This chapter explores the role of the crowd in journalism for sourcing information and resources. Through crowdsourcing, journalists can harness the crowd’s knowledge for journalism. Crowdsourcing has been widely used as a mechanism for problem solving by companies, and professional journalists have more recently begun to crowdsource knowledge for their stories.

Crowdsourcing channels information to journalists quickly and from a large number of people, thus contributing to the journalistic process in several ways. Crowdsourcing requires opening up the journalistic process, and thus, crowdsourcing becomes an open journalistic practice. Crowdsourcing enhances both knowledge search and discovery, strengthens reader-relationships, and provides journalists with a window to readers’ world and tacit knowledge about their preferences (Aitamurto, 2013, 2015a).

Crowdfunding is a subtype of crowdsourcing, and it channels the crowd’s resources to journalism as funding, with a growing number of stories financed this way. When used to fund journalism, the consequent journalism product becomes a more open product with journalism being a shared economic venture (Aitamurto, 2015b).

Crowdsourcing, however, brings also costs and complexities to the journalistic process. The method requires substantial human resources and also challenges traditional journalistic practices, norms, and ideals. The balance between the benefits and costs create the value of crowdsourcing in journalism. In this chapter, I introduce a framework for analyzing the value of crowdsourcing as an open journalistic practice. I first define crowdsourcing and its boundaries between other methods of large-scale online collaboration and participatory journalistic practices, going on to illustrate how crowdsourcing has been used in journalism, before presenting a framework for analyzing the value of crowdsourcing in journalism.

**Crowdsourcing: definition and boundaries**

Crowdsourcing is increasingly used in journalistic contexts (Aitamurto, 2015c). Crowdsourcing is an open call online for anybody to participate in a task (Afuah and Tucci, 2012; Brabham, 2008, 2013; Howe, 2008), developed and defined by the crowdsourcer, which in the case of journalism is often a journalist or a group of journalists. Crowdsourced eyewitness stories about corruption and election fraud are shared on platforms like Ushahidi (Meier, 2011), and organizations across the world have initiated crowdsourced fact checking, including Full Fact
in the United Kingdom and Chequeado in Argentina. National and local governments have applied crowdsourcing to identify solutions to issues through policies and laws (Aitamurto and Landemore, 2013, 2015; Brabham, 2015). Emergency management organizations crowdsource information in crises such as earthquakes and hurricanes (Liu, 2010; Starbird, 2011), and companies like Eli Lilly and Procter & Gamble also use crowdsourcing to find solutions to their more difficult R&D problems.

Crowdsourcing can be either paid or voluntary. In paid crowdsourcing, members of the crowd are paid for accomplishing tasks typically on virtual labor markets like Amazon’s Mechanical Turk or on oDesk. On these platforms, crowdsourcers post tasks and crowdworkers select which ones they want to accomplish. The compensation is typically very small—a couple of cents per task. Another type of paid crowdsourcing is innovation challenges where companies and organizations reward certain submissions. Organizations post these challenges to innovation intermediaries and solvers are asked to post solutions either in public or in private to the company. In voluntary crowdsourcing, tasks are completed without financial compensation. Crowdsourced journalism, as well as crisis mapping, crowdsourced policymaking, and crowdsourced science, is typically based on voluntary contributions. The crowd participates in voluntary crowdsourcing because of intrinsically driven motivations, such as a sense of civic duty to contribute, ideological reasoning, or the fun derived from the activity (Aitamurto, 2015c; Brabham, 2010, 2012; Lietsala and Joutsen, 2007; Nov, Arazy, and Anderson, 2011).

Crowdsourcing differs from other popular forms of online collaboration, such as commons-based peer production in several fundamental ways. Commons-based peer production is used in open source software production and in Wikipedia creation (Benkler, 2002) and differs from crowdsourcing in terms of the locus of power and the mode of participation. In crowdsourcing, power is always with the crowdsourcer. In crowdsourced journalism, the power lies within the journalist who decides when, where, and how crowdsourcing happens. The journalist also decides how the crowd’s input is used. In contrast, for commons-based peer production, the locus of power is with the online participants, and they have more say over the production process than the crowd does in crowdsourcing. Contributors, such as Wikipedia editors or open source software producers, are often self-organized, and there is less of a hierarchy than in crowdsourcing. It is debatable, however, whether commons-based peer production can ever be fully free from managerial hierarchies. Wikipedia, for instance, while relying on volunteer editors in content production, has a managerial hierarchy in place. Wikipedia is also not a wholly independent endeavor absent any sustaining forces behind it; the encyclopedia is managed and sustained by the Wikimedia Foundation, which has its own managerial hierarchy in place.

The mode of participation in crowdsourcing also differs from commons-based peer production. In crowdsourcing, the crowd is not part of the actual making of the end result—the crowd contributes their knowledge, but it is the crowdsourcer who uses the crowd input to produce the end result. In crowdsourced journalism, the journalist treats crowd input as raw material and weaves it into the story in the ways that she or he thinks is appropriate, but—it is the journalist who writes the story. In commons-based peer production, in contrast, the participants are active in the production of the end product, as with a Wikipedia article or open source software. They collaborate almost in real time and follow as the end product comes to reality. If commons-based peer production was to be used in journalism, it would mean that the contributors would be collaboratively writing a story together using collaborative spaces, such as a Wiki, Google Docs, or Etherpad type of a platform.
Crowdsourcing as an open journalistic practice

When journalists use crowdsourcing in journalism, they open up the journalistic process to the public. Crowdsourcing, thus, can be considered an open journalistic practice (Aitamurto, 2015a). In open journalism, openness comes into play in several parts of the journalistic process. The process can be open to the public in the beginning, like in the previous examples, or it can also be open later in the process, when the journalist wants to source more information from the crowd. The journalist can also ask for ideas for interviewees, subjects of visualizations, and so on. However, in open journalism, the journalistic process is accessible to the public only in certain parts, and the process is never fully open.

When journalists use crowdsourcing, they publish a call online for the information they want from the crowd. One of the best-known examples of crowdsourcing in journalism happened in 2009, when the British newspaper the Guardian crowdsourced a part of the investigation of the British Members’ of Parliament’s expenses scandal. The crowd was asked to categorize digitized expense receipts and thousands of people participated in the process. Since then, crowdsourcing has been used in several high-profile cases across the world. In Sweden, Svenska Dagbladet, one of the leading daily newspapers, successfully crowdsourced mortgage interest rates for an investigation about mortgages in 2013. About 50,000 Swedes submitted their information on a crowdmap on the newspapers’ website, and dozens of stories were written based on the crowdsourced data. In Finland, the leading daily newspaper, Helsingin Sanomat, crowdsourced an investigation of stock short-selling in 2012. The investigation resulted in a scoop about a questionable holding company’s arrangement with a Finnish bank, leading to the firing of a bank executive (Aitamurto, 2015). The British Broadcasting Company (BBC) in the United Kingdom has crowdsourced information for real-time traffic updates and for tracking the consequences of the public budget cuts in cities. The New York Times in the United States successfully crowdsourced information for identifying people in images during the Boston bombings coverage in 2013. WNYC, a public radio station in New York, crowdsourced information for prediction about re-emergent of cicada swarms, by asking the crowd to use soil monitors to track the soil temperature in 2014. Participants submitted the temperature information, which was then displayed on a crowdmap online. Thus, it was possible to track the soil temperature in hundreds of locations simultaneously. Apart from these, largely visible crowdsourcing efforts, newspapers, magazines, and journalists also crowdsource information, such as pictures, about news events and story tips regularly on their websites. Journalists also ask for story ideas and tips through their social networks and beyond.

Crowdsourcing can also be used as a knowledge search method in participatory and citizen journalism. Readers participate in participatory journalism as commentators or content producers (Bruns, 2005; Domingo, 2011; Holmes and Nice, 2012). In citizen journalism, people who are not journalists by profession can create content that can be considered as journalism (Gillmor, 2004). This content can be published on independent blogs and citizen-run news sites or even in publications run by professional journalists. Participatory or citizen journalism, however, does not use crowdsourcing by default as a knowledge search method. Citizen journalism often resembles commons-based peer production more than crowdsourcing as there is not necessarily a crowdsource who controls the process. Furthermore, the locus of power in citizen journalism published on institutionally independent websites (not affiliated with professional news sites), is within the peer producers who are more equal than in crowdsourcing.
Crowdfunding in journalism: aggregated donor judgments

Crowdfunding is a subtype of crowdsourcing. In crowdfunding, the crowdsourced task is to gather money for a certain purpose, rather than information, and in crowdfunded journalism, the task is to gather funding for a story pitched by a journalist. Donors show their support for a journalistic story idea by donating money, and donors’ judgments are aggregated and accumulated into funding. These aggregated judgments are one manifestation of collective intelligence in crowdfunded journalism (Aitamurto, 2011, 2015b). Collective intelligence refers to the distributed talent and knowledge of large crowds (Lévy, 1997; Landemore, 2013).

Crowdfunding in journalism is typically ex ante crowdfunding: The crowd is asked to contribute to the story before it is produced. A journalist organizes a campaign for crowdfunding a story and asks the crowd to donate money so that the journalist can accomplish the project. The campaigns are often run on dedicated crowdfunding platforms like Beacon, Kickstarter, and IndieGoGo. The funders often get rewards for their donation. The value of the rewards rises in tiers, and the more the funder gives, the more valuable the reward is. The rewards are typically related to the article, for example, a handwritten note from the journalist, a signed photograph, or a T-shirt. In ex post facto crowdfunding, the funds are raised for a completed product. Ex post facto crowdfunding is more of a digital tip jar (Kappel, 2008). Funds can be raised for a story that is available for free online anyway, for instance, by using digital tip jars like Kachingle or Flattr.

There are four types of crowdfunding in journalism: Fundraising for a single story, fundraising for continuous coverage or beat coverage, fundraising for a new platform or publication, and fundraising for a service that supports journalism.

When a journalist raises funds for a single story, she or he sets up a campaign online for crowdfunding. In this model, the journalist pitches a story online for potential funders. The pitching often happens on a specific platform like Kickstarter, Contributoria, or Beacon. (Aitamurto, 2015b). Raising funds for a single story used to be the most common form of crowdfunding, but fundraising for a continuous beat is becoming increasingly common. In this type of crowdfunding, a journalist funds a more continuous coverage by crowdfunding. This type of crowdfunding is thus both ex ante and ex post facto crowdfunding: The funds are raised both for future stories about the beat, but also based on already published work on the beat. On the crowdfunding site Beacon, journalists raise funding for continuous coverage of topics like environmental issues and mass incarceration. For instance, during the Ferguson unrest in Missouri in the United States in 2014, a journalist gathered funding on Beacon to travel to Ferguson and undertake reporting there. The fundraising was ongoing across several weeks of unrest, and she was able to cover the news event for the whole period, thanks to successful crowdfunding. Crowdfunding for a continuous beat recalls artistic patronage in that the journalists ask the crowd to support their work in a more long-lasting form than just one story at a time. In the traditional artistic patronage, typically a small number of supporters donate large sums. In crowdfunded patronage, the support comes from a large crowd in small amounts instead.

Benefits of crowdsourcing: knowledge discovery, peer learning, and deliberation

Through crowdsourcing, journalists have access to a larger number of people than ever before. The online crowd is a huge potential source pool for the journalist with the size of the participant
crowd, in theory, being almost limitless. In practice, limited access to the internet, language, and skill barriers make the process accessible only to a certain number of people. It is also limited in terms of the number of people who have knowledge about the open journalistic process and can thus participate.

Using crowdsourcing, journalists can tap into the collective intelligence of the crowds and channel that to their articles. The larger the number of participants, the more likely the journalist is to find useful information. Crowdsourcing hence supports knowledge search in journalism. Empirical studies show that crowdsourcing in journalism can bring in relevant information that helps journalists to proceed with their investigations and that the crowdsourced knowledge search can lead to a fast knowledge discovery and even to a discovery of knowledge that the journalist did not know to search for (Aitanturko, 2013, 2015a).

Journalists can enhance the knowledge search aspects in crowdsourced journalistic processes by using several methods, depending on the goals of their investigations. By answering the question or prompt, online participants contribute to journalistic investigations. For example, in the crowdsourced mortgage interest investigation in Sweden, the crowd was asked to share their home loan interest rate by filling out a simple Google form online. The rates would then appear on the newspapers’ website as a large crowdmap. In another case, a science journalist in a science magazine in Finland crowdsourced an investigation of gender bias in math and science education. She developed provocative prompts about common beliefs regarding the different genders’ ability to learn math and science. She then asked the crowd to react to these and share their experiences about math and science education. After analyzing the crowd’s input, she asked experts’ opinion about the common beliefs and then incorporated the crowd’s and experts’ contributions to her article.

Journalists can also publish documents related to the investigations online and ask the crowd to check those documents and then submit their observations to the journalist. For example, Svenska Dagbladet, one of the leading daily newspapers in Sweden, crowdsourced an investigation of development aid use in Swedish municipalities. As part of their investigations, journalists published online documents about development aid and asked the crowd to read those documents and report back to the journalist if anything interesting was found. Similarly in the investigation of stock short-selling, the leading daily newspaper in Finland, the Helsingin Sanomat, asked the crowd to check hundreds of stock trading reports that the newspaper had published online. In these cases, the crowd identified relevant issues in the documents, and by reporting those back to the journalist they enabled the journalist to proceed with their investigations (Aitanturko, 2015a).

Journalists can also apply co-creation, a type of crowdsourcing, in their open journalistic process. Co-creation is a more interactive method in which the crowd and the crowdsourcer work collaboratively to create content or solutions (Piller, Ihl, and Vossen, 2011). In co-creation, the crowd is invited to participate in several parts of the story process—in contrast to basic crowdsourcing where the crowd participates in only one part of the process, and participation is a one-time act. One example of this is the Finnish lifestyle magazine Olivia, which co-created several magazine issues with their readers. The crowd was asked to participate in several parts of the story process, starting from developing topics, choosing the topic, transforming sources, and shooting contexts (Aitanturko, 2013). Co-creation involves a lot of interaction between the participants and the journalists and strengthens the reader relationship with the publication and the journalists. The online participants feel a strong sense of belonging to the journalistic process and product, even prompting participants to subscribe to the magazine (Aitanturko, 2013).
Crowdsourcing when used as an open journalistic practice is never free. The method comes with many costs. Most of the costs are incurred in the synthesis and aggregation part of crowdsourcing, as analyzing and synthesizing a large number of submissions requires a good deal of human resources, especially when the data are unstructured. Other aspects of a crowdsourced journalistic process are also demanding on human resources, along with technical costs incurred. In this section, I introduce a five-point framework outlining the benefits and costs in certain parts of a crowdsourced journalistic process. The framework is summarized in Table 18.1.

First, the better the crowdsourced process is prepared, the more likely it will succeed. However, preparing a crowdsourcing initiative requires resources. To mention just a few of the preparative tasks: there might be documents that need to be digitized for publishing; the crowdsourcing platform has to be set up, whether it is a specific platform or a website on the publications’ website; and when preparing a crowdfunding process, the campaign and the rewards have to be produced, and so on.

Second, the overall process requires attention. If the process is designed for interaction between journalists and participants, the journalists need to be present on the platform and interact with the participants. If there is horizontal transparency in the process, there needs to be either pre- or post-moderation on the platform for removing inappropriate input. The prompt for participation may require iteration based on early user feedback, or, if nobody participates or the participation is low, the prompt might need to be revised. Fast and smart iteration can save the whole process.

Third, the crowd's input needs to be analyzed, evaluated, and filtered. This is typically the most laborious part in crowdsourcing. Journalists often crowdsource unstructured data—that

<table>
<thead>
<tr>
<th>Process part</th>
<th>Description</th>
<th>Benefit</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preparation</td>
<td>Process planning, digitizing documents, and setting up the technology</td>
<td>The more and better designed probes, the more likely there will be useful crowd input</td>
<td>Human resource costs, technical cost</td>
</tr>
<tr>
<td>Interaction, moderation, and iteration</td>
<td>Interaction with online participants, moderating the comments, and iterating the prompt</td>
<td>Interaction leads to a stronger reader-relationship, deliberation, peer learning, and iteration to a better process</td>
<td>Human resource costs and technical costs</td>
</tr>
<tr>
<td>Analysis and evaluation</td>
<td>Analyzing the crowd input, evaluating what are useful and relevant data</td>
<td>Analysis and evaluation filters out useful information</td>
<td>Human resource costs and computational analysis</td>
</tr>
<tr>
<td>Verification</td>
<td>Verifying the accuracy of the crowd input</td>
<td>Cross-checking shows what is correct data</td>
<td>Human resource costs</td>
</tr>
<tr>
<td>Synthesis and aggregation</td>
<td>Synthesizing the crowd’s input, channeling input to article</td>
<td>Synthesis brings out the diverse voices of the crowd shows diverse perspectives and contradicting facts</td>
<td>Human resource costs and technical costs</td>
</tr>
</tbody>
</table>
is free-form online comments as responses to their questions. The data are qualitative, and the format and content varies, so the analysis is typically conducted so that the journalist peruses all of the comments. When the data are structured quantified data, they can be analyzed faster by automated computations. In the home loan interest investigation in Sweden, for example, the data were people’s home loan interest rates, so they were all quantified. Thus, the journalists could analyze the data at the newsroom through simple and fast computations. Similarly, in crowdfunded journalism, the aggregation of the crowd’s input—the donations—is easy, because it is all quantified contributions and they are all in the same format.

Fourth, the crowd’s input needs to be verified before journalists can use the information in news articles. Journalists use traditional cross-checking procedures to verify the accuracy of the crowdsourced information; they call the sources, check the documents, and so on. For instance, in the crowdfunded stock short-selling case, the journalist checked the accuracy of the crowd observations of the selling reports by investigating on his own the particular stock reports that the crowd highlighted. When the volume of crowd submissions is not too high, it is possible to undertake fact checking using traditional human resource-intensive journalistic practices. However, when the number of submissions is high, verification proves to be impossible—as with the home loan interest case where the journalists received 50,000 contributions. In this case, journalists first tried to verify the data by calling the participants, but after about 80 phone calls they had to give up. Nevertheless, they decided to use the data and were clear in the articles that the data were unverified, thus pushing the responsibility about accuracy onto the public. Consequently, journalists took the risk of compromising one of the core norms of journalism, that the reported facts are accurate. However, this has resulted in a new norm that blends responsibility between the crowd and the journalist (Aitamurto, 2015a). This also illustrates that the volume of participation in crowdfunded journalism can turn into a paradox: The more actively the crowd participates, the better, but the more difficult it makes the data verification.

Finally, the synthesis of the crowd’s input is a very important part of the process. This is the point when the journalist decides what information to use and how to use it. For instance, in a story on gender bias in math and science education, the journalist categorized the crowd’s input and mapped the controversies, attempting to create a balanced and accurate overview of the crowd’s input to the article. Successful synthesis requires understanding that the crowd’s input is rather like raw material for the story and cannot or should not be used ‘as is’ in the stories. When the volume of submissions is high and there are many controversies and broad diversity in the voices and opinions, it is challenging and time-consuming to synthesize the input. Synthesis also requires understanding that crowdfunded input always comes from self-selected group of people, so there is a selection bias. This means that the crowd’s input cannot be taken as the de facto public opinion about an issue, as it is not a random sample of the population.

The value of crowdfunded journalism: balance between benefits and costs

When used as an open journalistic practice, crowdfunding generates undeniable benefits to professional journalism. It can accelerate knowledge gathering, widen the journalists’ perspective, and create a stronger reader-relationship, as the examples presented in this chapter illustrate. The crowd also learns from each other, and that peer learning contributes to wider knowledge about the issue the journalist is investigating. Crowdsourcing thus has a potential to create value in several ways as an open journalistic practice.

However, crowdsourcing also comes with challenges. One of the main difficulties concerns the analysis and evaluation of the crowd’s input and verifying accuracy. This is where traditional
journalistic practices are in conflict with the nature of crowdsourcing. Crowdsourcing delivers information quickly from large crowds, yet the journalistic practices used to process and verify inputs can only approach small amounts of data at a time. Therefore, analysis, synthesis, and verification of crowdsourced input in open journalism often collide with traditional journalistic practices. They also collide with the ideals of journalism, which strive for verified information to ensure data accuracy.

The benefits of open journalism are partially compromised, then, by the costs of open practices. As shown in Table 18.1, planning, implementing, analyzing, verifying, and synthesizing the crowdsourced input require a considerable amount of human resources for post-processing of that input, especially when the amount of submissions is high.

New technologies have proved helpful in resolving the bottleneck of analysis and synthesis of the crowdsourced input. By training algorithms in machine learning and developing better Natural Language Processing technologies, some of these issues may be resolved, as the processing can be made more automatic. However, even those technologies do not resolve the challenges in data verification. Perhaps it is the case that to benefit from open journalistic practices like crowdsourcing certain norms and ideals of traditional journalism need to be left behind, to be replaced by new ideals, like that of blended responsibility in open journalism.

Further reading


References


journalistic practices are in conflict with the nature of crowdfunding. Crowdsourcing delivers information quickly from large crowds, yet the journalistic practices used to process and verify inputs can only approach small amounts of data at a time. Therefore, analysis, synthesis, and verification of crowdsourced input in open journalism often collide with traditional journalistic practices. They also collide with the ideals of journalism, which strive for verified information to ensure data accuracy.

The benefits of open journalism are partially compromised, then, by the costs of open practices. As shown in Table 18.1, planning, implementing, analyzing, verifying, and synthesizing the crowdsourced input require a considerable amount of human resources for post-processing of that input, especially when the amount of submissions is high.

New technologies have proved helpful in resolving the bottleneck of analysis and synthesis of the crowdsourced input. By training algorithms in machine learning and developing better Natural Language Processing technologies, some of these issues may be resolved, as the processing can be made more automatic. However, even those technologies do not resolve the challenges in data verification. Perhaps it is the case that to benefit from open journalistic practices like crowdsourcing certain norms and ideals of traditional journalism need to be left behind, to be replaced by new ideals, like that of blended responsibility in open journalism.

Further reading


References


