



<http://www.diva-portal.org>

This is the published version of a paper published in .

Citation for the original published paper (version of record):

Appelgren, E., Salaverría, R. (2019)

Interacting, but not contributing: fruitless news crowdsourcing in Spain

*Estudios sobre el Mensaje Periodístico*, 25(2): 639-655

<https://doi.org/10.5209/esmp.64792>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Attribution 4.0 International (CC BY 4.0)

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:sh:diva-38696>



## [en] Interacting, but not contributing: fruitless news crowdsourcing in Spain

Ester Appelgren<sup>1</sup>; Ramón Salaverría<sup>2</sup>

Received: July 2, 2018 / Accepted: December 10, 2018

**Abstract.** Trusting the audience to contribute data, a practice called crowdsourcing, is one of several procedures of contemporary data journalism. While previous research has embraced the idea of crowdsourcing as a transparent way of engaging the audience, the journalistic practice of crowdsourcing is currently still limited. The Spanish case is particularly illuminating of the gap between theoretical expectations and reality. Although online news media in Spain stands out because of its commitment to exploring interactive possibilities, and its audience is especially prone to comment on news, this qualitative study presents evidence that Spanish data journalists remain quite reluctant to embrace crowdsourcing as a research method. Based on semi-structured interviews with five Spanish data journalists, this paper explains the motivations behind this reluctance as well as the challenges of activating the audience.

**Keywords:** Crowdsourcing; citizen journalism; data journalism; interactivity.

## Interactuando sin aportar: el infructuoso crowdsourcing periodístico en España

**Resumen.** Confiar en la audiencia para que contribuya con datos, una práctica llamada *crowdsourcing*, es uno de los procedimientos del periodismo de datos contemporáneo. Si bien las investigaciones anteriores han adoptado la idea de *crowdsourcing* como una forma transparente de involucrar a la audiencia, esta práctica periodística todavía es limitada. El caso español es particularmente esclarecedor de la brecha entre las expectativas teóricas y la realidad. Aunque los medios de comunicación digitales españoles se destacan por su compromiso con la exploración de las posibilidades interactivas, y su audiencia es especialmente propensa a comentar noticias, este estudio cualitativo presenta evidencias de que los periodistas de datos en España siguen siendo bastante reacios a adoptar el *crowdsourcing* como un método de investigación. Basado en entrevistas semiestructuradas con cinco periodistas de datos, este artículo explica las motivaciones detrás de esta renuencia, así como los desafíos de la activación de la audiencia.

**Palabras clave:** *Crowdsourcing*; periodismo ciudadano; periodismo de datos; interactividad.

**Summary.** 1. Introduction. 2. Crowdsourcing, UGC and citizen journalism as part of “journalism as a conversation”. 3. Spanish media and data journalism. 4. Method. 5. Findings; 5.1. The difficulty of

<sup>1</sup> Södertörn University (Sweden)  
E-mail: ester.appelgren@sh.se

<sup>2</sup> Universidad de Navarra (Spain)  
E-mail: rsalaver@unav.es

encouraging the audience to participate; 5.2. The lack of collaborative culture between journalists; 5.3. The value of open journalism initiatives, such as citizen journalism and crowdsourcing. 6. Discussion and conclusions. 7. References.

**How to cite this paper:** Appelgren, Ester & Salaverría, Ramón (2019): "Interacting, but not contributing: fruitless news crowdsourcing in Spain". *Estudios sobre el Mensaje Periodístico* 25 (2), 639-655.

## 1. Introduction

During the past decade, journalism has been one of several disciplines that has embraced open source models. One such practice is crowdsourcing, a research method that is composed of three key elements: the “crowd”, outsourcing, and advanced Internet technologies (Saxton et al., 2013). Phenomena of such participatory culture, which were previously confined to the “geek culture”, are now becoming mainstream (Jenkins, Ito and boyd, 2016) and scholars have sought to categorize crowdsourcing. The focus has primarily been on its commercial aspects as in the taxonomy of crowdsourcing models suggested by Saxton et al. (2013). Non-commercial aspects of crowdsourcing however, have also been used by governments and authorities where public participation is used as a means for transparent and accountable problem-solving (Brabham, 2011). A third use for crowdsourcing can be found in the media in a form of “open journalism” (Aitamurto, 2017), where journalists apply various methods to collect data from the audience (Muthukumaraswamy, 2010).

The use of crowdsourcing in journalism mainly involves the public acting as assistants for journalists in their story processes in that the data submitted by audience members contribute to news research (Aitamurto, 2016). Howe (2006: 284) considered crowdsourcing projects such as these, where an individual acts as a “benevolent dictator”, collaborating with and guiding the crowd, to be the most successful. The practice of crowdsourcing in journalism is nevertheless often criticized (by journalists) for its lack of trustworthiness and problems with representativity; for example, a fake majority supporting certain opinions creates “a false sense of we” (Aitamurto, 2016) since journalists might not be able to reach a representative group or people in the audience may hijack crowdsourcing projects and submit fake data (see, for example, Vehkoo, 2013: 29). Aitamurto (2016, 2017) also finds that open journalistic practices, such as crowdsourcing, are in conflict with traditional journalistic norms, which rely on ownership of stories and content. Nevertheless, in journalism, trusting the audience to contribute data, or inviting them to participate (Lewis et al., 2014), has also developed into one of several techniques to collect information used in data journalism projects (Reference deleted).

In this paper, we use the broad definition of data journalism as suggested by Hermida and Young (2016), spanning computer-assisted reporting, data journalism and computational journalism, even though these three expressions represent somewhat distinct approaches. The practice of data journalism has often been described as a highly transparent form of journalism and one that activates the audience and uses interactive features often based on open data (Karlsen and

Stavelin, 2014; Reference deleted). Furthermore, data journalism represents one form of journalistic convergence between the open-source movement and professional journalistic norms (Coddington, 2015). In this context, Borges-Rey (2017) accounts for how several studies have pointed out that ideas from engineers and computer hackers had the potential to revolutionize the newsroom, for example by bringing along ideas of transparency originating from the computer science world, while other studies (see, for example, Bruns 2011) suggest that the inflow of such ideas in practice is limited. In recent research however, data journalists have been found to abandon interactivity when creating stories based on data, returning to linear forms of presenting content (reference deleted). Even though new interactive features are available, journalists thus appear to actively be steering the audience away from the possibility to engage with and affect journalistic content.

In this context, Spain is particularly interesting as the Spanish audience has a high tendency to engage with news. According to *Digital News Report 2016*, Spain stands out as being one of the countries where the residents are more prone to comment on news. With up to 80 percent of Internet users saying that they comment on online news at least once a week, Spain ranks fourth worldwide (among the 26 countries in the survey) and second among the EU countries behind Greece (Reuters Institute, 2016: 99). However, this high tendency to engage with news, which is especially intense on social media where Spain ranks first (p.100), does not mean that Spanish Internet users are likely to share news or trustworthy content. On the contrary, the vast majority of comments published by Internet users in Spain are merely commentary, i.e. opinion, so the online media outlets are interested in them not because of their contribution of relevant content but mainly because they help increase the outlets' traffic (Ruiz et al., 2010). Spain therefore demonstrates an interesting paradox: although residents are happy to share opinions publicly on the Internet, they are quite reluctant to collaborate on fact-based crowdsourcing projects. Over the past decade, Spanish online media outlets have only published a few examples of news that was based on crowdsourcing, and we are therefore analyzing in this paper the motives behind why journalists in Spain are not engaging with the audience in data journalistic projects. Our two research questions are as follows:

RQ1: What is it about the audience that makes Spanish data journalists reluctant to use it as a data producer?

RQ2: How do Spanish data journalists value open journalism initiatives such as crowdsourcing and citizen journalism?

In the following sections, we introduce the concept of crowdsourcing in journalism and discuss journalism as a conversation with the audience. We introduce the specific conditions of the Spanish media landscape and then continue with a section on the methodology used in the study, followed by the results, analysis and conclusions.

## 2. Crowdsourcing, UGC and citizen journalism as part of “journalism as a conversation”

Over the past decade, researchers have emphasized audience engagement as part of the future of journalism, in particular in the wake of success stories such as *The Guardian* (see, for example, Bruns, 2001; Aitamurto, 2016; and Manosevitch and Tenenboim, 2016). Due to the rise of mobile media and the use of social media, Pavlik (2015) suggested that public engagement would increase even further, and, according to Bruns (2011: 126), journalism needed to take on forms of conversation between equals. The relationship was no longer that of privileged producers and dutiful news consumers. This phenomenon of “journalism as a conversation” is defined by Marchionni (2013) as “the manner and/or degree to which interpersonal, reciprocal exchanges between journalists and everyday citizens, whether mediated or unmediated, enhance news work on matters of public import for the common good” (p. 136).

One example of such reciprocal exchange is crowdsourcing, which, according to Bruns (2011), challenges the gatekeeping model by creating a more equal environment between journalists and consumers of news. The history of crowdsourcing in digital media can be traced back to the time when pioneering Internet companies were designed to take advantage of the networked world, making little distinction between professionals and amateurs as labor for creating content (Howe, 2006). Back then, media scholars such as Jenkins (2008) also turned their attention to activities involving crowdsourcing, including it in the overall phenomena of the convergence culture. Here, activities carried out by media consumers were central to the convergence between the media, participatory culture and collective intelligence. Other related concepts such as user-generated content (UGC) and so-called “citizen journalism” were also increasingly discussed both by academics and practitioners, referring to the blurring of lines between producers and consumers of mediated content (Wahl-Jorgensen, 2015; Bruns, 2011).

Media scholars who have studied different types of open journalistic initiatives have however found that journalists are reluctant to share control (Usher, 2017). For example, in 2008, Domingo et al. carried out content analyses in six European countries but did not find journalistic projects that completely involved the audience in open journalism projects. The journalists kept the role of the “gatekeeper” and mainly let the audience participate in the interpreting stage, i.e. commenting on articles and debating current events. Singer (2010) found that British journalists using UGC still had their “elbows out” in terms of how they received different introductions of UGC. Their self-perception, the idea of what is quality content and their own role in safeguarding quality were found to be resilient to change and even challenged by it. Building on such previous findings, Usher (2017) proposed a four-step model for how citizen journalism content is currently used in professional news content. According to this model, news companies apply direct or passive appropriation to use content created by non-professional journalists. In a series of steps, the appropriated content is then amplified to a

larger news audience, either as native (offered on the news media website as an incorporated part of the news content) or deliberate amplification (shared across social media to a wider audience) (Usher, 2017: 253). Without this appropriation/amplification process, which is greatly aided by tools created for journalistic selection of UGC, Usher (2017) believes that citizen content would likely not be seen by a wider audience. This is a contradictory finding to Bruns (2011), which argues that active audience members increasingly will skip news organizations since they have their own channels for press releases and data from public institutions or corporations, and they furthermore share their insights independently of the news media (p. 118). Usher (2017) suggests, however, that such a development is unlikely to scale to wider audiences without mainstream media since news media outlets are bigger in size and have bigger networks than citizen journalists.

### 3. Spanish media and data journalism

Since the beginning of 21<sup>st</sup> century, the Spanish media landscape has undergone a significant transformation (Arrese et al., 2009; reference deleted; Baceiredo, 2017). Legacy newspapers have steadily lost their previous editorial hegemony, although they have managed to transfer part of that influence to their digital editions. Meanwhile, national and regional public broadcasting corporations have lost their former leadership in the audiovisual market to private television companies. However, the most salient phenomenon has been the increasing impact of digital media, both in terms of the growing influence of social media and the emergence of successful native Internet media (reference deleted). As of 2017, Spain is indeed one of the European countries in which the digital-born media has reached greatest diversity and, above all, greatest influence in public opinion (Nicholls et al., 2016).

The early initiatives in data journalism in Spain correlate with the introduction of the Spanish Transparency Act (BOE, 2013) in 2013 (La-Rosa and Sandoval-Martín, 2016). The law permitted citizens to request information from public authorities and use data published by the authorities. As the law entered into force in 2014, data journalism units were initially set up at *El Confidencial* in 2013 and *Eldiario.es* in 2014, both digital native publications. More units were opened later, for instance by the Vocento media group in 2014 and the *La Vanguardia* and *El Mundo* newspapers and the La Sexta television channel in 2015. During this time, according to Freedom House (2015), the decline of traditional media and the increased use of digital media empowered social minorities in Spain and supported political pluralism and digital activism. This surge of digital media has taken place within a journalism culture that, compared to other European countries, more intensively embraces the “convergence culture” (reference deleted). Nevertheless, as far as the interactivity with the public is concerned, research has found that there is a big gap in Spanish online news media between the theoretical possibilities (Schultz, 2000; Masip et al., 2010) and the real achievements (García-Avilés et al., 2014; Meso Ayerdi, 2015; Suau and Masip, 2015). In a cross-national study about the typology of comments published on five online editions of top quality newspapers (The New York Times, U.S.; The Guardian, UK; Le Monde, France;

La Republica, Italy; and *El País*, Spain), Ruiz et al. (2011) found that the majority of contributions published on the Spanish website fitted into the “dialogue of the deaf” (p. 482), having the second highest rate of derogatory references and insults. Thus, the online news media outlets in Spain seek user participation not so much for its potential contribution to the news supply, but mainly because it boosts the traffic. As a result, it may be understandable that Spanish data journalists find few incentives within their newsrooms to put a journalism based on crowdsourcing into practice. Nevertheless, compared to the other European countries, Spain remains a country where “editorial departments exert more effort toward online content (reference deleted). and Spanish online news media differ from many other countries because of its commitment to multimedia, i.e. interactive and hypertextual possibilities (reference deleted) Infographics in particular are one form of multimedia that has been found to be especially well developed and prominent on Spanish news sites (Herrero-Solana and Rodríguez-Dominguez, 2015).

#### 4. Method

A central theme in this study is the phenomenon of crowdsourcing as used by data journalists in Spain. We use what Bertrand and Hughes (2005: 109-111) refer to as a contemporary institutional perspective that looks at how technology is managed, what effects this has on institutional products and how the institution interacts with its audience. In this case, the institution is the newsroom working with the evolving practice of data journalism. To understand the motivations to use or avoid crowdsourcing and the challenges this method poses to data journalists, we used semi-structured interviews with five key informants working at five Spanish news organizations: two print newspapers with online editions (*El Mundo*, the third largest Spanish national general news newspaper, founded in Madrid in 1989, and *Diario de Navarra*, a centenary regional newspaper, published in Pamplona since 1903), two digital native publications (*El Confidencial*, the biggest national digital native publication of Spain, founded in Madrid in 2001, and *Nació Digital*, a digital native Catalan news company, founded in Barcelona in 2005) and a news start-up (Datadista, a project-based start-up specialized in data journalism, launched in Madrid in 2016).

Initially, a database of data journalists was set up, consisting of a non-probability purposive sample of 96 identified Spanish data journalists. They were sent a link on Twitter to participate in a short survey about the current status of data journalism in Spain (response rate 45%). From this survey, five key informants were selected using the following survey question: In an average week, how much time do you spend working with data journalism projects? We selected five subjects who stated they devote more than 70 percent of their working time to data journalism. Four of the chosen journalists work in the newsrooms or were working with data journalism at the time the specialization first appeared on the Spanish media market; the fifth journalist, younger than the rest, works in a regional newspaper and, so far, has focused her entire career on data journalism. Four of the

interviewed journalists have titles that reveal they work with data journalism, such as coordinator of the data desk or data journalist, and one held the title Director.

According to Kvale (1996), informed consent in qualitative research interviews involves informing research subjects about the purpose of the study, the research design and the possible risks and benefits from participating. We obtained informed consent to record the interviews before each interview started and briefed the subjects about the overall aim of the study. We furthermore asked for their permission to use the results for research purposes and to use their names in our research publications. Two interviews were conducted in English and three in Spanish. All interviews were transcribed and the three Spanish interviews were translated into English. After the interviews were conducted, we debriefed the subjects, as suggested by Kvale (1996: 128), about the general aim. We highlighted some of the most interesting points of their interview and asked once again for permission to use their names in our research results in case they felt they had shared something more sensitive in nature; they all consented they had not.

The semi-structured interviews took place between February 7 and February 23, 2017, in Madrid, Barcelona and Pamplona, at locations chosen by the subjects, such as a conference room at the news organization or a nearby café. The interviews were between 40 and 50 minutes long, and the interview guide consisted of 15 questions. We began by introducing questions about the definition of data journalism, the subjects' personal experiences with data journalism and the current main challenges in Spain. We continued with probing questions about three themes, of which one is used for this paper. The three themes were: 1) ethics in data journalism, 2) the Spanish Transparency Act and, finally, 3) the use of crowdsourcing. For each theme, the subjects were asked to give examples and, when necessary, we asked follow-up questions mostly by repeating what the subjects had already stated (see, for example, Kvale, 1996: 133).

## 5. Findings

Our first research question addressed the audience and what makes the journalists reluctant to use crowdsourcing to gather data. All five informants, who ranged from young journalists to pioneers in data journalism, expressed admiration for crowdsourcing projects from international media companies such as The Guardian, but confessed that they have limited experience with using crowdsourcing themselves. We identified two themes during the interviews as to why crowdsourcing was not a preferred working method for the data journalists: the difficulty of encouraging the audience to participate and a lack of collaborative culture between journalists. These two themes echo what Jenkins (2016) discusses as the convergence between participatory culture, media convergence and collective intelligence, and will be discussed more in detail in the following two sections.

Our second research question addressed how journalists value crowdsourcing in data journalism and previous forms of audience engagement methods in journalism, such as citizen journalism. The informants stated that they had not previously linked these two phenomena together, but they all agreed there were



similarities between the two. Given the outcome of citizen journalism as a previously hyped, but now abandoned, concept, they tended to discuss citizen journalism as a failure and crowdsourcing as a promise they would like to engage in more, if preconditions were a bit different.

### 5.1. The difficulty of encouraging the audience to participate

Motivations for using crowdsourcing methods in data journalism are contradictory. The five interviewees emphasized the importance of keeping the audience engaged, yet they were reluctant to use audience members as data producers. As Singer (2010) previously found, safeguarding quality can partly explain this reluctance; even though journalists may be attracted to use crowdsourcing methods, there are also potential risks.

I really like crowdsourcing data projects, but they are also quite dangerous. Well, not dangerous, but you need to be very careful. (Daniele Grasso, *El Confidencial*)

The danger described in the quote refers to the cumbersome process of verifying crowdsourced material. To exemplify, Grasso mentions a successful, yet time-consuming, project that he and his two colleagues on the data team published in 2014 called “¿Madrid da asco? 'Mapeamos' la basura en las calles de la capital” [Does Madrid suck? We map the rubbish in the streets of the capital]. The project was based on images from the people of Madrid, originally posted in social media channels picturing garbage on the streets as the result of a three-day long strike affecting the garbage collection.

*El Confidencial* published a city-wide rubbish map based only on pictures verified using Google Street View, which was a highly time-intensive task:

For our mental health, [we were lucky that] it was just a three-day strike. We would have gone crazy [if] we had to verify photos all day long for one month. That would have been totally crazy.

This manual verification process is generally part of how data journalists refine and clean data, regardless of whether the data consists of numbers, text or images. Refining the data is mandatory for quality journalism, and the data in this example was non-professionally produced images. Riding this wave of audience engagement, where content is highlighted by journalists practicing what Bruns (2011) denoted as “gatewatching”, it was possible for *El Confidencial* to create not only a good piece of news, but also a perfect example of when crowdsourced material is beneficial for journalism. The process of obtaining this data set fits into what Usher (2017) would denote as “passive appropriation”, where journalists purposefully search for citizen content on social media, rather than asking citizens to create exclusive content for them.

This passive approach to crowdsourcing seems to be more effective than the active approach. Ana Cordobés at the local newspaper, *Diario de Navarra*, which is based in Pamplona in northern Spain, highlighted the risk of not engaging audiences to contribute in crowdsourcing initiatives. When, in the past, this newspaper asked citizens to create exclusive content for them, what Usher (2017)

labels as “direct appropriation”, participation levels from the local audience were low. Thus, an ambitious crowdsourcing project where the newspaper encourages and instructs users to submit and create content for them that results in no answers could be a real problem for the newspaper and be viewed as a risk not worth taking. However, even though participation was found to be low at *Diario de Navarra*, Cordobés found that the audience is not all that passive. In fact, the newsroom does reach beyond the newspaper whenever it finds something that is wrong in visualizations or if it does not understand how to use interactive features. In such instances, where readers can see and correct journalists’ mistakes, the power balance between the audience members and the journalists is briefly equal, creating something of a more transparent dialogue (Aitamurto, 2016).

Karma Peiró, the director of Catalan news outlet *Nació Digital*, emphasizes that it is easy to get Spanish people to be active on social media. It is making them interact in a context that has proven itself to be difficult.

Participation is the pending subject for the digital media, followed by business models. Participation is now much more complex than before, mainly because of the breakthrough of social networks, such as Facebook and Twitter. Before this breakthrough, you could expect people to comment on news stories when you asked for their participation. However, now [social media] is very immediate; it’s in the mobile phone... You can ask them to write their opinion in a tweet or on a Facebook page - there you will find hundreds of comments. However, there are few reflexive comments and comments with almost no thought at all, sometimes destructive or violent. [...] People are no longer interested in having a relationship with the news.

The difficulty of getting the audience to interact in a news context is one example of how Spain demonstrates an interesting paradox: although citizens are happy to share opinions publicly on the Internet, they are quite reluctant to collaborate in fact-based news projects. This may be because audience engagement in news contexts in Spain is still in its infancy (Masip, 2016), similarly to what has been found in other countries almost a decade ago (Bergström, 2008).

According to Antonio Delgado at *Datadista*, however, crowdsourcing topics can be newsworthy in Spain if the topic itself is sufficiently interesting:

In Spain, when the topic is corruption, you get thousands of people ready to spend an hour if necessary, you know? [...] I think it’s more about the emotional aspect of the general public than the inter-personal relationship with the medium or the journalist. Why do the people collaborate with *The Guardian* on receipts [referring to The Guardian’s crowdsourcing project where people helped journalists make sense of the members British Parliament’s expenses documented in 5,500 PDF files]? Because of the widely appealing nature of the topic, not because of *The Guardian*. If any other medium would have done the same, it would also have worked. If a civilian would have done that, it also would have worked, because the topic is above the platform, or who was leading.

Previous research has pointed to the imbalance of power between journalists and citizens as the reason why both sides have not yet collaborated more closely on content production (see, for example, Bruns, 2011). The above quote emphasizes

the importance of an engaging topic rather than the identity or authority of the person in charge of a crowdsourcing initiative. According to this view, the role of a “benevolent dictator” naturally taken by journalists when they put forward a crowdsourcing project is not the obstacle to overcome. A meaningful topic that can turn into an exciting story will attract the audience to participate and thus determine the success of the project.

## 5.2. The lack of collaborative culture between journalists

At *El Mundo*, Hugo Garrido also admits that his data journalism team has promoted few crowdsourcing experiences. In his case, crowdsourcing methods were not used because of the competitiveness between journalists and the lack of a collaborative culture:

It might be a personal self-censorship, but sometimes I don't even suggest the idea [of crowdsourcing] because I know that there's a lack of collaborative culture. In fact, if you share information that you have, you are giving it away for free to your rival. The information is a bit restrictive, something exclusive and valuable that you keep just for your own medium, due to the culture of the company.

The quote refers to what Aitamurto (2016) denotes as crowdsourced projects with horizontal transparency, where audience submissions are visible to anyone online and, thus, if you are a competitor, also possible to “steal” during the process of gathering the data for the story. Such an open approach is naturally in contrast to how data journalists normally work with large data sets, where a story is published after the data has been gathered, cleaned and refined rather than published as an ongoing project. It is possible to see in this quote yet another aspect that does not involve the power imbalance between the audience and the journalists, i.e. the competition between journalists as a hinder towards using the crowd. For structural reasons, open journalism initiatives are avoided in a more traditional media company while digital native media companies may be more prone to take on the risk of having content stolen before a story is published in order to adapt to new media practices. Instead, Garrido prefers what Aitamurto (2016) denotes as vertical crowdsourcing, a more closed version where only the journalist can see submissions. Garrido mentions the project “Congreso de los Diputados”, which was based on declarations by members of the Spanish Parliament (PM) about the properties they own. The documents proved to be incomplete, and journalists therefore asked the PMs to provide their missing declarations. Not everyone was willing to cooperate, and some responded that they had no obligation to provide the information, even though they are obliged by law to respond to such requests. This particular example mixes the two types of appropriation described in Usher's model; on the one hand, there is a passive appropriation of the content since the PMs have generated data themselves in their property declarations, which are then used by journalists, and, on the other hand, there is also a direct appropriation since the PMs are also asked to fill in the gaps. In this case, the data didn't come from ordinary citizens but officials. However, regardless of the profile of the sources, they react similarly to such journalistic requests. At the time of this research,

Garrido was working with a similar methodology on a project related to Senate members and expressed yet another anticipated side-effect of publishing this kind of partly crowdsourced dataset that is related to the challenge of competition between journalists:

We would like to publish that information, and when that happens I guess other people will work with our data. In fact, I'm sure there have been people from other news media outlets who have entered our piece [about the PMs] to analyze the data because it's more accessible and who have elaborated news based on our data.

In this quote, Garrido expresses typical values of data journalists that are in conflict with traditional investigative journalism. Aitamurto (2016) argues that a closed process prevents competing news organizations from stealing the topic and potential sources from hiding information. However, open data publishing for peers is in line with journalists' goals of making information public.

Delgado at Datadista gave an example of how journalists may publish news based on data they did not gather while they aid the development of new stories. He referred to the project "Los papeles de Bárcenas" [Barcenas' papers], a corruption scandal unveiled in 2013 within the People's Party (PP), which was then ruling the national Government. Findings from the analysis of these documents were first published by the newspaper *El País* in print format, and, a few days later, a database was added, where it was possible to filter the documents according to a limited number of criteria. Delgado, who is not employed by *El País*, channeled a request on Twitter asking his tens of thousands of followers to collaboratively create an online spreadsheet of the printed material. This direct appropriation to his followers was soon retweeted by other Twitter users, reaching an even larger network of potential volunteers. This type of action is comparable to the prediction of Burns (2011), where active audience members, in this case coinciding with Delgado being a journalist, bypassed traditional media channels to crowdsource content that was later to be used as journalistic content. Delgado explained his intention as follows:

If I look at the data in that way [refers to the way *El País* released the data], it could only be used to tell anecdotes, but I felt a need to analyze everything as a whole to find journalistic stories. That is roughly the philosophy of data journalism.

As other newspapers started to publish parts of the files of Luis Bárcenas, the PP treasurer, Delgado continued to ask for help to upload the data in a shared Google sheet. He could eventually compare different datasets published in print by various newspapers, finding errors in the overlapping data. Because the data originated from handwritten documents, there were errors in how journalists deciphered the manuscripts. With the help of a graphologist and thanks to a specific technology for handwriting recognition, Bárcenas' documents could then be understood more easily and errors in several news stories were identified and corrected. In short, using social networks as a mobilizing tool, Delgado managed to create a collaborative dataset in an accessible and reusable format, which was verified and refined thanks to the help of mainly fellow journalists and expert

researchers rather than just ordinary people. In his direct appropriation of data, Delgado used what Usher (2017) denotes as the “professionalism gate”. Journalists using this gate would typically “have more time to choose content that adheres most closely to professional standards” (p. 256). Delgado explains his methodology as follows:

The form that I chose, which is slightly like a craftsmanship, was done, first, by creating a circle of trust. In other words, allowing everybody to edit, but giving more trust to some people than others. I had a ring of close people; I mean, not close to me but close to the project, who were involved in the project since the very first moment and who verified it. Secondly, I checked the dataset myself, by hand. It wasn't that much; at that stage of the process the verification was possible. And not only by me; other people checked the data as well. In this case, this could be done because Bárcenas' papers, while quite large in size, had only hundreds of cells rather than millions of cells.

The example given by Delgado shows an almost rebellious attempt to change the culture of how legacy Spanish media were working with data. On his own initiative, he did exactly what, for example, Garrido mentions as one of the reasons he does not engage in crowdsourcing; he re-uses already published data. Such a boot-strapping method indeed signals a more open form of journalism, putting the ownership of the originally published stories into question by expanding them. But Delgado's methods are still consistent with maintaining journalistic control since he uses a more vertical approach to open journalism through a smaller group, many of them journalists, that he trusts to refine the data.

### **5.3. The value of open journalism initiatives, such as citizen journalism and crowdsourcing**

The three more experienced interviewees had been working as digital journalists since the mid-2000s when citizen journalism was a hyped concept. They judged citizen journalism as a failure mainly because of the scarcity of contributors to such platforms. Delgado, one of the most experienced, recalls the still-active CNN initiative, iReporter:

iReporter! It was all bullshit. There was nothing interesting. It was like... “Well, since everyone has a little camera, let's invite them to bring us things”. But, in fact, I think it is always seen from the perspective of the reader...

Two of the interviewed journalists worked in journalism less than five years. They have no direct relationship with the concept of citizen journalism, even though they are familiar with the phenomenon.

All informants agreed that Spanish attempts at citizen journalism failed, but exemplified instead a similar mechanism flourishing on social media. In Usher's Appropriation/Amplification model (2017), the publishing channels of media companies are crucial to amplify citizen content in order to reach a wider audience. In contrast to this model, the activity people undertake when creating and posting content online was not seen by the interviewed Spanish data journalists as

something that media companies could actively encourage; it was already happening outside of the control of the legacy media sphere.

When crowdsourcing methods are used in data journalism for verifying data, however, the informants believed that this is a form of citizen journalism. Within this context, all five informants described the challenges brought on by social media in relation to citizen journalism. Because everyone can use social media to channel their opinions, people are not in an apparent need of the publishing channels of legacy media for communication. Delgado describes this development also as one of the reasons why citizen journalism did not take off:

As it happens with Wikipedia and other sectors, a very small number of active users participate actively on these kinds of platforms. People go there to watch, but very few participate. Secondly, the boom of social networks. When I enter the social networks and everybody has that loudspeaker, their own medium, why should I publish in another media to spread my communication if I can be my own medium without the help of anyone else? If the fall of citizen journalism could be quantified on a graph, I think that it coincides with the rise of Facebook. I think that citizen journalism and Facebook are very closely related.

In the quote, Wikipedia is on par with citizen journalism initiatives, perhaps because Wikipedia represents an initiative where the community involved in creating content works with a minimal amount of hierarchy (Aitamurto, 2016). Early citizen journalism attempts were similarly focused on users creating the stories, and, as Burns (2011) expressed it, had no gatekeeping at all. However, in the above quote, once again, the power balance is not blamed for why citizen journalism attempts failed. Rather, the motivations for why people create content was transferred onto a new medium, from creating non-professional content as citizen journalism stories to the content created when interacting on Facebook. While it may be problematic that social media websites can be used as loudspeakers for the general public, Grasso, who did not work as a journalist when citizen journalism was at its peak, believes that journalists should be thankful that people today voluntarily act as citizen journalists on social networks:

[..] they are not journalists, but if something happens, they [citizens] can record it, send it or put it on their social media. They are doing something that is journalistic. I think that it is easier today for journalists to check that content, verify it, use it, spread it on other networks, and compare it to what other people are saying in the same zone.

Thus, we have not found evidence that journalists do not want to use the audience because of them not producing newsworthy content or content of low quality. Rather, all informants seem to long for a more active audience in terms of people taking a greater interest in news, and pessimistically do not expect that they, with the help of their media brand, can influence people to have a more active relationship with news. In the following section, we will finally ponder how this longing for an active audience and the need for relevant topics have affected the use of crowdsourcing in data journalism.

## 6. Discussion and conclusions

The names and features of open journalistic initiatives have varied over the past few decades. According to Bruns (2011), new generations of citizen journalists appear from time to time and challenge the traditional “gatekeeping monopoly” of journalists.

RQ1 was focused on the audience and why Spanish data journalists are reluctant to use it as data producers. We found that, since citizen content is used by the data journalists as a source, they treat citizen contributors as sources whose content will be harvested into large data sets, refined and then elevated into a story with many data points. When using citizen content in this manner, Spanish data journalists show no reluctance. However, they are reluctant to blur the barriers between professionals and citizens because of the risk that they will not engage the audience to contribute fully. Contrary to previous research, our study has not found a deep concern about the power balance between the gatekeeping journalists and the content-producing audience. One reason Spanish data journalists may not fear the audience as contributors is given by Self reference (year), which found that Mediterranean journalists, perhaps because newspaper distribution systems in their countries were never as strong as in the northern and central European countries, were more likely to embrace change in newsrooms and welcome innovation.

Other studies have found that data journalism methods are often enhanced by the use of algorithms to find new stories in massive datasets and visualize content (See for example Diakopoulos, 2015). Such convergence of the technology and journalism often involving the open-source movement, typical for data journalism (Coddington, 2015), could not clearly be found in our results. Rather, an opposite culture was described. The idea of open collaborations was mentioned as occurring occasionally, and usually with positive experiences, but in reality the Spanish data journalism practice was more often described as a semi-manual and closed practice limited to a smaller group of data journalists within an organization. However, we propose that this closed culture could easily benefit from algorithmic or automated audience participation since it does not present the challenges to ownership of stories and content that are typical for horizontal and transparent open journalism initiatives (Aitamurto 2016), and future research could investigate reasons for why algorithmic data journalism methods is not yet used in newsrooms in Spain.

RQ2 focused on how Spanish data journalists value open journalism initiatives, such as crowdsourcing and citizen journalism. Recent research emphasizes the importance of the relationship between the audience and journalists in open and conversational news forms, giving examples of how to increase user involvement and engagement (see, for example, Boyles and Meyer, 2016; Aitamurto, 2016). We found, however, that the journalists believe that an open and active relationship with the audience, in terms of horizontal crowdsourcing or citizen journalistic content gathered with direct appropriation (Usher, 2017), even though desirable, may be lost in a country such as Spain due to the widespread use of Twitter and Facebook. In such a media landscape, the question is if there is a need for forms of open journalism based on audience-created content that are for the exclusive use of

journalists. Perhaps user involvement in this context could be reinvented as something that takes less effort, at least from the audience's point of view; for example, as previously suggested by increasingly replacing manual steps of the journalistic processes with the use algorithms as suggested by Diakopoulos (2015). Crowdsourcing based on content the audience has created on its own and that was selected by journalists would furthermore reduce the critique of trustworthiness and representativity from the fake submission of data as described by Aitamurto (2016) and Vehkoo (2013).

In line with Bruns' ideas about gatwatching (2011), we have found that the interviewed journalists are coming to terms with the fact that Spanish people no longer want an active relationship with news media since they now have their own channels to communicate with the world. Instead of being hindered by this development, we have found examples of how the journalists are trying to use citizen content that was not originally submitted exclusively to the news media. Such content is used as their source and from this they create data sets that can be used in data journalism projects. We argue that this form of data journalism is an updated form of gatwatching, not affecting the powerbalance between the journalist and audience, as both sides get to keep a sense of being in control.

## 7. References

- Aitamurto, T. (2016): Crowdsourcing as a Knowledge-Search Method in Digital Journalism: Ruptured ideals and blended responsibility. *Digital Journalism* 4(2): 280–297.
- Aitamurto, T. (2017): Crowdsourcing in Open Journalism: Benefits, Challenges and Value Creation. In: Franklin B and Eldridge II SA (eds.) *The Routledge Companion to Digital Journalism Studies*. London: Routledge.
- Arrese, Á; Artero, JP; and Herrero-Subías, M. (2009): The media in Spain: Three decades of transformation. In: Albarran, A. (ed.): *The handbook of Spanish language media* New York: Routledge, pp.17–33.
- Bergström, A. (2008): The reluctant audience: Online participation in the Swedish journalistic context. *Westminster Papers in Communication and Culture* 5(2): 60–79.
- Bertrand, I. and Hughes, P. (2005): *Media Research Methods: Audiences, Institutions, Texts*. Basingstoke: Palgrave Macmillan.
- BOE (2013): Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno.
- Boyles, J.L. and Meyer, E. (2016): Letting the Data Speak: Role perceptions of data journalists in fostering democratic conversation. *Digital Journalism* 4(7): 994-954.
- Borges-Rey, E. (2017): Towards an epistemology of data journalism in the devolved nations of the United Kingdom: Changes and continuities in materiality, performativity and reflexivity. *Journalism*. Epub ahead of print 1 February 2017. DOI: 10.1177/1464884917693864.
- Brabham, D.C. (2012): Motivations for participation in a crowdsourcing application to improve public engagement in transit planning. *Journal of Applied Communication Research* 40(3): 307–328.



- Bruns, A. (2011): Gatekeeping, gatewatching, real-time feedback: new challenges for journalism. *Brazilian Journalism Research* 7(2): 117–136.
- Domingo, D.; Quandt, T.; Heinonen, A.; Paulussen, S.; Singer, J.B.; and Vujnovic, M. (2008): Participatory journalism practices in the media and beyond: An international comparative study of initiatives in online newspapers. *Journalism practice* 2(3): 326–342.
- García-Avilés, J.A.; Kaltenbrunner, A.; and Meier, K. (2014): Media convergence revisited: lessons learned on newsroom integration in Austria, Germany and Spain. *Journalism Practice* 8(5): 573–584.
- García-de-Torres, E. (2010): Contenido generado por el usuario: aproximación al estado de la cuestión. *El profesional de la información* 19(6): 585–594.
- Hermida, A. and Young, M.L. (2016): Finding the Data Unicorn: A hierarchy of hybridity in data and computational journalism. *Digital Journalism* 5(2): 159-176.
- Herrero-Solana, V. and Rodríguez-Domínguez A.M. (2015): Periodismo de datos, infografía y visualización de la información: un estudio de *El País*, *El Mundo*, *Marca* y *El Correo*. *BiD: Textos universitaris de biblioteconomia i documentació*, 34.
- Howe, J. (2006): The rise of crowdsourcing. *Wired magazine* 14(6): 1–4.
- Jenkins, H. (2006): *Convergence culture: Where old and new media collide*. New York: NYU press.
- Jenkins, H.; Ito, M.; and Boyd, D. (2015): *Participatory culture in a networked era: A conversation on youth, learning, commerce, and politics*. Malden, MA: Polity Press.
- Karlsen, J. and Stavelin, E. (2014): Computational journalism in Norwegian newsrooms. *Journalism Practice* 8(1): 34–48.
- Kvale, S. (1996): *InterViews. An introduction to qualitative research writing*. Thousand Oaks, CA: SAGE Publications.
- La-Rosa, L. and Sandoval-Martín, T. (2016): The Transparency Law's insufficiency for Data Journalism's practices in Spain. *Revista Latina de Comunicación Social* 71: 1208–1229.
- Lewis, S.C.; Holton, A.E.; and Coddington, M. (2014): Reciprocal journalism: A concept of mutual exchange between journalists and audiences. *Journalism Practice* 8(2): 229–241.
- 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): 383–393.
- Marchionni, D.M. (2013): Journalism as a Conversation: A concept explication. *Communication Theory* 23(2): 131–147.
- Masip, P. (2016): Investigar el periodismo desde la perspectiva de las audiencias [Investigating journalism from the perspective of journalism]. *El profesional de la información* 25(3): 323–330.
- Manosevitch, I. and Tenenboim, O. (2017): The Multifaceted Role of User-Generated Content in News Websites: An analytical framework. *Digital Journalism* 5(6): 731–752.
- Meso Ayerdi, K. (2015): Journalism and audiences: debate on the quality of user-generated content. In: Meso Ayerdi K, Agirreazkuenaga Onaindia I and Larrondo Ureta A *Active audiences and journalism: analysis of the quality and regulation of the user generated contents*. Bilbao: Universidad del País Vasco, pp. 65-82.

- Muthukumaraswamy, K. (2010): When the media meet crowds of wisdom: How journalists are tapping into audience expertise and manpower for the processes of newsgathering. *Journalism practice* 4(1): 48–65.
- Pavlik, J. (2015): Journalism: From delivering information to engaging citizen dialogue. In: Cantoni L and Danowski JA (eds) *Communication and technology*. Berlin: Walter de Gruyter GmbH, pp.513–526.
- Rodríguez Gómez, E.F. and Sandoval Martín M.T. (2013): Tendencias del periodismo de investigación: ONG's, crowdfunding y crowdsourcing. In *Sociedad Española de Periodística, XVIII Congreso Internacional: Los nuevos desafíos del oficio del periodismo*. Madrid: Sociedad Española de Periodística, pp.659–684.
- Ruiz, C.; Domingo, D.; Micó, J.L.; Díaz-Noci, J.; Meso, K.; and Masip, P. (2011): Public sphere 2.0? The democratic qualities of citizen debates in online newspapers. *The International Journal of Press/Politics* 16(4): 463–487.
- Ruiz, C.; Masip, P.; Micó, J.L.; Díaz-Noci, J.; and Domingo, D. (2010): Conversación 2.0. y democracia. Análisis de los comentarios de los lectores en la prensa digital catalana [Conversation 2.0. and democracy. An analysis of reader's comments in Catalan online newspapers]. *Communication & Society* 23(2): 7–39.
- Saxton, G.D.; Oh, O.; and Kishore, R. (2013): Rules of crowdsourcing: Models, issues, and systems of control. *Information Systems Management* 30(1): 2–20.
- Schultz, T. (2000): Mass media and the concept of interactivity: an exploratory study of online forums and reader email. *Media, culture & society* 22(2): 205–221.
- Singer, J.B. (2010): Quality control: Perceived effects of user-generated content on newsroom norms, values and routines. *Journalism Practice* 4(2): 127–142.
- Suau, J. and Masip, P. (2015): Models of online media participation and active audiences – a comparison of what the media are offering and what citizens are looking for. In: Meso Ayerdi, K.; Agirreazkuenaga Onaindia, I; and Larrondo Ureta, A.: *Active audiences and journalism: analysis of the quality and regulation of the user generated contents*. Bilbao: Universidad del País Vasco, pp.119–142.
- Usher, N. (2017): The Appropriation/Amplification Model of Citizen Journalism: An account of structural limitations and the political economy of participatory content creation. *Journalism Practice* 11(2–3): 247–265.
- Wahl-Jorgensen, K. (2015): Resisting epistemologies of user-generated content? Cooptation, segregation and the boundaries of journalism. In Carlson M and Lewes SC (eds) *Boundaries of Journalism*. New York: Routledge, pp.167–183.
- Vehkoo, J. (2013): *Crowdsourcing in investigative journalism*. Oxford: Reuters Institute for the Study of Journalism.

---

Ester Appelgren: Journalism Department, School of Social Sciences, Södertörn University, Huddinge, Sweden

Ramón Salaverría: Center for Internet Studies and Digital Life, School of Communication, Universidad de Navarra, Pamplona, Spain.