Digital Journalism

ISSN: 2167-0811 (Print) 2167-082X (Online) Journal homepage: http://www.tandfonline.com/loi/rdij20

Crowdsourcing as a Knowledge-Search Method in Digital Journalism

Tanja Aitamurto

To cite this article: Tanja Aitamurto (2016) Crowdsourcing as a Knowledge-Search Method in Digital Journalism, Digital Journalism, 4:2, 280-297, DOI: 10.1080/21670811.2015.1034807

To link to this article: http://dx.doi.org/10.1080/21670811.2015.1034807

Published online: 08 May 2015.

CROWDSOURCING AS A KNOWLEDGE-SEARCH METHOD IN DIGITAL JOURNALISM
Ruptured ideals and blended responsibility

Tanja Aitamurto

This article examines crowdsourcing as a knowledge-search method and an open journalistic practice in digital journalism. The study draws on data from four cases in which professional journalists used crowdsourcing in their investigations. Crowdsourcing resulted in an efficient knowledge discovery and a continuous flow of tips to journalists and thus benefited journalistic investigations. The horizontal and vertical transparency in crowdsourced journalism supported the knowledge-search process. However, the high volume of submissions in some cases made the journalists compromise the journalistic norm of data verification, which resulted in publishing unverified information. Crowdsourcing as an open journalistic practice thus ruptures journalistic norms and creates pressure for new ones to emerge, such as blended responsibility, in which the responsibility for data accuracy is shared by the journalists and the readers. The article extends the examination of open journalistic practices and contributes to the understanding of their impact on digital journalism.

KEYWORDS co-creation; crowdsourcing; digital journalism; knowledge search; knowledge management; open journalism; open knowledge; transparency

Introduction

Crowdsourcing is increasingly used as a knowledge-search method in digital journalism (Bradshaw and Brightwell 2012; Muthukumaraswamy 2010). Crowdsourcing as an open journalistic practice is becoming more common in gathering information for news articles and in co-creating stories with readers (Aitamurto 2013). Open practices hold the promise of effective knowledge searches, thus supporting journalistic investigations, yet little is known about their impact on knowledge-search and management mechanisms in journalism. While open journalistic practices in digital journalism are becoming more common in newsrooms, there is a lack of research about the impact of crowdsourcing and co-creation on the knowledge-search and management practices in journalism.

By drawing on data from several story processes in which crowdsourcing was used by professional journalists in established newspapers and magazines, this article examines crowdsourcing as a knowledge-search practice in digital journalism. The focus

Digital Journalism, 2016
Vol. 4, No. 2, 280–297, http://dx.doi.org/10.1080/21670811.2015.1034807

is on the impact of crowdsourcing on journalistic norms, practices, and ideals governing and shaping knowledge-search and management mechanisms in professional journalism. The research questions are the following:

RQ1: How does crowdsourcing function as a knowledge-search method for professional journalism?
RQ2: How does crowdsourcing impact on journalistic norms, practices and ideals that govern knowledge-search and management procedures in professional journalism?

The article is structured in three sections. The first reviews relevant literature about crowdsourcing, co-creation, open journalism, and knowledge-search practices, drawing on theories and evidence in journalism studies and information systems science. The second presents the case profiles, methods, and data. The third presents the findings, followed by conclusions.

Theoretical Framework and Key Concepts

Crowdsourcing and Co-creation in Journalism

Crowdsourcing is an open call for anyone to participate in a task online (Brabham 2008, 2013; Howe 2008). "The crowd" refers to an undefined group of people who participate in the open call online. Outsourcing, on the other hand, means that the task is assigned to a specific agent. In crowdsourcing applications, the crowd is invited to complete an online task, for instance by submitting information that the crowdsourcer —in the context of this article, a journalist or a newspaper—is searching for. Crowdsourcing is used in several contexts ranging from ideation challenges for companies' research and development needs (Jeppesen and Lakhani 2010), to urban planning (Brabham 2010), public policy-making (Aitamurto 2012; Aitamurto and Landemore 2015), identifying corruption (Brito 2008; Meier 2011), to tagging photos for mapping the consequences of a hurricane (Liu 2014). In crowdsourced journalism, the crowd is invited to participate in the journalistic processes in various ways, such as submitting knowledge, sharing opinions, and sending pictures. Well-known examples of crowdsourced journalism include the British newspaper *The Guardian*'s initiative to use crowdsourcing to examine hundreds of thousands of documents related to the British politicians' 2009 expenses scandal (Aitamurto 2011; Daniel and Flew 2010; Vehkoo 2013) and the 2011 Hurricane Irene crowdsourced effort for journalism in the United States (Dailey and Starbird 2014). In these instances, the journalists used crowd-generated input for the purposes of professional journalism, typically conducting the standard fact-checking procedures before using the crowd's input in published stories.

In crowdsourcing, a task initiator—for example, a newspaper—invites readers to conduct tasks online, asking the readers to help discover and share information that is relevant to the investigations. This typically entails a one-time response or an act by the participant (Afuah and Tucci 2012), in which there is very little, if any, interaction between the journalists and the readers. In contrast, in co-creation, the process is shared between the crowd and the organization (Piller, Ihl, and Vossen 2011). Reader-participants and journalists work together in a two-way interaction, including peer-to-peer communication among readers, resulting in a shared experience (Banks and Deuze 2009; Hatch and Schultz 2010). Co-creation is a series of crowdsourcing moments

following each other, and the crowd-input received in preceding phase determines the use of the method in the following phase.

Crowdsourcing and co-creation function as tools in open journalism (Aitamurto 2013). Openness refers to both public accessibility and transparency during various phases in the journalistic process. The degree of openness varies depending on the method used, but the journalistic process is rarely, if ever, fully open to public participation. Rather than practicing full openness in a story process, journalists make visits "to the open," which means opening up only certain parts of the story process to public participation. Figure 1 illustrates these moments of openness. A journalist can crowdsource story ideas, ideas for visualizations in the story, story angles, and reactions to a published story. Crowdsourcing also functions as need-finding in that it allows the journalist to sense the readers' interest in various topics and their needs to know more about those, and can thus guide decision-making about which topics will be covered. Open journalistic practices have been found to generate raw material for the journalists to build on, develop further, and to weave to their stories. Open journalism also strengthens the reader-relationship,[1] but comes with high human resource and organizational adaptation costs (Aitamurto 2013.)

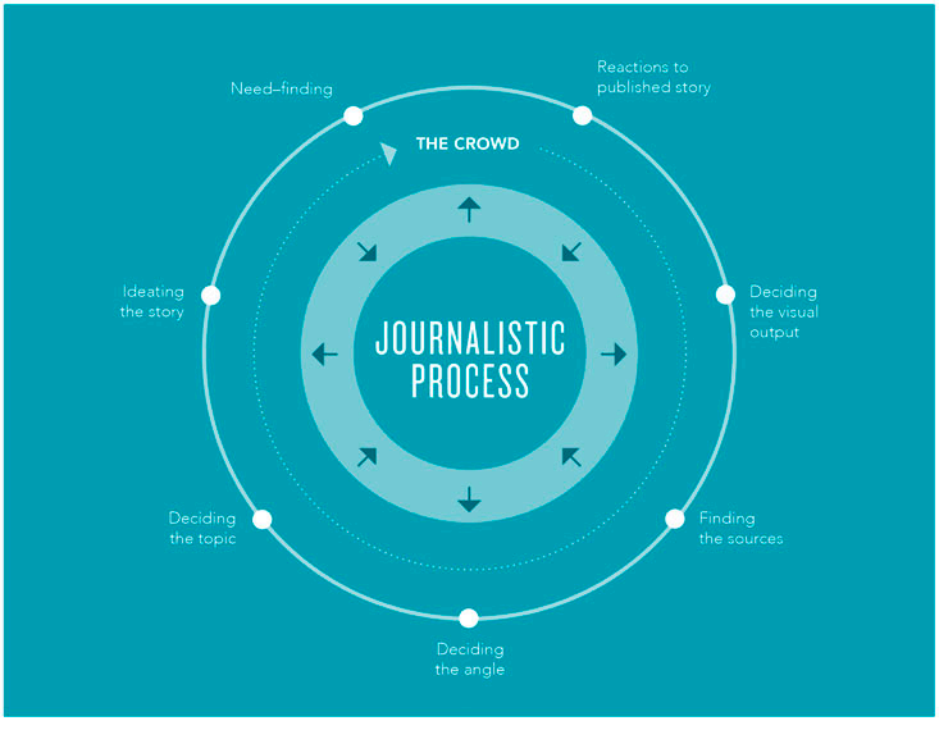

FIGURE 1
In open journalism, journalists make visits from the traditional, more closed journalistic process to the more open process in several modularized stages. The arrows in the inner circle illustrate those visits. The inner circle depicts the more closed journalistic process, and the outer circle illustrates the open moments that are shared with the crowd. The journalists determine the moments of openness: when and how the public is invited to join a journalistic process

In open journalism, journalists make visits from the traditional, more closed journalistic process to the more open process in several modularized stages. The arrows in the inner circle illustrate those visits. The inner circle depicts the more closed journalistic process, and the outer circle illustrates the open moments that are shared with the crowd. The journalists determine the moments of openness: when and how the public is invited to join a journalistic process.

Openness is a particularly radical shift from traditional investigative journalism, which is typically a closed process conducted in secret until the end result—the story—is published. Investigative journalism is a specific form of reporting that is often focused on revealing corruption, corporate wrongdoing, or social inequalities (Aucoin 2005; Ettema and Glasser 1998). The opaqueness of the process prevents competing news organizations from stealing the topic. It also prevents potential sources from hiding relevant information.

As a knowledge-search method, crowdsourcing has specific characteristics that differentiate it from other large-scale online collaboration architectures, such as commons-based peer production. In crowdsourcing, the locus of power lies within the process organizer—the crowdsourcer—who conducts the crowdsourcing initiative. The crowdsourcer decides when, where, and how the process takes place and how the input will be used. In this way, crowdsourcing differs from Wiki-type collaboration, which is commons-based peer production. In commons-based peer production (Benkler 2002), the power lies with the commons or community, such as in Wikipedia article creation or in free and open-source software production, and the process involves a minimal amount of hierarchy.

Open journalistic practices can be used as knowledge-search methods for both participatory and citizen journalism. In participatory journalism, readers participate in the journalistic process (Domingo 2011; Domingo et al. 2008) as, for instance, commentators or content producers (Bruns 2005; Holmes and Nice 2012). In citizen journalism, people who are not formally journalists produce news and content that can be perceived as journalism (Gillmor 2004),[2] which can appear on blogs and news sites run by citizens or on established news sites. However, not all participatory journalism is citizen journalism, and vice versa. Nor do participatory journalism and citizen journalism always use crowdsourcing or co-creation as a knowledge-search method. On citizen journalism sites, there is necessarily not a crowdsourcer, who masters the process, but the production is organized from the bottom-up. In those cases, citizen journalism sites resemble more commons-based peer production, in which the contributors are equal and there is minimal hierarchy in the production process.

Local and Distant Search in Crowdsourcing and Co-creation

An important element helping in understanding crowdsourcing and co-creation as knowledge-search mechanisms is a knowledge neighborhood, which refers to the knowledge space each individual operates in. Models and frameworks based on the metaphor of a knowledge landscape have been applied to a wide variety of contexts, such as evolutionary biology (Kauffman 1993), organizational sciences (Nelson and Winter 2009), and problem solving in general (Hong and Page 2004). When an agent—here, a journalist—looking to solve a problem searches for knowledge near his or her current position, the search is framed by the agent's existing knowledge. This is called a local search. In contrast,

extending the search beyond the known neighborhood is called a distant search. Crowdsourcing and co-creation can improve problem-solving performance in cases in which distant knowledge is needed (Afuah and Tucci 2012). Instead of performing a distant search, the problem solver may crowdsource the task to a large number of participants, all of whom perform only local searches. If the number or diversity of participants is high enough, the chances rise that someone is already in the neighborhood of the solution and can find the solution by using a local search (Page 2008). Thus, a distant search is transformed into a local search, and problem-solving performance can be improved.

Crowdsourcing is a way for journalists to fill gaps in their knowledge. A metaphor for illustrating the mechanism in practice is fishing with nets. By crowdsourcing, journalists cast their nets into the water. The net is larger and wider than in a traditional journalistic knowledge search, which involves the journalist calling potential sources one by one. In crowdsourcing, the journalist's call for information goes out to a massive number of people simultaneously, and thus, can result in an effective discovery of knowledge. Crowdsourcing can be particularly useful to journalists who are new to a topic or to news journalists who are not covering their regular beat, because news journalists are time-pressured for finding relevant information very fast.

Methods, Case Profiles, and Data

Case Profiles

A multiple case study methodology was chosen to emphasize discovery in the study (Eisenhardt 1989; Huberman and Miles 2002). This study draws on seven story processes which were opened up to the public for participation. Readers were invited to participate in several pre-editorial stages as they were asked to find information, and share experiences and expertise as a basis for journalists' stories. In all cases, crowdsourcing was initiated by professional journalists working for the publications in which the stories were ultimately published. The cases included investigative journalism, news journalism, and service journalism. The case profiles are detailed below.

Case A. The leading daily newspaper in Finland, the *Helsingin Sanomat*, with a readership of over one million throughout Scandinavia, crowdsourced information in order to investigate the short-selling of stocks in the fall of 2011. The investigations were provoked by a bill proposed by the Finnish government that, if passed, would decrease the transparency of the stock markets. The newspaper wanted to investigate the pervasiveness of short-selling under the current regulations. The newspaper acquired documents that reported stock selling, published them online, and then instructed readers on how to investigate the reports and track irregularities. The readers submitted information by filling out an online form and by sending emails to the journalist investigating the case. Based on the crowdsourced information, several unethical or questionable short-selling activities were uncovered. But most importantly, the reporter received information beyond his initial inquiry that led to the discovery of a tax evasion scheme within the largest co-operative bank in Finland. The story provoked a nationwide discussion and led to the firing of a bank executive in early 2012. The journalist later received an award from the Finnish Association of Investigative Reporters for the best investigative story of the year, and the Ministry of Finance later abandoned its plan to decrease transparency in stock markets.

Case B. Three story processes were opened up for online participation in several publications' blogs, websites, and social media, as well as on a specific platform for crowdsourcing and co-creation in journalism. The stories were produced and published in Finnish magazines in 2011 and 2012. The first story examined incorrect information in physics schoolbooks, the second story examined issues with quality in Finnish products and services, and the third investigated gender inequality in mathematics and science education. The reporters opened several parts of the story process. For example, they asked readers to identify incorrect information in schoolbooks and to share sources that could be interviewed and their experiences with problematic products and services. The journalists published updates about the investigation's proceedings during the story process. The story about quality in services was published in a women's lifestyle magazine (*Kotiliesi*), and the schoolbook investigation story and the story about gender differences in mathematics and science education were published in a science magazine (*Tiede-lehti*). These three stories are grouped together because they were completed by the same journalists and because the story processes were similar.

Case C. Three journalists working for the *Svenska Dagbladet*, one of the leading daily newspapers in Sweden, opened up two story processes to the public in 2009 and 2010. The first story covered the efficiency of development aid allocations by the Swedish government. The second examined the quality of senior care. The readers were able to follow and participate in the process on a blog on the newspaper's website. The journalists set up tasks for readers to participate in, such as submitting information about corruption in the development aid projects and about problems in senior care and suggesting sources to interview. The investigations led to several stories in print and online, exposed problems in the systems, and provoked discussion in Sweden and public responses from local and national governments. For the story about corruption in the development aid system, the journalists were presented the Best Innovative Entry of the Year award by Schibsted, the corporate owner of the newspaper. These two stories are grouped together because they were completed by the same journalists and because their story processes were similar.

Case D. The *Svenska Dagbladet* also opened an investigation into mortgage interest rates in Sweden in 2012. People were asked to submit information about their mortgages and interest rates by filling out an electronic form. The information was then displayed on a constantly updating map on the newspaper's website, all the information being geographically located and visualized. This type of crowdsourcing is called crowdmapping (cf. Furtado et al. 2012). By July 2013, about 40,000 interest rate submissions had been placed on the map. The map broke the website traffic records for the newspaper. The process resulted in 30 articles about mortgage interest rates, the inequalities within them, and the lack of regulation of banks, provoking nationwide discussion. People used the information on the map to renegotiate their interest rates, and some received discounted rates. Other news outlets used the data gathered on the map, and external Web developers built applications based on the data. Prior to the investigation, there had been discussions of the process by which the banks determined interest rates and potential problems therein.

For each of the cases profiled above, the data were gathered through in-depth interviews, elaborated below.

Methods and Data

In-depth interviews with journalists leading the crowdsourced investigations were conducted. In case A, the journalist was interviewed three times—first at the beginning of the story process, in October 2011, again in January 2012, when the stories were published, and again in February 2012. In case B, there were three journalists involved in the story processes, of which two journalists were interviewed three times from 2011 to 2012—at the beginning of the investigations, in the middle, and again when the stories were published—and one journalist who was interviewed once in 2012. In case C, the three journalists involved in the process were interviewed once in 2010. In case D, a journalist involved in the process was interviewed in 2012 and another journalist was interviewed in 2013. Altogether, 15 in-depth interviews with nine journalists were conducted. The questions in the semi-structured interview scheme focused on the challenges and benefits of the open process, knowledge-search, and discovery methods, and changes in journalistic norms and practices. The interviews were recorded and transcribed.

The interview data were analyzed following Strauss and Corbin's (1998) analytical coding system. In the first round of coding, open coding was used, allowing key themes and patterns to emerge from the data and thus guide further analysis. In the next round of coding, axial coding was used to relate the emerging categories to subcategories (Saldana 2009, 159; Strauss and Corbin 1998, 123). Finally, selective coding was applied to integrate and synthesize the subcategories into the following categories: (1) the quality, volume, and type of knowledge; (2) accuracy and data verification; and (3) transparency. Qualitative data analysis software was used in coding. The numbers 1–9 below refer to individual interviewees.

Findings

Extended Search and Unexpected Tips

Journalists acquired information by crowdsourcing and co-creation that assisted them in their story processes, and they received crucial information that helped bring the stories to completion. In cases A (the stock short-selling case) and C (the development aid case and the senior care quality investigations), the journalists were able to navigate a larger number of documents than they could have without the readers, as is described in the following excerpt from a journalist working on the story about problems in development aid:

> We couldn't look into every municipality in Sweden. No way! There was no time for that, but they actually did that for us. Maybe they didn't look into every municipality, but they looked into their own, where they lived. They wrote to us and said, "Oh! This is looking kind of fishy. What is this? Can you look into it?" (1)

The documents enabled the crowd to contribute to journalistic investigation, thus bridging knowledge asymmetries between the participants and journalists. By reading the journalists' instructions for checking the documents and perusing the documents, the readers were able to contribute to the journalistic investigation, even without knowing much about the topic in advance. By sharing information about the

development aid projects with the journalists, the readers helped the journalists choose which projects were worth investigating further, as was also the case in the senior care quality investigation and the stock short-selling case. By publishing those thousands of documents online, journalists enabled the readers to access information that normally would be out of their reach, and thus helped readers to perform a distant search—they could access knowledge that was not initially within their reach. The readers would not have had access to those reports without the newspapers' initiative to publish the documents, and the journalists would not have been able to peruse as many documents as the readers collectively could. The extension of readers' knowledge neighborhoods and coupling that to the journalistic investigations made the journalists' knowledge-search process successful.

The crowdsourced tasks were based on the assumption that the reader either has the knowledge that the journalist is searching for or that the reader can find the information in the documents published with the task, or a combination of those two. In case A, the short-selling story, the journalist was particularly hoping he would receive help from those readers who had expertise in the area. Thus, he hoped the readers would be able to reveal something that was, as of yet, unknown and unseen, as described in the following excerpt from the journalist working on the story:

> Knowledgeable readers can identify things in the reports that I can't see or don't understand. There are over 5 million readers, 800,000 small investors, and over 800,000 shareholders in mutual funds in Finland. They have more information than I do. (2)

Furthermore, the journalist saw crowdsourcing as a way to extend the information search to all readers and thus increase the possibility of finding relevant information. As a result, the journalist received about 50 tips and reports about short-selling via the online form and via emails. One of the reports included an unexpected tip about holding companies related to a Finnish co-operative bank. The journalist investigated this matter further, and as a result, he discovered a questionable holding company arrangement owned by executives in the bank. The holding company was operating as a way of avoiding taxes on bonuses for the executives. The tipster has proven to be a good source for information even after this particular story process. The open journalistic process lowered the threshold for the source to approach the journalist, and thus created long-lasting value for the journalist in a form of a valuable regular tipster.

In other cases as well, journalists received information that they did not even know about. To illustrate one example, in case C (the development aid story), journalists were contacted by a source they had not sought out—a former Swedish ambassador in Namibia who had vital information about the matters under investigation. Thus, by crowdsourcing, the journalists moved beyond their own knowledge neighborhoods and those of the known sources by reaching out into the readers' knowledge neighborhoods. Because of the widespread call online, the search is conducted simultaneously in thousands, tens of thousands, and hundreds of thousands of knowledge neighborhoods. The result was information that contributed to and even steered the investigations.

The story processes in all cases triggered a flow of tips to the journalists, and it continued even after the story process was over. Open journalistic practices thus lower the threshold for the public to contact the journalists—journalists become more visible

and accessible in the investigations, which are opened up to public participation, and therefore the readers connect to the journalists more easily with their tips and hints.

Horizontal and Vertical Transparency in Crowdsourced Knowledge Searches

There are two types of transparency in crowdsourced journalism: horizontal and vertical, similarly to crowdsourced policy-making (Aitamurto and Landemore 2015). When there is horizontal transparency, the crowd's submissions are visible to anybody online. When there is only vertical transparency in the process, the crowd's input is visible only to the crowdsourcer, the journalist who initiates the crowdsourcing process. All cases studied for this article had horizontal transparency, apart from the stock short-selling case, which had only vertical transparency in the crowd's submissions, as only the journalist was able to see the readers' submissions. In the stock short-selling case, however, there was horizontal transparency in the documents, which were publicly accessible online.

Even in the horizontally transparent cases the journalists implementing crowdsourcing were not afraid of the publicity of the process or of being scooped. They perceived their investigation to be unique enough that no other news outlet would dare or bother to pick up the same topic. They perceived that openness would actually protect the stories from being stolen, because it was publicly known that they were the original instigators of the investigations. Instead of worrying about protecting information from being scooped, after publishing the stories, journalists hoped that the open story process would result in wider coverage of the matters by other news organizations. In cases in which the crowdsourced information resulted in massive datasets (case D: mortgage interest rate) or included documents published online (cases A and C: stock short-selling, problems in development aid projects, and the senior care quality), journalists hoped that other journalists would also use the data, as the following excerpt from an interview with a journalist describes:

> I actually hope that other media will find something that we haven't discovered so that this issue would move forward in order to show the public what the initiative [to decrease transparency in monitoring stock-trading] in the Ministry really means for society. (2)

Journalists viewed their stories as agents of change, and themselves as actors in this change process, and the more other actors use their data and the more they build on those data, the better it is for the society. Journalists' attitudes toward open journalistic process are similar to those of open-source coding communities, in which code is shared in the community for other developers to build on. However, journalists do want to protect their right to the scoop and to the topic. They use several procedures for doing that. For instance, in the stock short-selling case, before launching the crowdsourcing initiative the journalist examined some of the trading documents himself. He then wrote a small news story about the issues that he discovered in the documents. The story was published at the launch of the crowdsourcing process, so the journalist covered the topic first. But in case D, the mortgage interest rate case, no such protective mechanism was used, and the story idea was "stolen" from *Svenska Dagbladet*. The

national news agency in Sweden scooped a story based on the mortgage interest rate map before *Svenska Dagbladet* published such a story. The journalists who were working on the story for *Svenska Dagbladet* viewed this as an unavoidable consequence of open investigations, as described in the following excerpt:

> but we said from the start that we just wanted to be very open with this project, and actually, we would never hunt for any competitors or any other people who were using the data. We just said that you could use the data if you acknowledge the source. If you want to use it, go right ahead. (5)

However, the principle of openness was initially in conflict with the traditional journalistic norm of owning the story process and the data, as described in the following excerpt from a journalist. He describes an instance in which a student created a Web application to search for interest rates using the data from the map in the newspaper:

> At first, we were like, "Shit, who is this guy? He stole our thing." Then, of course, we were like, "No, no. Now we're going to be very open with it." (5)

This reflects the constant challenge of balancing the open and closed paradigms, practices, and norms. New, adapted journalistic norms, which allow more openness, are constantly evolving as a result of this conflict. Open journalistic practices challenge the traditional, more closed paradigm in journalism, which relies on ownership of stories and content. The more horizontal transparency there is in the story process, the more accessible the crowd-generated data are and the more conducive it is for others to use the data. Paradoxically, this supports the journalists' goal of sharing the data and letting others build on that; yet, in parallel, horizontal transparency exposes the information to the public and makes it possible for others to "steal it," thus amplifying the unexpected nature of open journalistic initiatives.

The crowd also contacted the journalists by means other than the crowdsourcing platform; they emailed and phoned them. For instance, in the development aid and the quality in senior care investigations, the readers took the initiative to call the journalists very actively, and journalists received most of the information that way. A similar pattern occurred with the schoolbook errors investigation, the quality of services story, and the gender differences in mathematics and science education story. The tipsters and sources also often wanted to remain anonymous, and journalists informed them that anonymity was a possible option. This shows that even when horizontal transparency is available to the participants, they may choose a more protected communication channel. Therefore, when applying open journalistic practices, it is necessary to have a variety of communication avenues available to the participants to secure the efficiency of the knowledge search.

Journalists see the transparency of the open process benefiting journalism as a means of increasing credibility and accountability. The data on which the investigations are based are publicly available online so that if anybody wants to check the facts, the documents are accessible online. As one journalist put it, "The evidence for the accuracy of the story is all online. Just like in a trial" (2). For instance, in case C, the politicians whose activities were questioned in the aid story claimed that the article was incorrect; the journalists and the readers then referred to the online documents, proving the politicians wrong. The journalist elaborated the process in the following excerpt:

> You can say: "Are we right, or are they right? Who is wrong? You can make up your own mind. We don't have to tell you." I don't think anyone thought they were right. We didn't get any responses like that. People said, "Yes, it's in the papers. It's in the documents." (1)

Furthermore, journalists embrace the transparency of the process with regard to the readers. Readers also get to see the journalists' mistakes and monitor the journalists, who are another power-holder in society. The dialog with readers has the potential to build trust in journalism, as described in the following excerpt from a journalist: "This is a way of showing what we do and why we do it and sharing the process with other people. I'd say that's a good way to improve our reputation" (6).

Data Verification and Blended Responsibility

Open journalistic practices did not save time or resources in any of the cases studied. If time was saved in one part of the process, some other stage required more time. In cases A and C, which included publishing documents online, the process of acquiring, formatting, and digitizing those thousands of document pages was a time-intensive effort. In case D, the huge number of reader submissions (over 40,000) required quick and constant technical iteration in order to manage the large amount of data. In most cases, processing the crowd-submitted information took a lot of time because of the volume of the submissions.

The larger the amount of data gathered through the open journalistic process, the more challenging it is to follow traditional journalistic practices of verification. In most cases (A, B, and C), the journalists crosschecked the crowd-submitted information used in the stories, unless it was a participant's opinion or perspective about an issue. Responsibility for the accuracy of the facts stayed with the journalists, as described in the following excerpt: "The responsibility is ours. It's we who are going to do all the journalistic work. We do so with help from the readers, but it's still our responsibility" (1).

The verification of crowdsourced information created the largest challenge in case D, in which 40,000 reader submissions were displayed on the newspaper's website. At the beginning of the process, the newspaper attempted to act according to traditional journalistic norms, which stipulate that all information should be checked before being published—including, in this case, the identity of those submitting information and whether their reported mortgage interest rates were correct. Journalists reached out to about 80 readers who had submitted their contact information when sharing the mortgage interest rates to get a sense of who was using the tool and whether there was any hint of fake submissions. The newsroom also removed some obviously false interest rates from the data. However, journalists soon realized that the volume of submissions was too large to be verified. The newspaper told readers that the mortgage interest map was not a verified database, allowing the readers to draw their own conclusions about the correctness of the data:

> We have been very clear from the beginning that we have opened this up to the public and that we try to get rid of answers that are deliberately false. However, we have never said that we try to verify all the facts here. This is a tool for the users and created by the users, and it's up to the people how much they rely on this. (5)

The situation reflects not only a balance of traditional journalistic norms with new open procedures, but a balance of responsibility between readers and journalists—a kind of blended responsibility. Open journalism provided a space for citizens to participate in a nationwide effort to map interest rates and, by making the submissions public, provided an opportunity to observe others' interest rates, thus increasing horizontal transparency. Instead of taking the responsibility for verifying the information, journalism pushed the responsibility onto the public, thus implementing new norms caused by the pressure that a new form of information gathering—crowdsourcing—had created. Yet, the lack of clarity about how accurate the information is continues to be a weakness in crowd-generated content, and journalists also acknowledge that weakness. Because it was impossible to verify the interest rate information, journalists used the society's reaction as an indicator of the accuracy of the information:

> The impression has been that the result [the crowd-generated map] has been quite balanced, and no banks have contacted us and said to us, 'Come on guys, this is so wrong. You have to check your materials because this is real wrong because you could see that directly because of that or that." (9)

When the interest rate data were analyzed in the newsroom, it turned out that the wealthier areas had lower interest rates than the other areas, one reason being the increased ability of wealthier people to negotiate a better deal with the bank, perhaps because they understand the economy better. Furthermore, it turned out that the state-run Bank of Sweden had the highest interest rates, which was in contrast to the bank's image. This enabled the journalists to question the Finance Minister's activities, as described in the following excerpt from a journalist:

> he's been criticizing the banks for their high mortgages. Then, of course, we could put the question to him like this: "Your own bank is by far the most expensive. Why is that?" (5)

The mortgage interest rate map thus led to a nationwide discussion about inequalities in interest rates. Some people renegotiated their interest with their bank based on the information in the mortgage interest map and got a discount on their rate. Crowdsourcing enabled journalists to gather a massive amount of information, and allowed readers to participate by submitting a small piece of data that accumulated into an efficiently collected mass of data. The process thus allowed journalists and citizens to challenge power holders—in this case banks and politicians—in society. However, in the case of a massive amount of reader submissions, this was only possible by amending the journalistic ideal of data verification.

As a result of rendering to the ideal, the efficiency of the knowledge-search mechanism in open journalism increases journalism's vulnerability: it is ever easier to feed incorrect information to journalists through crowdsourcing and co-creation. It would even be possible to mobilize people to do so on a massive scale and thus make journalists spend their time filtering the false data from the correct data. Thus, the efficiency of the knowledge-search mechanism in open journalism can compromise the benefits of the search, because a large amount of data makes knowledge discovery more time-consuming, or even impossible, and can thus compromise the information verification.

Scoops Are Lost in Co-created Journalism

The co-created open investigations (cases B and C) required more resources throughout the process than the crowdsourced ones did. While crowdsourcing is a one-time shot of acquiring information, co-creation requires a constant dialog with the readers, which takes a significant amount of time, as reflected in an excerpt from a journalist in case B:

> Oh yeah, the process took so much more time than a typical story process. Although the process guided the sketching of the story outline and was helpful … But writing blog posts all the time and engaging in the dialog with the readers takes a lot of time. (3)

The co-created journalistic process is iterative; adjustments and changes have to be made midstream based on readers' feedback and interactions. The nature of shared information in co-creation is different than that of crowdsourcing. In co-creation, the readers write about their experiences, and the journalists have to then synthesize the most important from often long comments and build the next phases based on the synthesis.

The real-time nature of the Web has also created pressure to constantly publish blog posts and move on with the investigations. Furthermore, transparency in the co-created process also leads to unexpected consequences, altering the traditional ideals of "scooping" in journalism. In co-created stories, the notion of big scoops is lost in the process. Everything is published immediately, as elaborated in the following excerpt from a journalist:

> I think that if we had investigated in the old-fashioned way, if we just had gathered all the information and then gathered it all together in one or two big publications, all the material at the same time, then it would have had a greater impact on other media. (4)

This creates a situation in which news does not come in one cluster, but rather in a flow of information. Traditionally, if a news article creates a splash, that is regarded as success in journalism, and in co-created journalism that splash is lost in the open process. Thus, traditional journalistic practices collide with the new open practices, impacting the evaluation of the project's success.

Open Journalism and the Digital Divide

As the evidence from the cases studied shows, the topics most germane to open journalism, when measured by the amount and quality of participation, are those in which the readers have first-hand experience and knowledge, like home loan interest rates, or experiences with gender inequality in classrooms. The threshold for sharing such information is low. If the knowledge relates to our loved ones, we might also share the knowledge simply because we care. This was undoubtedly the case in the senior care story, in which relatives shared information about senior care homes based on the experience of those close to them. Having said that, journalists might be inclined to pass on topics that are not that conducive to online sharing, such as highly sensitive issues like high-level corruption. The more sensitive the knowledge is, the less likely the sources are to share the information in public, or by using digital means like

private emails that leave a trace, particularly in this age of pervasive digital mass-surveillance, in which governments and other actors are increasingly using sophisticated monitoring mechanisms, as, for instance, the revelations of the surveillance conducted by the National Security Agency, in 2013, show.

Furthermore, when considering the future of open journalism, we must understand the limits of open practices. One major restriction is the fact that not everybody has equal access to the digital realm. There are also groups which are not able to speak for themselves—people like the disabled, illegal immigrants, and victims of human trafficking. Therefore, the entrance to all knowledge landscapes is not equally open: societal structures create barriers to entrance and to the transmission of knowledge from the landscape. Crowdsourcing, however, is a method relying on participation, so if a population whose voice is relevant in the news article is not able to participate, crowdsourcing can bring in only a part of the information needed in the story. Furthermore, the access to open journalism requires crossing many divides, including those of digital divide and literacy. This can lead to a false sense of majority online, of "we," if the excluded are forgotten, as warned by Fraser (1993). Therefore, journalists need to remain alert to filters, which restrict the participation in open journalistic practices.

Moreover, open journalism is vulnerable to manipulation by somebody mobilizing his or her forces to influence the process. In this so-called "astroturfing" (Monbiot 2011), fake personas with credible-appearing backgrounds try to steer online sites by pursuing their agenda by participating in online conversations. Journalism needs to remain alert to that and develop tools to ensure the accuracy of the information used in the stories. Moreover, crowdsourced input should not be treated as the public opinion of the people. Crowdsourcing is a self-selecting mechanism, and thus, it does not lead to a statistically representative sample of the population, and in that, it does not replace opinion polls.

Conclusions

By applying open journalism practices, journalists extended the knowledge search from their local knowledge neighborhoods to the distant knowledge neighborhoods of the crowd. As a result, one person's—in this case, a journalist's—search was turned into multiple local searches in the crowd's knowledge neighborhoods, leading to successful knowledge discovery. By publishing documents about the topics under investigation and asking readers to peruse them, journalists push readers to conduct a distant search in the provided data. The extended search contributed to the journalists' investigations and helped them find information that they did not even know to look for. The open journalistic process with horizontal and vertical transparency thus amplified journalists' information search power, increasing the likelihood for knowledge discovery.

However, the efficiency of the knowledge search easily produces a large amount of information, which requires processing power, that processing power involving journalists' time and resources. The crowd's submissions are often qualitative in nature, and journalists have to read all of them to find nuggets of valuable information in the submissions. The resources required for information processing diminish the possibilities for knowledge discovery and mitigate in part the benefits of the open journalistic process, as is found also in earlier studies about open practices in journalism (cf. Aitamurto

2013). This challenge might be addressed at least partially in the future when Natural Language Processing technologies (for NLP technologies, cf. Convertino, Sándor, and Báez 2013) will be developed further and they can successfully automate the analysis of the crowd's input.

Moreover, open practices sometimes conflict with traditional journalistic norms and ideals. The efficient knowledge-search mechanism in crowdsourcing can result in tens of thousands of submissions, as happened in one of the cases studied in this article. Journalists cannot verify that much information. Instead, they have the option either to abandon crowd-generated information, or they can push the responsibility of information accuracy back onto the readers, attempting to make clear how the mechanisms of open journalism work: the crowd must submit correct information, but nobody can verify that they actually do so. It is also the crowd's responsibility to understand that the data, which are published on the newspaper's website and are used as a basis for the story, are not verified by the newspaper. This leads to compromising the journalistic norm of data verification, which is supposed to ensure accuracy in journalism. This is how the old norms are being ruptured by the pressure of the logic of the crowd, which manifests itself as an uncontrollable amount of unverifiable information. In open journalism, new norms and ideals are thus emerging, one of them being shared responsibility for the truth of what is published. It is not only journalists who have the responsibility of data accuracy, but the responsibility is shared with the readers, the collaborators in the journalistic process. This type of blended responsibility is an example of how the existing norms and ideals need to be modified so that journalism can use crowdsourced information. These new ideals, however, lead journalism to vulnerability, that vulnerability being the potential of publishing inaccurate information.

Open journalistic methods such as crowdsourcing appeal to journalists because of the provided efficiency in the knowledge search. By casting the nets widely in the water by using crowdsourcing, the likelihood that at least one of the readers is in the knowledge neighborhood and can access the information that the journalist needs to pursue her or his story increases. However, open methods have serious limits, like the digital divide, sensitivity of the topic, and making journalism vulnerable to false information. Therefore, the journalistic agenda should not be molded simply to fit the methods of open journalism. Instead, the agenda should follow news criteria and the guideline of relevancy to society, covering also those that are the most vulnerable and who cannot speak for themselves, and thus cannot participate in crowdsourced journalism.

To this end, open journalistic practices are a double-edged sword for professional journalism: open journalism is both a benefit and a challenge. It is a benefit in that it creates undeniable value to journalism by bringing in relevant information and helping journalists to make progress in their investigations. But open journalism also challenges the journalistic norms and ideals, and it pressures journalists to compromise some of the most fundamental ideals of journalism, like that of accuracy. Open journalism is also limited to the online crowd as an immediate pool of participants, and it can most easily be applied in investigations that exclude sensitive issues and suppressed groups, thus narrowing down the agenda of topics. Therefore, open journalistic practices should be used carefully in digital journalism and with a deep awareness of the method's risks and limitations, although taking risks can be worthwhile because of the undeniable benefits of open practices to journalism.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author.

FUNDING

This work has been in part enabled by a grant from the Helsingin Sanomat Foundation.

NOTES

1. Reader-relationship refers to the connection between the readers and the publication (Ferguson 1983; Gough-Yates 2003).
2. Sometimes the term citizen journalism is used broadly to refer to commenting on and linking to news rather than producing full journalistic pieces. Thus the boundaries of citizen journalism are blurry (Goode 2009; Lasica 2003; Nip 2006). Participatory journalism typically refers to professional journalism, which involves readers; however, in Bowman and Willis's (2005) definition, for instance, participatory journalism is based on independent civic activity, and there is no journalistic intervention.

REFERENCES

Afuah, Allan, and Christopher L. Tucci. 2012. "Crowdsourcing As a Solution to Distant Search." *Academy of Management Review* 37 (3): 355–375.

Aitamurto, Tanja. 2011. "The Impact of Crowdfunding on Journalism." *Journalism Practice* 5 (4): 429–445.

Aitamurto, Tanja 2012. *Crowdsourcing for Democracy: New Era in Policy–Making*. Publications of the Committee for the Future, Parliament of Finland 1/2012. Helsinki, Finland.

Aitamurto, Tanja. 2013. "Balancing Between Open and Closed." *Digital Journalism* 1 (2): 229–251.

Aitamurto, Tanja, and Hélène Landemore. 2015. "Five Design Principles for Crowdsourced Policymaking: Assessing the Case of Crowdsourced off-road Traffic Law in Finland." *Journal of Social Media for Organizations*. 2 (1): 1–20.

Aucoin, James. 2005. *The Evolution of American Investigative Journalism*. Missouri: University of Missouri Press.

Banks, John, and Mark Deuze. 2009. "Co-creative labour." *International Journal of Cultural Studies* 12 (5): 419–431.

Benkler, Yochai. 2002. "Coase's Penguin, or, Linux and 'The Nature of the Firm'." *The Yale Law Journal* 112 (3): 369–446.

Bowman, S., and C. Willis. 2005. "The Future is Here, But Do Media Companies See It?" *Nieman Reports* 59 (4): 5–9.

Brabham, Daren C. 2008. "Crowdsourcing as a Model for Problem Solving: An Introduction and Cases." *Convergence: The International Journal of Research into New Media Technologies* 14 (1): 75–90.

Brabham, Daren C. 2010. "Moving the Crowd at Threadless." *Information, Communication & Society* 13 (8): 1122–1145.

Brabham, Daren C. 2013. *Crowdsourcing. The MIT Press Essential Knowledge Series*. Cambridge: Massachusetts Institute of Technology.

Bradshaw, Paul, and Andrew Brightwell. 2012. "Crowdsourcing Investigative Journalism: Help Me Investigate—A Case Study." In *The Handbook of Global Online Journalism*, edited by Eugenia Siapera and Andreas Veglis, 253–271. West Sussex: Wiley-Blackwell.

Brito, Jerry. 2008. "Hack, Mash & Peer. Crowdsourcing Government Transparency." *The Columbia Science and Technology Law Review*. 119 (9): 119–157.

Bruns, Axel. 2005. *Gatewatching: Collaborative Online News Production*. New York: Peter Lang.

Convertino Gregorio, Ágnes Sándor, and Marcos Báez. 2013. Idea Spotter and Comment Interpreter: Sensemaking Tools for Idea Management Systems. Presented at the ACM Communities and Technologies Workshop: Large-scale Idea Management and Deliberation Systems Workshop.

Dailey, Dharma, and Kate Starbird. 2014. "Journalists as Crowdsourcerers: Responding to Crisis by Reporting with a Crowd." *Computer Supported Cooperative Work (CSCW)* 23 (4-6): 445–481.

Daniel, Anna, and Terry Flew. 2010. "The Guardian Reportage of the UK MP Expenses Scandal: A Case Study of Computational Journalism." In *Record of the Communications Policy and Research Forum 2010*, 186–194. Sydney: Network Insight Pty..

Domingo, David. 2011. "Managing Audience Participation: Workflows, Strategies and Motivations." In *Participatory Journalism: Guarding Open Gates at Online Newspapers*, edited by J. B. Singer, et al., 94–121. Chichester: Wiley-Blackwell.

Domingo, David, Thorsten Quandt, Ari Heinonen, and Steve. Paulussen, Jane Singer and Marina Vujnovic. 2008. "Participatory Journalism Practices in the Media and Beyond." *Journalism Practice* 2 (3): 326–342.

Eisenhardt, Kathleen M. 1989. "Building Theories from Case Study Research." *Academy of Management Review* 14 (4): 532–550.

Ettema, James, and Theodore Glasser. 1998. *Custodians of Conscience. Investigative Journalism and Public Virtue*. New York: Columbia University Press.

Ferguson, Marjorie. 1983. *Forever Feminine*. Women's Magazines and the Cult of Femininity. London: Heinemann.

Fraser, Nancy. 1993. "Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy." In *The Phantom Public Sphere*, edited by B. Robbins, 1–32. Minneapolis, MN: University of Minnesota Press.

Furtado, Vasco, Carlos Caminha, Leonardo Ayres, and Henrique Santos. 2012. "Open Government and Citizen Participation in Law Enforcement via Crowd Mapping." *IEEE Intelligent Systems* 27 (4): 63–69.

Gillmor, Dan. 2004. *We the Media. Grassroots Journalism by the People, for the People*. Sebastopol, California: O'Reilly Media.

Goode, Luke. 2009. "Social News, Citizen Journalism and Democracy." *New Media & Society* 11 (8): 1287–1305.

Gough-Yates, Anna. 2003. *Understanding Women's Magazines*. New York: Routledge.

Hatch Mary Jo and Schultz Majken. 2010. "Toward a Theory of Brand Co-creation with Implications for Brand Governance." *Journal of Brand Management* 17 (8): 590–604.

Holmes, T., and L. Nice. 2012. *Magazine Journalism*. London: Sage.

Hong, Lu, and Scott E. Page. 2004. "Groups of diverse problem solvers can outperform groups of high-ability problem solvers." *Proceedings of the National Academy of Sciences* 101 (46): 16385–16389.

Howe, Jeff. 2008. *Crowdsourcing: Why the Power of the Crowd is Driving the Future of Business*. New York: Crown Business.

Huberman, A. Michael and Matthew B. Miles. 2002. The Qualitative Researcher's Companion. Sage Publications.

Jeppesen, Lars B., and Karim R. Lakhani. 2010. "Marginality and Problem-solving Effectiveness in Broadcast Search." *Organization Science* 21 (5): 1016–1033.

Kauffman, Stuart A. 1993. *Origins of Order: Self-organization and Selection in Evolution*. New York: Oxford University Press.

Lasica, J.D. 2003. "What is Participatory Journalism?" *Online Journalism Review*. Accessed August 2003. http://www.ojr.org/ojr/workplace/1060217106.php

Liu, Sophia B. 2014. "Crisis Crowdsourcing Framework: Designing Strategic Configurations of Crowdsourcing for the Emergency Management Domain." *Computer Supported Cooperative Work (CSCW)* 23 (4-6): 389–443.

Meier, Patrick. 2011. "Do 'Liberation Technologies' Change the Balance of Power between Repressive States and Civil Society?" PhD diss., Fletcher School of Law and Diplomacy, University of Medford, Massachusetts.

Monbiot, George. 2011. "The Need to Protect the Internet from 'Astroturfing' Grows Ever More Urgent." *The Guardian* 23. Accessed April 28, 2015. http://www.theguardian.com/environment/georgemonbiot/2011/feb/23/need-to-protect-internet-from-astroturfing

Muthukumaraswamy, Karthika. 2010. "When the Media Meet Crowds of Wisdom." *Journalism Practice* 4 (1): 48–65.

Nelson, Richard R., and Sidney G. Winter. 2009. An Evolutionary Theory of Economic Change: Harvard University Press.

Nip, Joyce. 2006. "Exploring the Second Phase of Public Journalism." *Journalism Studies* 7 (2): 212–236.

Page, Scott E. 2008. *The Difference. How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies.* Princeton: Princeton University Press.

Piller, Frank T., Christopher Ihl, and Alexander. Vossen. 2011. "A Typology of Customer Co–Creation in the Innovation Process." In *New Forms of Collaborative Production and Innovation on the Internet: Interdisciplinary Perspective*, edited by Volker Wittke and Heidemarie Hanekop, 31–61. Göttingen: Universitätsverlag.

Saldana, Johnny. 2009. *The Coding Manual for Qualitative Researchers*. London: Sage.

Strauss, Anselm L., and Juliet M. Corbin. 1998. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. 2nd ed. Thousand Oaks: Sage Publications.

Vehkoo, Johanna. 2013. *Crowdsourcing in Investigative Journalism*. Oxford: Report. Reuters Institute for the Study of Journalism.

Tanja Aitamurto, Brown Institute for Media Innovation, Stanford University, USA; E-mail: tanjaa@stanford.edu. Web: www.tanjaaitamurto.com