests (e.g. Czarniawska, 2011; van Dijk, 2008, 2009).

Citizen journalism maintains that open participation rather than professional journalism would rebalance the bias of large media by watchdogging the elites (Allan, 2013). Recent years have seen an increase of user-generated content in journalism and lay people's data production has been challenging current forms of journalism (Boczkowski, 2004; Boczkowski and Mitchelstein, 2013; Bruns, 2005, 2016; Conboy, 2004; Landert, 2014a, 2014b; Lewis, 2003; Newman, Dutton, and Blank, 2012; Ostertag and Tuchman, 2012; Papacharissi, 2009; Wardle, Dubberley, and Brown, 2014; Wardle, 2016; Wardle

and Williams, 2008). Indeed, nowadays moderating and editing content produced online by 'crowds' has become a central part of the work of journalists: user-generated content contributes significantly, directly and indirectly, to influential news publications. For instance, by the time reporters arrive at war or disaster sites, plenty of information and pictures are already available by those directly affected (Allan, 2013). Dutton (2009) argues that internet-based communication allows the consolidation of a 'fifth estate', i.e. bloggers, social media and online reporters, as distinct from the fourth estate (which refers to the press and mainstream media in general) and counterbalances its inequalities. Resonating with the ideal of an open cyberspace confronting large conglomerates that WikiLeaks in 2006-2013[2] appeared to have revived, Brevini et al.'s (2013a) volume emphasizes the prospects of transparency and free flow of information in contrast to secrecy and dominance of the few. Their tone is well-exemplified in the opening of the book: "Transparency and open access to information are the only real pressures on governments to remain true democracies." (2013: xvi). This enthusiasm for openness and democratization for every niche of society that the arrival of the World Wide Web promised (Poster, 1995; De Kerckhove, 1997) has since been curbed by revelations of global surveillance.

2.2 Whistleblowers and journalists

---

[2] We refer to this period because later developments of the WikiLeaks case, especially concerning the US presidential campaign in 2016, showed how its openness may have been exploited by Russian government to influence the US electorate.

The fourth estate has always relied on non-journalist informants; anonymous sources have always been central for investigative journalism. Whistleblowers have often found support and resonance on the press and news media. So, if an open democratic public sphere remains chimerical, and if investigative journalism has always been part of the fourth estate (Benkler, 2011), what new can we learn from the recent wave of whistleblowing? The traditional role of journalists used to be gatekeeping. They were in charge of and responsible for deciding what information to make accessible to the public and how. Contemporary whistleblowers can engage in a different way with the public opinion by making information public and taking the frontstage, without relying on journalists as intermediaries. With contemporary media, access to information is faster and less restricted to the extent that professional journalists are constantly challenged by competing sources. In this context, the recent cases of whistleblowing have transformed the established balance between frontstage and backstage maintained by journalists (see also Flew and Liu, 2011).

The relation between WikiLeaks and traditional media journalists is characterised by ambivalent attitudes. On the one hand, the relation is mutually beneficial (see Dunn, 2013). For journalists, the documents that are leaked on WikiLeaks can provide valuable material for news stories that otherwise would remain untold. At the same time, for WikiLeaks, the coverage of the documents in traditional media is crucial for achieving political impact. Without reports in major newspapers, most of the documents published on WikiLeaks would remain unnoticed by the general public, which is why WikiLeaks actively seeks the attention of journalists (Dunn, 2013; Lynch, 2010: 311). In addition,

WikiLeaks makes use of the infrastructure of established media organisations for the analysis of the raw data they receive (Brevini and Murdock, 2013: 49; Lynch, 2013). On the other hand, the ethos adopted by WikiLeaks stands in contrast to the established principles of investigative reporting in a number of crucial points. Ethical concerns that have been raised include the limited options of independent verification of the information (Lynch, 2010: 314) as well as the lack of redaction of leaked documents, which has been argued to have endangered lives in some cases (see Benkler, 2013: 24). This leads to various tensions between journalists and WikiLeaks.

Lynch (2010) describes these ambivalent attitudes during the early years of WikiLeaks. Her study is based on public reports by members of the WikiLeaks collective and on a survey among reporters. Members of the WikiLeaks collective are described as being frustrated with the perceived lack of press response to certain leaks and the fact that mainstream journalists do not grant enough authority to their analyses (Lynch, 2010: 312). One point that is of particular relevance to the present study is the claim that some of the documents received little attention from journalists because they were difficult to understand (Lynch 2010: 312). At the same time, the attitudes of journalists towards WikiLeaks varied greatly in Lynch's survey. Some of the journalists used the site regularly or at least occasionally as a valuable source for news stories, while others said that they had only come across the site during an ongoing investigation of a story. One of the greatest benefits journalists saw in the site was its use as a repository for leaked documents, especially in cases in which journalists come under legal pressure to keep them from publishing leaked information (Lynch, 2010: 315-316). Overall, Lynch

concludes that "Wikileaks has been only partly successful at appearing credible and newsworthy in journalists' eyes" (2010: 315).

Lynch's (2010) study reports on the early stages of the interaction between WikiLeaks and traditional journalism, during a time in which the general public had relatively little awareness of the platform. This changed in 2010, when WikiLeaks gained worldwide attention with the publication of a series of leaks, including the US army video known as Collateral Murder, the Afghan War Logs, the Iraq War Logs and the US diplomatic cables that led to "Cablegate" (Brevini et al., 2013b: 2-3). Together with the increase in attention, the controversies around WikiLeaks also gained momentum. If anything, the tensions between WikiLeaks and traditional media became stronger. Benkler (2013) traces some of these tensions in more detail. He points out that the attitude towards WikiLeaks varies across media organisations. While some, like *The Guardian*, try to maintain a partnership others, like *The New York Times*, are more critical towards WikiLeaks (Benkler, 2013: 23). Among the factors influencing the relationship Benkler (2013: 23) mentions different legal implications of a perceived partnership, different attitudes towards new networked models of journalism, as well as personal animosity between Assange and newspaper editors. Benkler (2013) also notes the emergence of a new model of watchdog function, which is neither purely networked nor purely traditional (see also Dreyfus et al., 2011). Instead it appears to originate from a mutual interaction between these two modes of producing and conveying news to society.

Most of the research literature dealing with the relation between WikiLeaks and traditional journalists focuses on ethical and legal questions (see, for instance, the papers

in Brevini et al., 2013a, especially Elliott, 2013). How much redaction and editing is needed before publishing leaked documents? What steps need to be taken to protect (innocent) individuals mentioned in leaked files? Can one justify the publication of information from an anonymous source whose identity cannot be verified? How can journalists mediate the ideal of journalistic objectivity and political motives behind leaks? And how can legal restrictions concerning the publication of leaked material be circumvented? In addition, Eldridge (2016) explores metadiscourses of journalistic identity in the coverage of WikiLeaks and Snowden.

In contrast, our study focuses more closely on news writing. We ask how different practices of managing and publishing leaked data affect the content and structure of news articles, and, as a consequence, how news stories are perceived by the readers. One point of departure is the observation mentioned in Lynch's (2010: 312) study that some of the leaked documents are not easy to understand. This is not surprising, given that leaked material tends to come from highly specialised and often very technical domains, like political and military reports. However, in order to have political impact, the information in the documents needs to reach an audience that is as large as possible. There are marked differences in the attitudes of how this should be realised between WikiLeaks and traditional investigative journalism. WikiLeaks is "[r]ooted in hacktivism and in ethics of radical transparency" (Brevini et al., 2013b: 4), which is reflected in the primary aim of maximising public access to data. This means that the WikiLeaks site provides access to large amounts of data with very little contextualisation or interpretation. In other words, WikiLeaks' attitude is largely one of letting data speak for themselves. In contrast,

investigative journalism gives considerably more room to the story behind data. Data are certainly crucial to support and verify the story, but the news story is much more than just a collection of factual information. The selection, presentation and contextualisation of the information for the reader are core tasks of the investigative journalist. Lovink and Riemens (2013: 248) aptly summarise the different attitudes as follows: "Traditional investigative journalism used to consist of three phases: unearthing facts, cross-checking these, and backgrounding them into an understandable discourse. WikiLeaks does the first, claims to do the second, but omits the third completely."

In Sections 3 and 4, we present two case studies that illustrate how the tensions between these different attitudes can be observed within news articles. In the first case, the Afghan War Logs, the WikiLeaks approach of letting data speak for themselves is dominant, while in the second case, PRISM, there are considerably more characteristics typical of traditional investigative journalism. We regard both cases as illustrations of ongoing attempts by journalists of developing novel approaches of dealing with new forms of leaking data.

2.3 The two cases

The two cases considered here are usually referred to as WikiLeaks and Snowden. "As the prefix 'wiki' suggests, Assange originally envisaged WikiLeaks as a demonstration of the principle of open collective production in practice, providing a space in which users could build on the base information by adding their own material and annotations." (Brevini and Murdock, 2013: 47). After years of activities, WikiLeaks

suddenly became a household name when it published thousands of internal US military documents that presented unknown sides of the Afghan war. This six-year archive of classified military documents leaked by Chelsea Manning was published by WikiLeaks initially in collaboration with *The New York Times*, *The Guardian*, and *Der Spiegel* (Dunn, 2013; Lynch, 2013). The whistleblower in the second case considered here is Edward Snowden, a NSA contractor who copied a comparable number of documents and passed them on to trusted investigative journalists of *The Guardian*.

As mentioned above, these two cases engender two approaches to information management, one more closely derived from the original culture of the internet, the other more sensitive to the broadly legitimized function of journalism. The former puts paramount emphasis on data to be openly shared for everyone to interpret it and make own judgments. This approach, sometimes labeled "scientific journalism" since it advocates for full disclosure of the sources, was championed by Assange and fueled data journalism by providing it with new tools (Allan, 2013). The latter instead informed Snowden's actions, whose leaks were handled by experienced investigative journalists. The contrastive analysis of the two cases shows the role of narratives as sense-making devices that established institutions like the press and free speech continue to rely upon, although in novel ways.

Brevini et al. describe WikiLeaks with the following words: "Rooted in hacktivism and in ethics of radical transparency, exploiting technological expertise and opportunities, and carrying the "wiki" concept of open publishing and collaborative work in its name, WikiLeaks connects with both an alternative countercultural and a digital

citizen media model" (2013b: 4). In accordance with its radical transparency orientation, having received documents leaked by Manning, WikiLeaks aimed at achieving maximum visibility for those materials. In a first phase Julian Assange, founder of WikiLeaks, engaged in partnerships with established media. Due to dissatisfaction with the slow pace of publication by his media partners, which was a consequence of their focus on careful redaction of the documents, he made the unrefined materials publicly available (Dunn, 2013; Elliott, 2013). This allowed anyone to see the data and contribute to their evaluation. This echoes the so-called Linus's Law about software development, according to which 'given enough eyeballs, all bugs are shallow'[3]. Underneath one can spot the belief in what Surowiecki (2005) termed the 'wisdom of the crowds' in his book of the same title, according to which an open-ended set of people can outsmart any individual or organization.

Three years later, Edward Snowden accorded upfront a prominent role to traditional investigative journalism to focus, select and publish data. He made his first contact with *The Guardian* journalist Glenn Greenwald in December 2012 under the pseudonym of Cincinnatus, an ancient Roman statesman and farmer who was made dictator to solve a crisis and resigned two weeks later, after resolving it (Greenwald, 2014). Things started to come together in April 2013 with the help of journalist and film-maker Laura Poitras and support from *The Guardian*. In a move that reminds of Cold War spy stories, Snowden flew from Hawaii to Hong Kong (China, but tied to the West

---

[3] More formally: Given a large enough beta-tester and co-developer base, almost every problem will be characterized quickly and the fix will be obvious to someone.

and its values of freedom) where he met trusted journalists. Shortly after the first release, a few video appearances of Snowden gave the public an image of him as a rational and trustworthy source, rather than the insane person that the NSA would have tried to depict to discredit the whole operation.

Of course, the media resonance of those cases exceeds by far what can be captured here. Thus, we focus on how one newspaper only, *The Guardian* – which played a major role in both cases – reported online on two datasets of the many they received: The Afghan War Logs from WikiLeaks and PRISM from Snowden. The Afghan War Logs were the first large batch of data from Manning that was released on WikiLeaks. Previous to this, only selected individual files were released, including a diplomatic cable known as Reykjavik13 and a video about an airstrike in Baghdad. Unlike the earlier files, the War Logs were released by WikiLeaks in collaboration with media partners, namely *The Guardian, The New York Times* and *Der Spiegel*, who had advance access to the data before their release.

PRISM was the first big media revelation that resulted from Snowden's data. The data were given to *The Guardian* and *The Washington Post*, who selectively published them together with their stories. In contrast to the War Logs, the data from Snowden were not (initially) published elsewhere on the internet. In the following two sections, we will analyse the articles that were published on the online news site of *The Guardian* when the two stories were revealed.

3. Coverage of the Afghan War Logs

On 25 July 2010, the Guardian published a selection of about three hundred files[4] from the War Logs on their website. Each file was presented on a separate page as a standardised report in a format similar to the reports published on WikiLeaks. The files were handpicked by Guardian analysts to represent a set of "significant incidents" (warlogs-01).[5] A separate article provides further explanations about the data selection and the motivations behind it (warlogs-02). One point that is emphasized is that *The Guardian* did not want to include sensitive information, such as the names of informants and information that might put the Nato troops in danger (warlogs-02):

> It was central to what we would do quite early on that we would not publish the full database. Wikileaks was already going to do that and we wanted to make sure that *we didn't reveal the names of informants or unnecessarily endanger Nato troops*. At the same time, we needed to *make the data easier to use* for our team of investigative reporters […] We also wanted to make it *simpler to access key information for you*, out there in the real world – as clear and open as we could make it. (warlogs-02, *italics* added)

This quote illustrates well the tensions involved in publishing these data. On the one hand, there is the aim of transparency by disclosing the data to the general public. This is

---

[4] The number of selected files is given as 300 in several of *The Guardian* articles (e.g. warlogs-01, warlogs-02). The Excel file that can be downloaded from *The Guardian* website (warlogs-05) contains 310 entries.

[5] *The Guardian* articles on which we base our analysis are referred to with abbreviations in the format warlogs-XX and prism-XX. The complete references of the articles can be found in the Data section at the end of the paper.

motivated by the idea that the public has a right to know about the numerous problems that occurred during the military operation, such as friendly fires and civilian casualties. On the other hand, publishing the data could endanger lives. *The Guardian* decided to be very restrictive in what they published on their website, but at the same time their reports about the leak helped publicise the data that were accessible on WikiLeaks, which were much more comprehensive and not redacted to the same extent.

The above quote also introduces two additional aims of *The Guardian*. First, they used the data to develop their own stories in the tradition of investigative journalism. On the day on which the War Log files were published, *The Guardian* published twelve articles in which they investigated various topics based on the data included in the files. For instance, one of the articles deals with shootings and bombings of civilians by British troops (warlogs-03). The article refers to several of the original War Log files that were published on *The Guardian* and the files are accessible from the text through hyperlinks. In addition, there is a separate page that lists all the relevant files on which the article is based in chronological order (warlogs-04). This is a good example of how the data were used by *The Guardian* to present readers with direct access to the evidence on which their stories are based.

The second aim is to let readers investigate the data, not only to verify *The Guardian*'s stories, but also to carry out their own analyses. Several of the articles that were published on *The Guardian* website instruct readers how to work with the data. The headlines of three such articles are given in (1) to (3).

(1) "Wikileaks Afghanistan files: download the incidents as a spreadsheet" (warlogs-05)

(2) "Afghanistan war logs: the glossary" (warlogs-06)

(3) "How to read the Afghanistan war logs: video tutorial" (warlogs-07)

The headline given in (1) comes from a page where readers can download an Excel file with a list of the 310 key events that were selected for publication by *The Guardian* journalists. The Excel file contains the same information that is included in the individual reports, but the format allows users to filter and sort the events according to date, geographical location, type of event, number of victims, etc. The page from which the file can be downloaded provides further information about the data, for instance pointing out that the data referring to victims are "highly unreliable" (warlogs-05). The glossary referred to in the second headline explains military abbreviations and acronyms that are used throughout the files. This glossary is indispensable for readers who want to investigate the data for themselves, since it is impossible to understand the files without understanding the highly specialized abbreviations. That the raw data are not easy to work with can be seen from the fact that *The Guardian* published a video tutorial in which they explain to readers how to read and interpret the data (warlogs-07). In addition to giving an overview about all the resources that are available on *The Guardian* website, the tutorial takes readers through one of the reports line by line, explaining each piece of information. A screenshot of the video is presented in Figure 1.
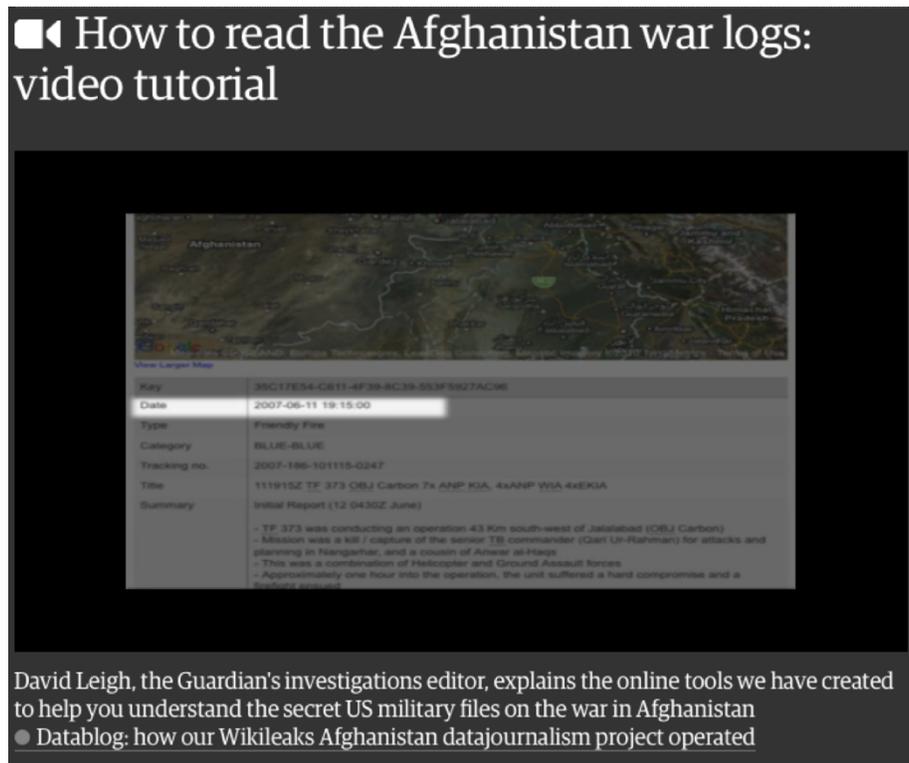
Figure 1: Screenshot from video tutorial (warlogs-07)

The publication of the War Log files was accompanied by several additional articles, including an opinion piece, an editorial, two background articles on Assange, a timeline of events after the publication (warlogs-08), and an article that the timeline article refers to as "the main story". This main article (warlogs-main) provides further insight into the framing of the news event by *The Guardian*. There is quite a strong focus on the leaking of data and on the leaked files, which is reflected in the article's title, reproduced in (4).

(4) Afghanistan war logs: Massive leak of secret files exposes truth of occupation

(warlogs-main)

Roughly half of the article deals with the leak as such, detailing the amount and nature of the leaked data, their publication and reactions to the leak. The other half of the article gives some examples of the kinds of problems that are revealed in the files. These examples are clearly presented as being based on the leaked data and include quotes from and links to the files, as illustrated in passage (5).

(5) Questionable shootings of civilians by UK troops also figure. The US compilers detail an unusual cluster of four British shootings in Kabul in the space of barely a month, in October/November 2007, culminating in the death of the son of an Afghan general. Of one shooting, they wrote: "Investigation controlled by the British. We are not able to get [sic] complete story." (warlogs-main)

Overall, the main article focuses at least as much on the leaked data as on the events in Afghanistan.

The Guardian refers to their own approach in covering the story as "datajournalism" (warlogs-02, warlogs-09). This term captures well the picture that emerges from analysing the published articles. Instead of telling one coherent story about the problematic events in Afghanistan, *The Guardian* journalists present themselves as data analysts and facilitators of reader investigations. They focus on explaining the data,

instructing readers how to study them, and providing tools for data analysis and visualisation. Their own analyses of the data are presented as examples that can be followed by readers by accessing the data for themselves. A very explicit declaration of this can be found at the end of one of the articles, given in extract (6).

(6) Have we published enough? Inevitably not. Have we started to make sense of an incredibly complex dataset? We hope so.

Now it's your turn. Can you help us make more sense of the raw info? (warlogs-02)

By explicitly asking readers to submit their own evaluations, *The Guardian* tries to leverage a sort of citizen journalism and make use of the 'wisdom of the crowds'. However, there is little evidence of reader participation. A later article, published in November 2010, presents six visualisations submitted by readers, which is a rather modest outcome, and there is no indication that these visualisations have led to new insight into the data.

4. Analysis of PRISM coverage

The story about the PRISM program broke three years after the Afghan War Logs, on 6 June 2013 (prism-main). The reporting on the two cases differs in several

respects. The first difference concerns the pace of publication (see Figure 2).[6] When covering the Afghan War Logs in 2010, *The Guardian* published 65 articles within only three days. After that, the coverage consisted of individual articles only. Overall the overview page lists 95 articles (not including the published files from the leaked data), the last of which was published on 18 April 2014. In contrast, the overview page of the PRISM story lists 172 articles. Compared to the War Logs coverage, these articles are more evenly distributed over a longer period. In the first three days, between 6 June and 9 June 2013, a total of 10 articles were published. The peak in frequency came on 10 June 2013 (day 5) with 15 articles and 11 June (day 6) with 9 articles. After this, several articles were devoted to the story almost every day until the end of June (day 25), and between 6 and 22 articles every month until the end of the year (month 7).

---

[6] The numbers are based on the articles that are listed on the overview page of each topic, warlogs-op and prism-op (not including the 300 war logs reports from the leaked data). While this includes all articles that are tagged as "The war logs" and "Prism", there may be a few articles that relate to these stories but that are missing from the overview pages. Nevertheless, the articles listed there provide a good basis for comparing the coverage of the two stories.
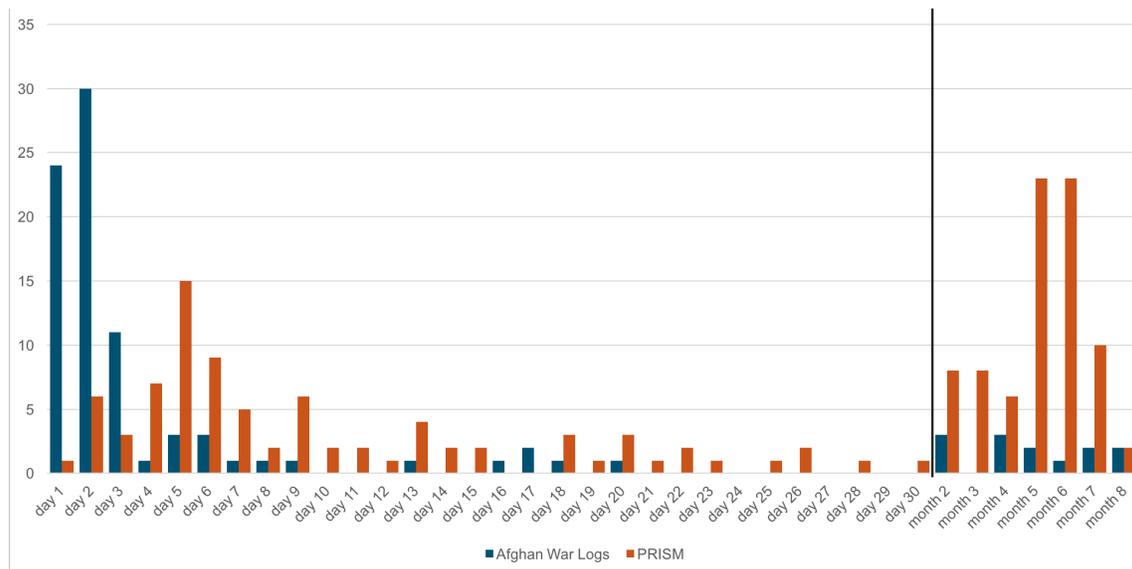
Figure 2: Number of articles published after the initial release of the story

To some extent at least, the higher number of articles after the initial release was due to a more active public and political debate on the PRISM program. The fact that the PRISM program affected UK residents – the main target readership of *The Guardian* – more directly than the Afghan War Logs certainly plays a role, too. However, there is one factor that relates to the mode of data release, namely that the leaked data were published very selectively and in small chunks. The original PRISM story is based on a 41-slide PowerPoint presentation, of which only 3 slides were published with the main article (prism-main). Compared to the War Logs, where hundreds of files provided material for dozens of stories, the PRISM slides supported only one single story. However, it is important to emphasise that this difference is not simply a difference in the amount and

21

nature of data acquired by the whistleblowers. Snowden is assumed to have been in possession of hundreds of thousands of intelligence files; but instead of being publicly published in large batches, the files were carefully selected and redacted by journalists, then published individually over a longer period of time, and each release was accompanied by its own story. In this way, the PRISM story was kept alive while related stories were released over the following months. These included, for instance, Britain's spy agency program Tempora (21 June 2013, prism-01), the XKeyscore program (31 July 2013, prism-02), an NSA datamining operation targeting the Indian embassy (25 September 2013, prism-03), and the monitoring of phone conversations of world leaders (25 October 2013, prism-04). At the end of November 2013, Greenwald talked about the stage of reporting in an interview, saying:

> [...] if I had to guess, we are still in the first part, the first half of the reporting. The majority of reporting on these documents for sure is reporting that has yet to be done but that will be done. (prism-05)

This statement clearly shows the intention of using the leaked documents to support reporting over a longer time frame.

The mode of data release is also responsible for the fact that there are no attempts to instruct readers on how to carry out their own investigations. The data that are made accessible to readers are already thoroughly analyzed and interpreted, and there is no explicit opening for further reader investigations.

Further differences between the reporting on the PRISM program and the War Logs can be found in how the story is presented in the main article (prism-main). In contrast to the releasing article on the War Logs case, there is very little focus on the data leak. Instead, the article focuses almost exclusively on the facts that are revealed about the PRISM program and on reactions by political actors and by the companies that are affected by the program. The article headline, given in (7), makes no mention at all of the data leak.

(7) NSA Prism program taps in to user data of Apple, Google and others (prism-main)

The leaked data are first mentioned at the end of the first paragraph, but without explicitly referring to them as "leaked". Instead, they are used in the attribution "according to a secret document obtained by *The Guardian*", where the term "secret" implies that *The Guardian* is in contact with a confidential source. The most explicit discussion of the data appears in the third paragraph, which is reproduced in (8).

(8) *The Guardian* has verified the authenticity of the document, a 41-slide PowerPoint presentation – classified as top secret with no distribution to foreign allies – which was apparently used to train intelligence operatives on the capabilities of the program. The document claims "collection from the servers" of major US service providers. (prism-main)

23

There is no detailed information in this passage or elsewhere in the article about how *The Guardian* obtained the data. However, the term "leak" is used twice in later parts of the article as illustrated in examples (9) and (10).

(9) Disclosure of the Prism program follows a *leak* to *The Guardian* on Wednesday of a top-secret court order compelling telecoms [sic] provider Verizon to turn over the telephone records of millions of US customers. (prism-main, our emphasis)

(10) The document is recent, dating to April 2013. Such a *leak* is extremely rare in the history of the NSA, which prides itself on maintaining a high level of secrecy. (prism-main, our emphasis)

Of these two passages, only the second uses the term "leak" in relation to the PRISM slides. In (9), the term "leak" refers to data about a related disclosure on which *The Guardian* reported earlier. The article does not reveal that both datasets are part of the same leak, namely Snowden's files, even though the formulation suggests some sort of connection. Snowden's identity as "the whistleblower behind the NSA leaks" is only revealed three days later, on 9 June 2013 (prism-06).

Overall, *The Guardian*'s reporting on PRISM is much more representative of traditional investigative journalism than of the data journalism approach that was adopted in covering the War Logs case.

5. Discussion and conclusion

Even though the motifs behind the two whistleblowing events are similar, their performances differ substantially. Comparing the cases, similarities between the two are quite evident: both leaked data about US international operations and they both heavily rely on new media in their actions. The size of the leaks is comparable and both became global media events. On the other hand, some significant differences are worth discussing here. Manning sent the files he accessed to WikiLeaks, who advocates for no curation of materials. So, WikiLeaks left the 'analytical labor' to make sense of data on the shoulders of the newspapers it partnered with initially, and ultimately on readers (Allan, 2013; Brevini et al., 2013b). Snowden chose carefully a different intermediary: investigative journalists with a credible record of supporting causes of public interest, especially regarding civil rights and state abuses of power.

WikiLeaks' communication relied on the assumption that data should be left speaking for themselves, therefore no one should edit them before they are publicly released for everyone to make up their minds, including when accompanying journalistic stories. Snowden instead relied on journalists to scrutinize data about their veracity (also double-checking with relevant government bodies) and relevance, then embed those leaks into narratives that the public could relate to. This approach to whistleblowing relates to

narratives and sense-making: "the narrative mode of knowing consists in organizing one's experience around the intentionality of human action. The plot is the basic means by which specific events, otherwise represented as lists or chronicles, are put into one meaningful whole" (Czarniawska, 1999: 14). Even if the events documented by data did not happen in dramatized forms, narratives are used as sense-making devices to the extent that they create an arch or tension towards a meaningful interpretation or prospect (see also Bell, 1991: 147). Journalists keep making use of this power of narratives.

As consequence of those distinct communication strategies, the two cases considered above spurred quite different reactions. WikiLeaks radical openness clashed with established social norms of news communication. On the other hand, Snowden's leaks prompted reactions from all levels, including a US Presidential speech on 17 January 2014 responding with an NSA reform to the accusations of an out of control surveillance state.

We started with the research question of "How do contemporary forms of online whistleblowing change the role of journalists as intermediaries between data and the general public?" In a nutshell, the answer is that journalists may no longer be needed to publish leaked data and guard the distinction between stage and backstage, but they are still needed to accompany the readers and to tell the stories of leaked data. Large datasets may simply be too complex, rich and multifaceted, therefore overwhelming and dismissed or misinterpreted by the readers. Without journalists, the message risks not to be effective. Indeed, our analysis shows how the roles of journalists shifted very strongly from gatekeeping to data management in the first case. Then a more active role in

interpretation, contextualisation and narration was recovered in the second case. In other words, editing, redacting and curating data can be seen as sanitizing data or, alternatively, as responsible journalism. These contrastive positions proved to have remarkable consequences on the public sphere and opinion, thus onto politics.

More recently, new, even bigger leaks were revealed. Referred to by *The Guardian* as "the biggest data leak in history", (Harding, 2016: title), the Panama Papers include 11.5 million files and a total of 2.6 terabytes of information – more than 1,500 times the amount of data that was leaked to WikiLeaks by Manning (Harding, 2016: paragraph 13, graphic 2). The identity of the whistleblower has not been revealed so far, but it seems clear that Snowden's mode of leaking data served as an example. The whistleblower contacted the German newspaper *Süddeutsche Zeitung* and made the data accessible to them. Due to the large amount of data, the newspaper collaborated with the International Consortium for Investigative Journalists to evaluate the documents. According to the *Süddeutsche Zeitung*, 400 journalists from 80 different countries – including journalists from *The Guardian* – investigated the documents for an entire year before making the leak public (Obermayer et al., 2016: paragraph 4). Despite the larger scale, the mode of operation shows strong parallels to how Snowden's data were handled and there is sufficient reason to speculate that a successful model has been established by now that will be used by whistleblowers and journalists in years to come.

Finally, it is worth mentioning that WikiLeaks and its non-curation policy have demonstrated that open participation does not necessarily counterbalance bias, but rather can be exploited. WikiLeaks has played a central role during the US presidential election

campaign in 2016 due to the publication of emails of prominent figures of the US
Democratic party. Allegedly, those emails were elicited by hackers close to the Kremlin
in an attempt to influence the presidential elections and leaked to WikiLeaks, who
published them without restrictions.

These conclusions can shed some novel light on contemporary journalism. While
we see little doubt for data journalism to be here to stay, we certainly see compelling
limitations to claims that supplying people with data inevitably leads to a re-
democratizion of the public sphere (Johansson, Lehti and Kallio, this issue). Open
information infrastructures allow motivated skillful individuals to access and use data as
never before, but without compelling storytelling, the broader social resonance seems
severely limited. This position echoes Joerges (1999), according to whom narratives are
the tools for politics. Indeed, established and widely legitimized strong narratives like
patriotism, democracy, and privacy vs. surveillance supported Snowden's reaffirmation
of free speech and investigative journalism across hugely diverse social settings.

An interesting contradiction can be noted: WikiLeaks relied on the wisdom of the
crowds, but when the huge crowd of global general public was addressed, this alleged
internet wisdom did not scale to this new level. Rather, the diversity of views created
confusion rather than better understanding of the situation. In sum, our comparison
suggests that it has been more effective to say 'we defend transparency' than to perform
transparency by releasing to the general public large, unstructured and unedited bulks of
secret documents. This is not to say that WikiLeaks' actions should be dismissed as not
influential. Firstly, they created the conditions for more effective revelations later on.

Secondly, the lack of a clear-cut narrative does not mean that many narratives cannot be derived from large and unstructured datasets. However, the open access to information granted by communication technologies does not appear to be shaping the public sphere towards a more democratic rationality (Johansson and Lehti and Kallio, this issue). Rather, it is storytelling that keeps journalism and public opinion going.

Acknowledgements

Data

War Logs	*Article* (title given as in the original)
Main article	
warlogs-main	Afghanistan war logs: Massive leak of secret files exposes truth of occupation. 25 July 2010. http://www.theguardian.com/world/2010/jul/25/afghanistan-war-logs-

	military-leaks

Overview page	

warlogs-op	Afghanistan: the war logs. http://www.theguardian.com/world/the-war-logs

Other articles	

warlogs-01	Afghanistan war logs: our selection of significant incidents. 25 July 2010. http://www.theguardian.com/world/datablog/interactive/2010/jul/25/afghanistan-war-logs-events
warlogs-02	Wikileaks' Afghanistan war logs: how our datajournalism operation worked. 27 July 2010. http://www.theguardian.com/news/datablog/2010/jul/27/wikileaks-afghanistan-data-datajournalism
warlogs-03	Afghanistan war logs: Civilians caught in firing line of British troops. 25 July 2010. http://www.theguardian.com/world/2010/jul/25/british-troops-afghan-civilian-shootings
warlogs-04	Afghanistan war logs: List of civilian shootings by British troops. 25 July 2010. http://www.theguardian.com/world/2010/jul/25/british-shootings-afghan-civilians-list
warlogs-05	Wikileaks Afghanistan files: download the key incidents as a spreadsheet. 25 July 2010. http://www.theguardian.com/world/datablog/2010/jul/25/wikileaks-afghanistan-data#data
warlogs-06	Afghanistan war logs: the glossary. 25 July 2010. http://www.theguardian.com/world/datablog/2010/jul/25/wikileaks-afghanistan-war-

	logs-glossary
warlogs-07	How to read the Afghanistan war logs: video tutorial. 25 July 2010. http://www.theguardian.com/world/datablog/video/2010/jul/25/afghanistan-war-logs-video-tutorial
warlogs-08	Afghanistan war logs: as it happened. 26 July 2010. http://www.theguardian.com/world/blog/2010/jul/26/afghanistan-war-logs-wikileaks
warlogs-09	"Data journalism" scores a massive hit with Wikileaks revelations. 26 July 2010. http://www.theguardian.com/media/greenslade/2010/jul/26/press-freedom-wikileaks
warlogs-10	Wikileaks data visualisations: what you did with our Iraq and Afghanistan spreadsheets. 8 November 2010. http://www.theguardian.com/news/datablog/interactive/2010/nov/08/wikileaks-data-visualisations-iraq-afghanistan

Prism	*Article* (title given as in the original)
Main article	
prism-main	NSA Prism program taps in to user data of Apple, Google and others. 6 June 2013. http://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data
Overview page	
prism-op	Prism. http://www.theguardian.com/us-news/prism
Other articles	
prism-01*	GCHQ taps fibre-optic cables for secret access to world's communications.

	21 June 2013. http://www.theguardian.com/uk/2013/jun/21/gchq-cables-secret-world-communications-nsa
prism-02*	XKeyscore: NSA tool collects "nearly everything a user does on the internet". 31 July 2013. http://www.theguardian.com/world/2013/jul/31/nsa-top-secret-program-online-data
prism-03	NSA spied on Indian embassy and UN mission, Edward Snowden files reveal. 25 September 2013. http://www.theguardian.com/world/2013/sep/25/nsa-surveillance-indian-embassy-un-mission
prism-04	NSA monitored calls of 35 world leaders after US official handed over contacts. 25 October 2013. http://www.theguardian.com/world/2013/oct/24/nsa-surveillance-world-leaders-calls
prism-05	Canada "allowed NSA to spy on G8 and G20 summits". 28 November 2013. http://www.theguardian.com/world/2013/nov/28/canada-nsa-spy-g8-g20-summits
prism-06	Edward Snowden identifies himself as source of NSA leaks – as it happened. 9 June 2013. http://www.theguardian.com/world/2013/jun/09/nsa-secret-surveillance-lawmakers-live

* article is not listed on overview page (and not included in quantitative overview), but still relates to the story.

References

Allan, Stuart, 2013. Citizen Witnessing: Revisioning Journalism in Times of Crisis.

Polity Press, Cambridge.

Bell, Allan, 1991. The Language of News Media. Blackwell, Oxford.

Benkler, Yochai, 2011. Free Irresponsible Press: Wikileaks and the Battle over the Soul

of the Networked Fourth Estate, A. Harv. CR-CLL Rev. 46, 311.

Benkler, Yochai, 2013. WikiLeaks and the networked fourth estate, in: Brevini,

Benedetta, Hintz, Arne, McCurdy, Patrick (Eds.), Beyond WikiLeaks. Palgrave

Macmillan, Basingstoke / New York, pp. 11–34.

Boczkowski, Pablo J., 2004. Digitizing the News. Innovation in Online Newspapers. MIT

Press, Cambridge, Mass.

Boczkowski, Pablo J., Mitchelstein, Eugenia, 2013. The News Gap: When the

Information Preferences of the Media and the Public Diverge. MIT press, Cambridge

Mass.

Brevini, Benedetta, Hintz, Arne, McCurdy, Patrick (Eds.), 2013a. Beyond WikiLeaks.

Palgrave Macmillan, Basingstoke / New York.

Brevini, Benedetta, Hintz, Arne, McCurdy, Patrick, 2013b. Introduction, in: Brevini,

Benedetta, Hintz, Arne, McCurdy, Patrick (Eds.), Beyond WikiLeaks. Palgrave

Macmillan, Basingstoke / New York, pp. 1–10.

Brevini, Benedetta, and Murdock, Graham, 2013. Following the money: WikiLeaks and

the political economy of disclosure, in: Brevini, Benedetta, Hintz, Arne, McCurdy,

Patrick (Eds.), Beyond WikiLeaks. Palgrave Macmillan, Basingstoke / New York,

pp. 35–55.

Bruns, Axel, 2005. Gatewatching. Collaborative Online News Production. Peter Lang,

New York.

Bruns, Axel, 2016. 'Random acts of journalism' redux: News and social media, in:

Jensen, Jakob Linaa, Mortensen, Mette, Ørmen, Jacob (Eds.), News Across Media:

Production, Distribution and Consumption. Routledge, New York / London, pp. 32–

47.

Cohen, Stanley, Young, Jock (Eds.), 1973. The Manufacture of News. Social Problems,

Deviance and the Mass Media. Constable, London.

Conboy, Martin, 2004. Journalism. A Critical History. SAGE, London.

Czarniawska, Barbara, 1999. Writing Management: Organization Theory as a Literary

Genre. Oxford University Press on Demand, Oxford.

Czarniawska, Barbara, 2011. Cyberfactories: How News Agencies Produce News:

Edward Elgar Publishing, Cheltenham.

De Kerckhove, Derrick, 1997. Connected Intelligence: The Arrival of the Web Society.

Somerville House, USA.

Dreyfus, Suelette, Lederman, Reeva, Bosua, Rachelle Milton, Simon, 2011. Can we

handle the truth? Whistleblowing to the media in the digital era. Global Media

Journal: Australian Edition 5 (1),

http://www.hca.westernsydney.edu.au/gmjau/archive/v5_2011_1/pdf/dreyfus_truth_

Essay.pdf.

Dunn, Hopeton S., 2013. "Something Old, Something New…": WikiLeaks and the

Collaborating Newspapers—Exploring the Limits of Conjoint Approaches to

Political Exposure, in: Brevini, Benedetta, Hintz, Arne, McCurdy, Patrick (Eds.),

Beyond WikiLeaks. Palgrave Macmillan, Basingstoke / New York, pp. 85–100.

Dutton, William H., 2009. The fifth estate emerging through the network of networks.

Prometheus 27 (1), 1–15.

Eldridge Scott, 2016. Hero or anti-Hero? Narratives of newswork and journalistic identity

construction in complex digital megastories. Digital Journalism,

http://dx.doi.org/10.1080/21670811.2016.1162105.

Elliott Chris, 2013. WikiLeaks and the Public Interest Dilemma: A View from Inside the

Media, in: Brevini, Benedetta, Hintz, Arne, McCurdy, Patrick (Eds.), Beyond

WikiLeaks. Palgrave Macmillan, Basingstoke / New York, pp. 78–84.

Fairclough, Norman, 1988. Discourse representation in media discourse. Sociolinguistics

17, 125–139.

Fairclough, Norman, 1995. Media Discourse. Arnold, London.

Flew, Terry, Liu, Bonnie Rui, 2011. Globally Networked Public Spheres? The Australian

Media Reaction to WikiLeaks. Global Media Journal: Australian Edition 5 (1),

http://www.hca.westernsydney.edu.au/gmjau/archive/v5_2011_1/pdf/flew_liu_RA.p

df.

Floyd, Alan, 2000. The reporting verbs and bias in the press. Revista Alicantina de

Estudios Ingleses 13, 43–52.

Fowler, Roger, 1991. Language in the News. Discourse and Ideology in the Press.

Routledge, London/New York.

Galtung, Johan, Ruge, Mari, 1973. Structuring and selecting news, in: Cohen, Stanley,

Young, Jock (Eds.), The Manufacture of News. Social Problems, Deviance and the

Mass Media. Constable, London, pp. 62–72.

Greenwald, Glenn, 2014. No Place to Hide: Edward Snowden, the NSA, and the US

Surveillance State. Metropolitan Books, New York

Harding, Luke, 2016. What are the Panama Papers? A guide to the biggest data leak in

history. Guardian, 3 April 2016,

http://www.theguardian.com/news/2016/apr/03/what-you-need-to-know-about-the-

panama-papers

Joerges, Bernward, 1999. Do politics have artefacts? Social Studies of Science 29 (3),

411–431.

Landert, Daniela, 2014a. Blurring the Boundaries of Mass Media Communication?

Interaction and User-Generated Content on Online News Sites, in: Tyrkkö, Jukka,

Leppänen, Sirpa (Eds.), Texts and Discourses of New Media. Studies in Variation,

Contacts and Change in English 15. VARIENG, Helsinki,

http://www.helsinki.fi/varieng/series/volumes/15/landert/.

Landert, Daniela, 2014b. Personalisation in Mass Media Communication: British Online

News between Public and Private. Pragmatics & Beyond New Series 240. John

Benjamins, Amsterdam/Philadelphia.

Lewis, Diana M, 2003. Online News: A New Genre?, in: Aitchinson, Jean, Lewis, Diana

M. (Eds.), New Media Language. Routledge, London/New York, pp. 95–104.

Locher, Michael A., Stanton E.F. Wortham, 1994. The cast of the news. Pragmatics 4 (4), 517–533.

Lovink Geert, Riemens Patrice, 2013. Twelve Theses on WikiLeaks, in: Brevini, Benedetta, Hintz, Arne, McCurdy, Patrick (Eds.), Beyond WikiLeaks. Palgrave Macmillan, Basingstoke / New York, pp. 245–253.

Lynch, Lisa, 2010. We're going to crack the world open. Wikileaks and the future of investigative reporting. Journalism Practice 4 (3), 309–318.

Lynch Lisa, 2013. The Leak Heard Round the World? Cablegate in the Evolving Global Mediascape, in: Brevini, Benedetta, Hintz, Arne, McCurdy, Patrick (Eds.), Beyond WikiLeaks. Palgrave Macmillan, Basingstoke / New York, pp. 56–77.

Newman, Nic, Dutton, William H,.Blank, Grant, 2012. Social media in the changing ecology of news: The fourth and fifth estates in Britain. International Journal of Internet Science 7 (1), 6–22.

Obermayer, Bastian, Obermaier, Frederik, Wormer, Vanessa, Jaschensky, Wolfgang, 2016. Das sind die Panama Papers. Süddeutsche Zeitung, 3 April 2016, http://panamapapers.sueddeutsche.de/articles/56ff9a28a1bb8d3c3495ae13/

Ostertag, Stephen F., Tuchman, Gaye, 2012. When innovation meets legacy: Citizen journalists, ink reporters and television news. Information, Communication & Society 15 (6), 909–931.

Papacharissi, Zizi (Ed.), 2009. Journalism and Citizenship. New Agendas in Communication. Routledge, New York / Oxon.

Poster, Mark, 1995. Cyberdemocracy: Internet and the public sphere. Internet Culture,

    http://www.humanities.uci.edu/mposter/writings/democ.html.

Rogers, Simon, 2013. Facts Are Sacred. The Power of Data. Faber and Faber / Guardian

    Books, London.

Stenvall, Maija, 2008. On emotions and the journalistic ideals of factuality and

    objectivity – Tools for analysis. Journal of Pragmatics 40 (9), 1569–1586.

Stenvall, Maija, 2014. Presenting and representing emotions in news agency reports. On

    journalists' stance on affect vis-à-vis objectivity and factuality. Critical Discourse

    Studies 11 (4), 461–481.

Sullivan, John, 2014. Uncovering the data panopticon: The urgent need for critical

    scholarship in an era of corporate and government surveillance. The Political

    Economy of Communication 1 (2),

    http://www.polecom.org/index.php/polecom/article/view/23/192.

Surowiecki, James, 2005. The Wisdom of Crowds. Anchor Books, New York

van Dijk, T.A., 2008. Discourse and Power. Palgrave Macmillan, Basingstoke.

van Dijk, T.A., 2009. Society and Discourse. How Social Contexts Influence Text and

    Talk. Cambridge University Press, Cambridge.

Wardle, Claire, Dubberley, Sam, Brown, Pete, 2014. Amateur Footage: A Global Study

    of User-Generated Content in TV and Online News Output. Tow Center for Digital

    Journalism, http://usergeneratednews.towcenter.org/.

Wardle, Claire, 2016. The Great Disrupter of the News Industry. The Huffington Post, 25

January 2016, http://www.huffingtonpost.com/claire-wardle/the-great-disrupter-of-

th_b_9069496.html.

Wardle, Claire, Williams, Andrew, 2008. User Generated Content. Understanding Its

Impact upon Contributors, Non-Contributors and BBC News. Cardiff School of

Journalism, Media and Cultural Studies, Cardiff,

http://www.bbc.co.uk/blogs/knowledgeexchange/cardiffone.pdf.

Wortham, Stanton, Locher, Michael, 1996. Voicing on the news: An analytic technique

for studying media bias. Text 16 (4), 557–585.

Zuboff, Shoshana, 2015. Big other: surveillance capitalism and the prospects of an

information civilization. Journal of Information Technology 30 (1), 75–89.