Confidentiality in a world of open data: How OSINT journalists safeguard privacy

Woude, Maartje van der

Confidentiality in a world of open data: How OSINT journalists safeguard privacy

Maartje van der Woude – s3674894
Leiden University

Abstract

As the volume of openly available digital data continues to grow exponentially daily, open-source intelligence methods have gained significant traction within the field of journalism, particularly for investigative news reporting. However, while this practice holds great potential for advancing journalism, it introduces a paradox between the ethos of transparency inherent in journalism and the privacy and security concerns of data subjects and brokers. This article delves into this contradiction by drawing on data from in-depth interviews with eight professional open-source journalists, shedding light on the underlying factors at play.

By examining how open-source journalists perceive and navigate privacy issues while striving to achieve their investigative goals, it becomes evident that these journalists possess considerable power and awareness regarding the security and privacy of human data subjects and brokers. The findings of this study reveal that open-source journalists rely heavily on personal assessments and ongoing dialogues with colleagues to make privacy-related editorial choices, as there is a dearth of established rules and guidelines in this domain. Consequently, open-source journalists frequently engage with their organizations' legal departments as they harbor concerns about potential legal ramifications.

This research provides valuable insights into the intricacies of open-source journalism, uncovering the delicate balance between journalistic transparency and privacy/security considerations. It underscores the need for further development of privacy-related frameworks and guidelines specific to the realm of open-source journalism while highlighting the pivotal role of open-source journalists in shaping the privacy landscape within their field.

Keywords; OSINT, journalism, privacy, transparency, open data

Introduction

Journalists are turning to open-source intelligence tools more frequently as digital spaces provide access to a greater amount of data (Westcott, 2019). These tools allow journalists, especially investigative reporters, to dive deep into big data sets to conduct advanced public interest investigations and tell stories that otherwise would not be told (Müller & Wiik, 2023). However, OSINT tools and the journalists using them are not exempt from the ethical debates surrounding big data. Although open-source initiatives in journalism aim to contribute to public value, questions about journalists' perceptions of transparency, privacy, and security when using open-source data and methods remain largely unanswered (Meijer et al., 2013).

The omnipresence of openly available civilian photos, audio recordings, and videos is a feature of our times. It forms a boundless visual archive of civilians' personal data and everyday lives, including those residing in conflict zones (Saugmann, 2019). Citizen-produced images and videos are digital eyewitness accounts of human conflict and are especially valuable to stay informed on areas that are hard to reach (Mast & Hanegreefs, 2015). Today, journalists can digitally report on the events of dangerous battlefields despite the remoteness and inaccessibility of these conflicts (Müller & Wiik, 2023). However, as Saugmann (2019) argues, exploiting civilians' images for conflict-related purposes - including journalistic reporting - turns these people and their images into active actants in these conflicts without guaranteeing the civilian's knowledge and consent. Not surprisingly, adopting open-source data in the form of civilian content can pose serious privacy challenges and harm to the individuals involved (Koops et al., 2013; Eijkman & Weggemans, 2013; Meijer et al., 2013).

Privacy issues emerging from the intersection between journalism and open-source intelligence are not limited to the use of civilian imagery but also arise when dealing with big data (White, 2016). In a computational sense, big data refers to large datasets that cannot be processed by standard computer memory and software (Lewis & Westlund, 2015). Big data sources can include governmental records and databases, commercial databases that aggregate individual data from commercial transactions and public records, and geospatial data (Agnellutti, 2014). Where numerical data sets are not automatically meaningful or comprehensible to everyone, they can contain journalistic value for they reveal undisputed truths when granular, complete and regularly updated (Parasie and Dagiral, 2013). Subsequently, it is the task of the (data) journalist to turn the static numbers into comprehensible news items.

Big datasets can contain sensitive information about individuals, such as consumption patterns, personal health, or sexual preference (Agnellutti, 2014). Therefore, even when dealing with legal and openly available big datasets, open-source journalists should be careful not to violate people's privacy when processing and publishing them (Pastor-Galindo et al., 2020).

Violating privacy rights becomes even more probable when open-source journalists deal with gray information, that is, data gathered from legal or semi-legal sources that are not widely distributed and often of a questionable nature (Hribar et al., 2014). Generally, gray information is acquired from sources that are defined as open but, in fact, are not entirely so. For example, gray information can include leaked databases of social media accounts, 'inside information' of a company's personnel, or private publications by researchers (Hribar et al., 2014).

Questions about responsibility and liability arise when open-source journalists employ sensitive content or gray information. Who decides whether to use certain content? Are there ethical or legal guidelines to fall back on? Do open-source journalists receive training to conduct investigations safely? According to Edwards (2023), the number of resources dedicated to the ethics of media open-source research is increasing. Workbooks, manuals, and guidelines are being developed to assist open-source journalists in navigating these ethical questions. Nevertheless, it is unclear if and how open-source organizations and newsrooms adopt these resources: online newsrooms, in general, remain "woefully under-researched" (Manninen, 2017) in these areas.

Furthermore, central to the open-source ideology is the emphasis on transparency of sources and methods (Hammond, 2017), which supports the replicability of open-source investigations and is expected to lead to increased public trust and legitimacy (Meijer et al., 2013). A tension exists between the transparent nature of the open-source journalistic movement and the privacy needs of civilians and open-source journalists themselves. Therefore, gaining insight into how open-source journalists balance privacy concerns with their investigative goals is ever more relevant. This article seeks to explore these balances by probing the following questions: (RQ1) How do open-source journalists understand and navigate privacy issues and balance them with their investigative goals? (RQ2) What measures do open-source journalists take to protect the privacy and security of open-source data providers and brokers? (RQ3) What stance do open-source journalists take when it comes to using gray open-source information? And (RQ4) What is the role of newsrooms

when open-source journalists deal with privacy issues? These questions are explored by interviewing eight open-source journalists.

Literature review

The practice of open-source investigation draws on research based on publicly accessible information (Edwards, 2023) and is characterized by the belief in social responsibility through openness and freedom (Coleman & Golub, 2008; Coleman, 2012; Lewis & Usher, 2013). Furthermore, open-source practices involve synthesizing freely available information into actionable forms (Glassman & Kang, 2012). In essence, open-source data entails all publicly available information, from Tweets to governmental reports and from selfies to satellite imagery. The vastly varying types of open-source data can serve the highly diverse purposes of national intelligence agencies, corporate entities, academia, start-ups, and journalists (Westcott, 2019).

During the 1990s, the open-source movement emerged from the hacker community, which placed a strong emphasis on experimentation, play, and democratic ideals. According to Hansen (2015) and Coleman (2012), these individuals were motivated by a pro-social interest in information liberation and the free flow of knowledge. Thus, the primary objective of the open-source movement is to ensure that information is freely available to all, without requiring individuals to pay for access to software codes or data (Hansen, 2015). Instead, the movement advocates for generating revenue by offering services and practices that are based on open-source information (Young, 1999). In the early days of the internet, easy access to data and tools predominantly interested state actors and corporate departments at one end, and social media hobbyists on the other end (Edwards, 2023). Corporate entities used open-source data and tools to compile risk assessments, whereas social media hobbyists scoured images and videos to fact-check the claims that parties in armed conflicts had made. From 2010 onwards, digital news aggregator initiatives (i.e., Forensic Architecture, Bellingcat, and Syrian Archive) were established and picked up on the possibilities that open-source data and tools offer to accurately map out human rights violations and report on them (Edwards, 2023).

The enormous amount of easily accessible data and the increase in technical possibilities that enable open-source intelligence brought an entirely new dynamic to journalism (Muller & Wiik, 2023), particularly to investigative journalistic practice and war reporting (Edwards, 2023). This new dynamic consists of new digital tools, methods, and meeting points to conduct advanced

investigations and collaborate with other investigative reporters across borders (Carson, 2021). Open-source methods have led to what Muller and Wiik (2023) describe as "the collaborative turn in investigative journalism"; online, investigative journalists are part of wide networks of colleagues, often working with the same data and creating stories collectively. The benefits of working collaboratively and digitally include sharing costs and information, increased story reach, and the allowance for more complex reporting on a global scale (Carson & Farhall, 2018).

Even though open-source ideology and methodology stem from 'hacker culture' and are intertwined with the internet's historical development (Kelty, 2020), some of its normative values correspond to those of journalistic culture (Lewis & Usher, 2013). One mutual key value between these two cultures is *participation,* which is not historically part of the normative framework of journalism but emerges as part of the journalistic ethos for the digital age (Lewis, 2012; Lewis & Usher, 2013). Just like within open-source ideology, journalistic participation translates into the suggestion that consumers take on a more active, monitorial, and interlinked role - helping to supervise the news, instead of merely commenting on post-publication (Lewis & Usher, 2013). Instead of treating news as an end-product, open-source practices turn journalism into a participatory process to which users can meaningfully contribute (Robinson, 2011). A second normative value that open-source ideology and journalism share is *transparency,* which constitutes notions of accuracy and sincerity amongst news consumers (Blood, 2002; Singer, 2007; Phillips, 2010). In open-source journalistic practice, *disclosure transparency* is the norm, which is when journalists explain how they select and produce news in detail (Karlsson, 2010). Through sharing which sources were used and which steps were undertaken, transparency leads to the replicability of the investigation, enabling people to trace back a story and fact-check it themselves (Phillips, 2010). Just like in academics, journalistic replicability leads to a sense of accuracy and legitimacy among consumers (Eijkman & Weggemans, 2013). At a time when the markers of journalistic authority – monopoly of news selection; objectivity; commitment to democracy – do not hold self-evident legitimacy anymore, transparency is increasingly viewed as able to retrieve this authority (Perdomo & Rodrigues-Rouleau, 2022). Enabling and innovating journalistic authority and legitimacy through transparency is a promising feature of open-source journalism. Despite transparency leading to notions of accuracy and sincerity, however, the increased openness that it causes can lead to security violations and privacy breaches (Meijer, Conradie & Coenni, 2013).

The civilian visual security paradox

Defining privacy as "freedom from unreasonable constraints on the construction of one's identity" (Agre & Rotenberg, 1997, p. 7) or "the ability to be free from disturbance or observation," (White, 2016, p. 3), calls into question the roles and responsibilities of open-source journalists and organizations when using personal data, because digital open-source investigations can lead to exposing people's identities and locations, making them prone to (governmental) observation or other forms of disturbance in their everyday lives (Dubberley & Ivens, 2022).

What Saugmann (2019) terms the *civilian visual security paradox* is exemplary of how open-source journalists can put civilians in danger. The *civilian visual security paradox* describes how images and videos that civilians in conflict areas post to call attention to their circumstances, can quickly turn into sources of danger for the civilian when open-source journalists fail to deal with this content in safe ways. Consequently, Saugmann (2019) argues that open-source investigators must "respect the protected status of civilians in their online collection practices – so far, however, there is little sign of such respect" (2019, p. 344).

When asking journalists if and how they protect the people that are behind the content that they use, clarification on potential privacy and safety risks that content producers face is at its place. First, people that are visible in (sensitive) videos or images are potentially at risk when an open-source journalist decides to use such content. For example, if a video of a protester against a dictatorial regime is posted on a social media account and embedded in an open-source news item, it might lead regime supporters to track and punish this person. Even when a subject of content posts it themselves, journalists that embed the content must respect the user's privacy (Pastor-Galindo et al., 2020), for making content publicly accessible is not equal to asking for it to be distributed, aggregated, or otherwise scaled (Boyd 2010). Especially when embedding content of relatively unknown internet users, what Bellingcat investigator Giancarlo Fiorella (2021) refers to as 'the spotlight effect' can occur. The spotlight effect takes place when 'unnoticed' content is embedded in open-source news stories and through reaching a large audience ends up going viral. Consequently, a 'spotlight' is cast on the publisher of the content, which, for example, can lead to unwanted exposure, privacy breaches and other safety risks. Apart from the subjects of data, people that are not visible in an image or video but are part of the subjects' group might be put at risk, for example when their geolocation gets tracked (Dubberley & Ivens, 2022).

Dealing with personal content in investigative processes and news items raises an ethical question around social media authors (Suomela, Chee, Berendt, & Rockwell, 2019): should their content simply available for anyone to use, or should it be treated as the product of human participants, making research on them subject to ethical boundaries and informed consent? (See e.g., Rambukkana, 2019). According to Gauthier (2002), journalists' decisions to publish potentially harmful content are made based on a 'balancing test' that compares the potential harms and the potential benefits. Answering RQ2 will point out how open-source journalists deal with ethical questions around the use of personal content.

Gray information

To complicate legal and ethical open-source issues even further, open-source journalists sometimes use data and methods that are considered (on the verge of) illegal (Hribar, Podbegrar, & Ivanuša, 2014). Although open-source journalism is all about reporting based on openly available data, there are cases in which the line between open and closed data becomes blurry. According to Hribar, Podbegrar, and Ivanuša (2014), there exists a 'gray zone of open-source intelligence,' where *gray information* resides: semi-legal information, that is generally not distributed widely, and often of a questionable nature. Examples of such information include 'inside information' from a company's personnel, the contents of leaked databases, and videos, images, and messages taken from closed or private digital networks, like Telegram groups. Generally, gray information entails data that was meant to be closed but made openly available through hacking and leaking. Naturally, gray data often contain personal information that is not meant to be publicly available – when used in investigative journalistic processes and news items, security breaches and privacy violations are no exceptions. In legal terms, the "gray zone of open-source intelligence" is an area where community interpretation and legal interpretation intersect (Hribar, Podbegrar & Ivanuša, 2014). So far, there is no consensus on ethical guidelines regarding the use of gray information (Rambukkana, 2019), and neither is it known how open-source journalists approach its use. General literature on news-making processes shows that journalists rely on their 'gut feeling,' meaning that their news-making judgments appear self-evident and self-explanatory to them (Schultz, 2007). Through investigating RQ3 – *what stance do open-source journalists have towards gray information?* – the ways in which open-source journalists deal with gray information will be explored.

The social structure of open-source journalistic organizations

The relatively new nature of open-source journalistic practice means that there is limited academic literature on the structures of open-source organizations and newsrooms (Ganguly, 2022). Research on open-source ethos, however, points out that journalistic organizations that conduct open-source investigations emphasize communication, transparency, and strong bonds between members (Belghith, Venkatagiri & Luther, 2022). Explicit rules about the carrying out of specific tasks, and which techniques or tools should be used, are often non-existent within such organizations. In correspondence with the overarching ethos of participation, opportunities to increase open-source investigation skills – i.e., through training and workshops – are often at hand within open-source newsrooms (Belghith, Venkatagiri & Luther, 2022). Whether open-source training and workshops contain instructions on dealing with ethical and privacy issues remains under-researched. General literature on the role of the newsrooms in dealing with journalistic privacy issues shows that codes of ethics serve as crucial accountability tools that "every major professional organization has adopted and revised: individual news organizations build their own codes to clarify ethical expectations for employees" (Whitehouse, 2010, p. 313). Answering RQ4 - *What is the role of the newsroom when open-source journalists deal with privacy issues?* - will attempt to point out whether this is also true in open-source organizations and newsrooms.

Methodology

To understand how open-source journalists balance privacy concerns with their investigative goals and navigate privacy-related ethical and legal considerations, the lead author for this article conducted semi-structured interviews with eight Dutch open-source journalists (n=8) during May of 2023. The interviews were, on average, 52 minutes long and conducted via Zoom. The interviews with all participants were conducted in Dutch due to the expectation that they would be able to express themselves more easily in their native language than in English.

Exploring how open-source journalists balance privacy concerns with their investigative goals asks for in-depth inquiry that can be provided by the results of qualitative interviews, which are guided conversations in which the researcher carefully listens to the meanings that participants attach to the research subject(s) (Gubrium & Holstein, 2002). Specifically, the semi-structured interview approach is adopted. Semi-structured interviews are conducted based on an interview guide, which

is a list of questions or topics to be covered (Bryman, 2016). Semi-structured interviews allow the researcher (interviewer) to depart from the used schedule or guide, for example, when interviewees 'ramble,' to gain a rich understanding of what the interviewee deems relevant or necessary (Bryman, 2016). The interview guide (see appendix A) was prepared before conducting the interviews, which were audio-recorded and transcribed verbatim. The audio recordings of the interviews were anonymized and remained confidential.

Open-source journalists had to fulfill some criteria to fit the sample. First, conducting open-source investigations had to be their main work-related practice. This essential condition stems from the expectation that full-time open-source journalists are likely to be experts that actively participate in the OSINT community, abide by the values and ethos of the movement, and are well-versed in its techniques (Belghith, Venkatagiri, & Luther, 2022). Whereas open-source practice is relatively new to the realm of journalism, it is likely that journalists that do open-source investigations 'on the side' - or more in a novice manner - might not be as aware of privacy issues as full-time open-source journalists would be. Second, open-source journalists must work for a journalistic organization that carries out open-source investigations. This criterium has been established to be able to answer RQ3 (What is the role of the newsroom when open-source journalists deal with privacy issues?).

In the manner of Belghith, Venkatagiri, and Luther (2022), the lead author of this paper recruited open-source journalists through purposive and snowball sampling. Recruitment of participants started with purposive sampling through online requests. Upon finding willing participants, snowball sampling was adopted: Participants were asked to nominate other expert open-source journalists. Despite serious attempts to recruit more female participants, the sample comprised seven men and one woman, all of whom are professional investigative journalists working with open-source data, methods, and tools at open-source newsrooms. Literature shows that women are underrepresented in open-source journalistic practices (De Vuyst, 2020), which again became evident during the sampling process of this study. The participants represented three different (Dutch) journalistic open-source organizations or newsrooms: NOS, Pointer and Nieuwscheckers. NOS is the national Dutch broadcaster and is funded by the Dutch government. NOS has a specific open-source editorial department, called NOS Osint. Pointer is the open-source and data editorial department of KRO-NCRV, a renowned Dutch public broadcaster. Nieuwscheckers is a Dutch

editorial initiative, funded by the University of Leiden. Nieuwscheckers' journalists use open-source data and methods to establish their news items.

The raw data gathered by the interview was processed and ultimately categorized into themes through open coding and axial coding. First, open coding took place, which yields concepts that consequently are grouped together and ultimately are turned into categories (Bryman, 2016). This way, textual data from the interviews were broken up into codifiable parts. Second, axial coding led to the establishment of connections between codes to create categories. The codebook that was established during this process is presented in Appendix B.

Table 1. Characteristics of the sample.

Participant code	*Organization*
P1	NOS
P2	Pointer
P3	NOS
P4	Pointer
P5	NOS
P6	Pointer
P7	Nieuwscheckers
P8	NOS

Results

This article explores how open-source journalists balance privacy concerns with their investigative goals. It does so by investigating what measures open-source journalists take to protect the privacy and security of data subjects and brokers, what their stance towards gray information is, and what the role of their organization is when dealing with privacy issues. The first subsection addresses RQ2, showing that open-source journalists are aware of their power over data subjects' and brokers' privacy and security situations. However, the degree to which they feel responsible for safeguarding others' privacy and security differs. Measures that open-source journalists take to protect privacy are the altering of images and the encryption of data they share with colleagues. The second subsection answers RQ3 and reveals three prevalent categories regarding open-source journalists' stances towards the use of gray information: supportive of, undecided towards, and opposed to. RQ4 is answered in the third subsection. Findings on the role of news organizations

in open-source journalists' efforts to deal with privacy issues include active inter-organizational involvement and a lack of privacy-related guidelines.

Altering images and encrypting shared data

As conveyed by the literature, open-source journalists collectively stress that the privacy and security situations of individuals that provide or star in open-source data can negatively be affected by their actions. Nonetheless, the degree to which open-source journalists feel responsible for protecting these individuals differs. One recurring participant opinion is that if other (international) news organizations have already embedded certain content, potential 'damage' has already been done, so the participant's responsibility of safeguarding the privacy and security of the data subjects 'expires.' This thought is expressed by P5:

Choosing to use a sensitive video depends on whether it has been shared before. We do look at other news organizations. If, for example, the BBC has posted the video before, what we do with it does not matter anymore. Of course, we think about the risk of endangering people, but if their content has already been featured, they could already be in danger.

Whereas the above statement is echoed by some participants, others feel the duty of protecting the privacy and security of data subjects or providers should be autonomous from the actions of other news outlets. P1, for example, thinks that "journalists should no matter what avoid playing an active role in enabling governments to track people," and P3 stressed that "even if the content is already circulating widely, we still attempt to protect the privacy of the content subject or provider."

Participants indicate they treat each privacy-related case as a unique one "that deserves and receives custom treatment and personal attention" (P3) and "is not benefited by a standardized solution" (P2). Together with clarifying that there are no official open-source privacy-protection guidelines, participants reveal that they enjoy relative freedom when deciding whether and how to protect data subjects and providers. This freedom leads some participants to draft guidelines for themselves to adhere to. Abiding by self-established rules offers guidance to cautiously go about people's privacy. When P4 refers to his personally established and self-imposed rule of never posting videos taken from within homes, he mentions the 'spotlight effect' - a concept presented in the literature review:

When something like a bombardment is filmed from within a home or apartment, it can be geolocated. We know how to do this, and so do other people. In the Ukraine-Russia conflict, it has happened that Russian intelligence officers geolocated a publicly available video, after which they bombed the place it was filmed in. That's the spotlight effect. So, I never ever embed videos with risky content filmed from within homes.

Just like P4, P3 mentioned never embedding content shot from within homes in his news items. The avoidance of using specific content rules out the risk of endangering data subjects and brokers, but "if you want to report on areas where there is no free press, you need methods to safely use personal content because sometimes it is all there is" (P5).

Participants mention using a wide array of measures to protect the privacy of data subjects and brokers when embedding their content. Removing content metadata, which contains geolocational details, is a recurring measure that decreases the possibility of localizing data subjects or brokers. Removing watermarks, for example, on TikTok videos, is another way to complicate finding the content creator. Often, participants mention blurring faces and usernames to conceal identities.

Apart from image altering and deleting personal details, data subjects' and brokers' security is considered when open-source journalists share collected data (sets) amongst colleagues and community members. P2 said: "I call it 'good data hygiene' - the idea that all the data that you deal with is secured. We established a kind of danger-handling model for this." This model contains questions that P2 and colleagues pose to themselves when sharing data: "What's the potential danger? What could happen in the worst case? How will we prevent this?" P2 treats the answers to the questions of the danger-handling model as guidelines to keep data from being intercepted by malicious parties.

Some unexpected results regarding privacy protection surfaced. First, P1 mentioned that civilians themselves are increasingly aware of the dangers of posting content, which in Ukraine specifically led to a decrease in sensitive content being posted: "Gradually, Ukrainians realized that there is a large online community constantly analyzing civilian war content – over time, they tell each other to stop posting content since people get arrested and tortured when they are being localized." Second, participants mentioned feeling responsible for the safety and privacy of "bad actors," like compilers of child pornography networks or proclaimers of hate speech. P2 said: "you don't want their identity to surface either, for they might harm or even kill themselves when that happens –

something that I don't want to contribute to." P7, who often uses Tweets to expose public discourses, illustrated his stance on this topic with an example:

To me, the size of someone's reach and their public position are determinants of whether I anonymize them in my news items. When a relatively unknown civilian Twitter user tweets hate speech, I generally do not expose their account. That's because I see journalistic relevance in reporting on assertions, not in reporting on a random person. But if a politician tweets hate speech, I will not keep them anonymous. Because then there is also journalistic relevance in featuring the person.

P7 elaborated on the above statement by clarifying his inability to foresee the consequences of embedding content of a relatively unknown individual, which leads him to carefully deal with their privacy situations. Politicians, however, are already in the public eye. P7 states that, therefore, whether journalists anonymize them or not is not important: "Politicians' comments are already widespread."

Attitudes toward the use of gray information

Open-source journalists' stances on the use of gray information (semi-legal data, often from a questionable nature) vary. Three dominant attitudes towards using gray information were identified during the coding process: supportive of use, undecided towards use, and opposed towards use. Despite their differing opinions, all participants admitted that gray information plays a significant role in their investigative processes. This was made evident by the provision of illustrative cases regarding the use of gray information by all respondents.

Open-source journalists that were supportive of the use of gray information generally underpin their views with the explanation that, often, the key to a story's crux is found within gray information. P1's statement illustrated this: "Often, [gray information] is a necessary form of information disclosure. I think it can and should always be used as long as [we] don't have to pay a bad actor – like a blacktop hacker – for it." In line with P1's view, P7 is also supportive of using gray information, stating that "the importance of the investigative goal is often greater than the legality of the means."

P4 is also in favor of using gray information: "If it is in the public interest and relevant to our research, [we] must use gray information." When asked to provide an example, P4 talked about infiltrating in invitation-only Telegram groups with a fake identity, "sort of as a digital undercover

agent," in which he retrieved information that served as key evidence in an open-source news item. Supporters of using gray information emphasize their close relations with their organizations' legal department, for their positive attitude towards using it does not imply they are willing to risk legal implications.

Some participants have an undecided stance toward using gray information. Generally, their opinions on using it or not differ for every case. P5 stressed that his undecidedness towards using gray information stems from the frequent inability to verify the data:

Often, semi-legal data is even harder to verify than legal, open data. If the data is already leaked, I believe we can use it, it's just that it's hard to know whether it's real. For example, we once refrained from using leaked audio recordings of Russian soldiers communicating with each other about losing a battle. It would really complement our story, but we didn't know if it was real or fake and potentially posted by Ukranian soldiers.

Apart from the difficulty of verification, participants indicate their skepticism towards using gray information is due to the risk of breaking the law. The balance between using gray information because it is relevant and necessary to tell a story and not using it because of legal constraints is what makes participants undecided. P3 illustrated the struggle to keep this balance:

If we need leaked or hacked data because the story cannot be told otherwise, we must choose between telling a story with illegal data or not telling it at all... We then must decide the importance of the story. How important is it that people know about this? Is this importance high enough to use illegal data? It's a continuous discussion and struggle.

Although participants with doubts about the use of gray information indicate they struggle with it, it seems that often, their final decisions do lean towards using it. This decision is generally made after dialogue with colleagues, editors-in-chief, and the legal department.

A minority of participants are principally against using gray information. The main argument to not use semi-legal or illegal data is put into words by P6: "I never use it, because to me it is not an option, and in the Netherlands, it is also unnecessary. I always find legal ways to get the information I need." The proclaimed unnecessity of using illegal or dubious sources is echoed by the other participants that are anti-using gray information. P8 is one of them and argues that "generally, the more tech-savvy you are, the better you are at getting all the information you need in a way that is actually legal."

Active inter-organizational collaboration and lack of guidelines

Participants unanimously mentioned the active and daily involvement of their editorial colleagues and their organization's legal department when dealing with privacy-related issues.

First, the collaboration between colleagues, often in the form of brainstorming sessions or quick meetings, serves to ensure that a fitting privacy measure is taken. Also, the apparent interchangeable and continuous conversations between colleagues and editors-in-chief contribute to assuring the legality of the open-source investigations that are conducted and the news items that they fuel. Often, participants prefer to keep the legal department tuned into their investigative processes to ensure that nothing can be held against them if trouble arises once their news items are published. About this, P2 stated:

If the identity of an individual surfaces inadvertently and it was not our fault but that of another news organization or person, we can still get accused of it. That can result in a legal hassle. It is very convenient to have discussed the entire process with the legal department, so they are aware of all my steps and know that I have been cautious. Conversations with the legal department are hugely important, even if they are time-consuming.

OSINT journalists prefer that the legal department monitors their investigative processes to ensure the legality of their methods and, as indicated above, to rely on them when legal accusations might arise. As P6 commented: "I always clarify my information sources, because otherwise, it is impossible for them [colleagues and legal departments] to trace back my steps – suppose someone drags you in front of the journalism council, you just want to be able to show where something came from." P8 thinks that close relations with the legal department are important to not only legally protect yourself, but also your colleagues. He states:

If you did something illegal, didn't communicate about it properly, and it surfaces, you put not only yourself but also your colleagues and maybe even the whole company in a bad light. You have to take responsibility for yourself, but also for the team around you.

Although open-source journalists are in close contact with their colleagues and the legal department, guidelines about privacy issues are non-existent within their organizations. P1 stated that "within [my organization], we have never found the time to draw up guidelines, since the open-source department was only established about a year ago." P7 also blames time constraints for the absence of guidelines: "Open source is developing so quickly that news organizations haven't found the time to reflect and create rules." Some participants indicated they would prefer

guidelines to adhere to. P4 said: “We don’t have any guidelines, and I think that’s crooked: I wish we did have rules. We do talk about privacy considerations with colleagues continuously, but we don’t have standardized rules, and I think we should have them.”

Apart from the absence of guidelines, open-source journalists indicated they are not obliged to receive training about how to deal with privacy-related assets of their work. Most participants have taught themselves how to conduct open-source investigations and turned to the occupation out of intrinsic motivation. Participants mention teaching and helping each other when dealing with privacy-related issues. Notably, reciprocal instructions are not limited to the protection of data subjects and brokers but also include tips and recommendations about the digital privacy situations of open-source journalists themselves. P3 explained:

We strongly advise [colleagues] to encrypt their hard drives and collected datasets. Honestly, I sometimes dread explaining how to do this, because manual encryption techniques are not very user-friendly. However, encryption and data protection are extremely important and necessary, also when sharing files with colleagues. It can feel exaggerated at times, but it also protects you from legal implications if something goes wrong: Then, you can prove that you tried everything to prevent privacy mistakes.

Although inter-organizational tutoring is common, two participants are dissatisfied with some colleagues’ lack of awareness about how to protect the security of data brokers and subjects. P6 is one of them and admitted: “Honestly, sometimes I feel the editorial responsibility about the security of people just isn’t what it should be: often, insufficient thought is spent on it.”

Discussion

Existing research on open-source journalism and privacy issues emphasizes the power that journalists hold over the safety and privacy situations of the individuals that provide and star in the content used for investigations (Dubberley & Ivens, 2022; Saugmann, 2019; Pastor-Galindo et al., 2020). The results of this article show that open-source journalists are aware of the influence of their editorial choices on the safety and privacy of data subjects and brokers. Yet, despite open-source journalists’ unanimous awareness of their power positions, the degree to which they feel responsible for safeguarding data subjects and brokers differs. A significant number of the open-source journalists that were interviewed take to the actions of renowned, international open-source news organizations (like the BBC) to base their privacy-protection choices. These journalists feel

that their good intentions and careful measures are futile if the concerning content has already been embedded by organizations that are bigger than theirs. Others, however, consider it their autonomous and non-debatable duty to protect individuals' safety and privacy.

Surprisingly, Fiorella's (2021) concept of 'the spotlight effect' has been cited by an open-source journalist - in connection to a case where the Russian military bombed a Ukrainian home from where the content was filmed and uploaded. Apart from this direct citation, the concept of the spotlight effect indirectly surfaced when an open-source journalist explained how he typically refrains from exposing the Twitter accounts of unknown users due to his inability to foresee potentially harmful consequences. The appearance of 'the spotlight effect' in the results of this research indicates that open-source journalists are aware of the power they hold over the privacy and security of data brokers and subjects. That is, they realize that their reach, and that of their organizations, can cause unwanted and dangerous publicity to individuals that are involved in uploading certain content.

The privacy-protection measures that open-source journalists employ can roughly be divided up into two categories: Firstly, the altering of audiovisual content and the removal of its metadata. Secondly, the encryption of collected and shared data. The altering of images and videos, for example, by blurring faces or removing watermarks, serves to directly hinder the identification of the individual the measure is aimed at. Encrypting collected and trafficked data is done to keep it from falling into the hands of malicious parties.

The concept of gray information, as formulated by Hribar, Podbegrar, and Ivanuša (2014), was recognized by all open-source journalists that participated in this research. The fast rate and broad range of examples that open-source journalists provided in the interviews show that gray information plays a significant role in their investigative processes. Open-source journalists' stances towards using gray information can be categorized into three groups: supportive of, undecided about, and opposed to.

Supporters of using gray information stress that, often, semi-legal or hidden information contains the crux of an investigative story and is, therefore, utterly necessary to use. Supporters believe that, generally, the importance of a gray information-based news item is greater than the legality of the means to establish it. Open-source journalists with an undecided stance towards using gray information are doubtful due to the frequent inability to verify such data and the fear of legal implications. Regardless, they indicate that often, they ultimately choose to use gray information

after the reassurance of their colleagues, editors-in-chief, and legal departments. A minority of open-source journalists are opposed to using gray information because they believe it is unnecessary: according to them, there are always legal and open ways to generate the information they search for.

Results show that open-source journalists' stances towards gray information are in a way connected to their "gut feeling," which Schultz (2007) conceptualizes as a (journalistic) news-making process based on self-evident and self-explanatory judgments. Both supporters and opponents towards the use of gray information base their attitudes on their personal, self-evident, and self-explanatory assessment of whether using such information is necessary. However, open-source journalists that are mostly undecided about using gray information base their ultimate decisions on whether the data is verifiable, as well as on the opinions of their colleagues, editors-in-chief, and legal departments.

Furthermore, the results of this research are in line with Belghith, Venkatagiri and Luther's (2022) finding that open-source newsrooms and organizations emphasize communication, transparency, and strong bonds between members. Open-source journalists are in daily, direct, and close contact with colleagues, editors-in-chief, and their organization's legal department. Meetings serve to brainstorm about privacy measures, discuss legal issues, and exchange knowledge.

Open-source journalists' narratives reveal that there are no privacy-protection guidelines that their organization obliges them to adhere to. This grants them freedom in their investigative processes and privacy-related choices, but also leads to dissatisfaction: Open-source journalists indicate they would prefer rules and guidelines to offer them guidance when making privacy-related decisions. Therefore, the lack of guidelines leads some journalists to establish rules and models themselves.

The most prevalent and unprecedented theme that came up when exploring the role of open-source newsrooms and news organizations in dealing with privacy issues is the close contact between open-source journalists and their organization's legal departments. Open-source journalists are particularly concerned with the legality of their investigative processes and attempt to avoid potential legal repercussions. They prefer to keep the legal department tuned in to their investigative processes to rule out the possibility of illegalities being held against them.

Finally, according to Saugmann (2019), open-source journalists fail to respect the protected status of civilians in their online data collection practices. Based on the findings of this paper, Saugmann's (2019) verdict seems unjustified: Although some open-source journalists' sense of

responsibility towards civilians' privacy is arguably deficient, the protection measures that open-source journalists take are well thought-out and effective. The referrals to the implications of 'the spotlight effect' are indicative of this, as well as the close inter-organizational contact, of which a fundamental goal is to collectively ensure that the most effective protection measures are taken.
In summary, a formal normative infrastructure that guides open-source journalists with privacy issues is non-existent. Therefore, open-source journalists are expected to decide for themselves how to navigate privacy issues and where the boundaries of transparency and fairness lay.

Conclusion

As research on open-source journalists dealing with work-related privacy issues is scarce, this article aims to fill that void. It examines how journalists using open-source intelligence technologies understand and navigate privacy issues and balance them with their investigative goals. Open-source journalists understand and navigate privacy issues through dialogues and brainstorming sessions with colleagues, and through self-evident and self-explanatory assessment (RQ1). They balance privacy issues with their investigative goals by deciding whether journalistic interest is greater or less than the privacy and security needs of data subjects and brokers (RQ1). Furthermore, the results show that open-source journalists are aware of the power they hold over the security and privacy situations of the subjects and brokers of the data they use. However, the degree to which open-source journalists feel responsible for protecting people's privacy and security differs. The privacy protection measures they take include altering images and videos and removing metadata, and encryption of collected data and data shared with colleagues (RQ2). Additionally, open-source journalists' stances towards the use of gray information vary from supportive of it, undecided towards it, and opposed to it (RQ3). These attitudes are dependent on open-source journalists' personal assessments of the necessity of gray information and the opinions of their colleagues, editor
s-in-chief, and members of their organization's legal department. Lastly, the newsroom or organization that open-source journalists work at plays a significant role in their privacy-related decisions and actions (RQ4). Contact between open-source journalists, editors-in-chief, and members of the legal department serves to brainstorm about privacy measures, discuss legal issues, and exchange knowledge. Results show that open-source newsrooms and organizations do not enforce privacy-related rules or frameworks on their journalists. Additionally, dissatisfaction with

the lack of privacy-related guidelines exists among open-source journalists. The absence of guidelines and rules leads some open-source journalists to establish personal frameworks and models, to offer them guidance. An unanticipated finding of this research is that open-source journalists appear particularly concerned with the legal coverage of their investigative processes and published news items.

With one exception, the open-source journalists that were sampled for this article originate from and reside in the Netherlands. This fact, together with the relatively small number of interviews conducted for this research, stands in the way of any generalization. Additionally, bias caused by the snowball sampling method must be recognized – when individuals suggest other individuals from their network, there is no such thing as randomization. The sample is skewed in terms of gender: one woman was interviewed, versus seven men. The fact that some interview questions asked open-source journalists to speak about their organizational structure translates into another limitation of this research. That may be because journalists would not want to shine a negative light on their organizations and the people that they work with.

Future research can point out whether the limitations of this research distort the findings in any way. More research is also needed to map out the ethical and legal problems that open-source journalists encounter and the ways they deal with them.

References

Agnellutti, C. (Ed.). (2014). Big Data: An Exploration of Opportunities, Values, and Privacy Issues. Nova Publishers.

Agre, P. E., & Rotenberg, M. (Eds.). (1997). Technology and Privacy: The New Landscape. The MIT Press. https://doi.org/10.7551/mitpress/6682.001.0001

Belghith, Y., Venkatagiri, S., & Luther, K. (2022). Compete, Collaborate, Investigate: Exploring the Social Structures of Open Source Intelligence Investigations. CHI Conference on Human Factors in Computing Systems, 1–18. https://doi.org/10.1145/3491102.3517526

Blood, R. (2002). The weblog handbook: Practical advice on creating and maintaining your blog. Basic Books.

Boyd, D. (2010). "Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications." In Networked Self: Identity, Community, and Culture on Social Network Sites (ed. Zizi Papacharissi), pp. 39-58.

Bryman, A. (2016). Social research methods (Fifth Edition). Oxford University Press.

Carson, A. (2021). The Digital Spotlight: Applying a Connective Action Framework of Political Protest to Global Watchdog Reporting. The International Journal of Press/Politics, 26(2), 362–384. https://doi.org/10.1177/1940161220912679

Carson, A., & Farhall, K. (2018). Understanding Collaborative Investigative Journalism in a "Post-Truth" Age. Journalism Studies, 19(13), 1899–1911. https://doi.org/10.1080/1461670X.2018.1494515

Coleman, E. G. (2013). Coding Freedom: The Ethics and Aesthetics of Hacking. Princeton University Press. https://doi.org/10.1515/9781400845293

Coleman, E. G., & Golub, A. (2008). Hacker practice: Moral genres and the cultural articulation of liberalism. Anthropological Theory, 8(3), 255–277. https://doi.org/10.1177/1463499608093814

De Vuyst, S. (2020). Hacking gender and technology in journalism. Routledge, Taylor & Francis Group.

Dubberley, S., & Ivens, G. (n.d.). Outlining a Human-Rights Based Approach to Digital Open Source Investigations.

Edwards (2023). Open source journalism in a wired world.pdf. (n.d.).

Eijkman, Q., & Weggemans, D. (2013). Open source intelligence and privacy dilemmas: Is it time to reassess state accountability? Security and Human Rights, 23(4), 285–296. https://doi.org/10.1163/18750230-99900033

Gauthier, C. C. (2002). Privacy Invasion by the News Media: Three Ethical Models. Journal of Mass Media Ethics, 17(1), 20–34. https://doi.org/10.1207/S15327728JMME1701_03

Glassman, M., & Kang, M. J. (2012). Intelligence in the internet age: The emergence and evolution of Open Source Intelligence (OSINT). Computers in Human Behavior, 28(2), 673–682. https://doi.org/10.1016/j.chb.2011.11.014

Gubrium, J. F., & Holstein, J. A. (Eds.). (2002). Handbook of interview research: context & method. Sage Publications.

Hammond, P. (2017). From computer-assisted to data-driven: Journalism and Big Data. Journalism, 18(4), 408–424. https://doi.org/10.1177/1464884915620205

Hansen, E. (2015). The Homo Sacer of open-source journalism. Empedocles: European Journal for the Philosophy of Communication, 6(1), 21–38. https://doi.org/10.1386/ejpc.6.1.21_1

Hribar, G., Podbregar, I., & Ivanuša, T. (2014). OSINT: A "Gray Zone"? International Journal of Intelligence and CounterIntelligence, 27(3), 529–549. https://doi.org/10.1080/08850607.2014.900295

Karlsson, M. (2010). Rituals of Transparency: Evaluating online news outlets' uses of transparency rituals in the United States, United Kingdom and Sweden. Journalism Studies, 11(4), 535–545. https://doi.org/10.1080/14616701003638400

Kelty, C. M. (2020). Two Bits: The Cultural Significance of Free Software (M. M. J. Fischer & J. Dumit, Eds.). Duke University Press. https://doi.org/10.1515/9780822389002

Koops, B.-J., Hoepman, J.-H., & Leenes, R. (2013). Open-source intelligence and privacy by design. Computer Law & Security Review, 29(6), 676–688. https://doi.org/10.1016/j.clsr.2013.09.005

Lewis, S. C. (2012). THE TENSION BETWEEN PROFESSIONAL CONTROL AND OPEN PARTICIPATION: Journalism and its boundaries. Information, Communication & Society, 15(6), 836–866. https://doi.org/10.1080/1369118X.2012.674150

Lewis, S. C., & Usher, N. (2013). Open source and journalism: toward new frameworks for imagining news innovation. Media, Culture & Society, 35(5), 602–619. https://doi.org/10.1177/0163443713485494

Lewis, S. C., & Westlund, O. (2015). Big Data and Journalism: Epistemology, expertise, economics, and ethics. Digital Journalism, 3(3), 447–466. https://doi.org/10.1080/21670811.2014.976418

Manninen, V. J. E. (2017). Sourcing practices in online journalism: an ethnographic study of the formation of trust in and the use of journalistic sources. *Journal of Media Practice*, *18*(2–3), 212–228. https://doi.org/10.1080/14682753.2017.1375252

Mast, J., & Hanegreefs, S. (2015). When News Media Turn To Citizen-Generated Images of War: Transparency and graphicness in the visual coverage of the Syrian conflict. Digital Journalism, 3(4), 594–614. https://doi.org/10.1080/21670811.2015.1034527

Meijer, R., Conradie, P., & Choenni, S. (2014). Reconciling Contradictions of Open Data Regarding Transparency, Privacy, Security and Trust. Journal of Theoretical and Applied Electronic Commerce Research, 9(3), 32–44. https://doi.org/10.4067/S0718-18762014000300004

Müller, N. C., & Wiik, J. (2023). From Gatekeeper to Gate-opener: Open-Source Spaces in Investigative Journalism. Journalism Practice, 17(2), 189–208. https://doi.org/10.1080/17512786.2021.1919543

Parasie, S., & Dagiral, E. (2013). Data-driven journalism and the public good: "Computer-assisted-reporters" and "programmer-journalists" in Chicago. New Media & Society, 15(6), 853–871. https://doi.org/10.1177/1461444812463345

Pastor-Galindo, J., Nespoli, P., Gomez Marmol, F., & Martinez Perez, G. (2020). The Not Yet Exploited Goldmine of OSINT: Opportunities, Open Challenges and Future Trends. IEEE Access, 8, 10282–10304. https://doi.org/10.1109/ACCESS.2020.2965257

Perdomo, G., & Rodrigues-Rouleau, P. (2022). Transparency as metajournalistic performance: The New York Times' Caliphate podcast and new ways to claim journalistic authority. Journalism, 23(11), 2311–2327. https://doi.org/10.1177/1464884921997312

Phillips, A. (2010). Transparency and the new ethics of journalism. Journalism Practice, 4(3), 373–382. https://doi.org/10.1080/17512781003642972

Rambukkana, N. (2019). The Politics of Gray Data: Digital Methods, Intimate Proximity, and Research Ethics for Work on the "Alt-Right." Qualitative Inquiry, 25(3), 312–323. https://doi.org/10.1177/1077800418806601

Robinson, S. (2011). "Journalism as Process": The Organizational Implications of Participatory Online News. Journalism & Communication Monographs, 13(3), 137–210. https://doi.org/10.1177/152263791101300302

Saugmann, R. (2019). The civilian's visual security paradox: how open source intelligence practices create insecurity for civilians in warzones. Intelligence and National Security, 34(3), 344–361. https://doi.org/10.1080/02684527.2018.1553700

Schultz, I. (2007). The journalistic gut feeling: Journalistic doxa, news habitus and orthodox news values. Journalism Practice, 1(2), 190–207. https://doi.org/10.1080/17512780701275507

Singer, J. B. (2007). Contested autonomy: Professional and popular claims on journalistic norms. Journalism Studies, 8(1), 79–95. https://doi.org/10.1080/14616700601056866

Suomela, T., Chee, F., Berendt, B., & Rockwell, G. (2019). Applying an Ethics of Care to Internet Research: Gamergate and Digital Humanities. Digital Studies / Le Champ Numérique, 9(1). https://doi.org/10.16995/dscn.302

Westcott, C. (2019). Academic research, journalism or spying?

White, Gwen. (2016). Big data and ethics: examining the grey areas of big data analytics. Issues in Information Systems. 17. 1-7.

Whitehouse, G. (2010). Newsgathering and Privacy: Expanding Ethics Codes to Reflect Change in the Digital Media Age. Journal of Mass Media Ethics, 25(4), 310–327. https://doi.org/10.1080/08900523.2010.512827

Young, R. (1999). Giving it Away: How Red Hat Software Stumbled Across a New Economic Model and Helped Improve an Industry. The Journal of Electronic Publishing, 4(3). https://doi.org/10.3998/3336451.0004.304

Open-source Intelligence (OSINT) by Giancarlo Fiorella, Investigator and Trainer at Bellingcat. (n.d.). Retrieved May 30, 2023, from https://www.youtube.com/watch?v=AYKRE9WGSV4

Appendix A – Topic List

Protection of privacy and security of data subjects and brokers	• Do you feel open-source journalists have an influence on the privacy and security of the people whose content they feature? • Do you feel responsible for the privacy and security of the people whose content you feature? • Do you adhere to standardized rules when it comes to the protection of the privacy and security of people whose content you feature? • What digital measures do you take to protect the privacy and security of people whose content you feature?
Gray information	• Have you ever used gray information in an investigative process or news item? • Can you provide an example of gray information? • What is your stance on the use of gray information?
The role of the organization	• Are you in contact with colleagues about privacy-related issues? • Do you collaborate on finding fitting privacy protection measures? • Have you received training within your organization to deal with privacy-related issues in your work?

Appendix B - Codebook

Code	Coded when	Examples
Responsibility of privacy and security situations	Participants talk about the sense of responsibility they feel towards protecting the privacy of subjects and brokers of data.	P6: ‘’I really check whether people are potentially harmed. Sometimes, I keep on checking over and over whether I am protecting someone’s identity well enough.’’ P3: ‘’Even if content is already circulating widely, we still attempt to protect the privacy of the content subject or provider.’’ P5: ‘’Choosing to use a sensitive video depends on whether it has been shared before. We do look at other news organizations. If, for example, the BBC has posted the video before, what we do with it does not matter anymore. ’
Involvement of colleagues and executives	Participants mention the involvement of their colleagues and editors-in-chief when dealing with privacy-related issues. Unexpectedly, participants mentioned connecting with the legal departments of their broadcasters.	P2: ‘’I have continuous conversations about whether and how we anonymize individuals with the rest of the newsroom and with the editor-in-chief. And with the legal department, so everyone can be tested legally.’’ P4: ‘’What we do stays within boundaries, I feel. We always have conversations with the legal department to ensure we are not making ethical or legal mistakes.’’ P5: ‘’There is coordinators and editors that we can always

		talk and ask questions about what we do with privacy problems.''
Rules and guidelines	Participants mention (the absence) of privacy-related rules and guidelines enforced by their organizations + Participants mention rules, frameworks or guidelines they established for themselves.	P1: ''If people shoot images or videos of Russian military vehicles in occupied Ukraine from their apartments or houses, it's a rule for me to not post the geolocation or to make parts of the footage unrecognizable.'' P4: ''We don't have any guidelines, and I think that's crooked: I wish we did have rules. We do talk about privacy considerations with colleagues continuously, but we don't have standardized rules, and I think we should have them.''
Digital privacy protection measures	Participants mention the digital measures they employ in an attempt to protect the privacy of data subjects and brokers + Participants mention how and why they protect collected data (sets) and data traffic between colleagues.	P3: ''Encryption and data protection are extremely important and necessary, also when sharing files with colleagues.'' P4: ''We posted pictures, but blurred his face. That was an editorial choice, because he was under-age and because he was committing crimes in some of his pictures. We did not want to bring more trouble to him.'' P5: ''We often blur faces of prisoners of war, like when they confess something on camera. I think that's according to war legislation what you have to do.''

Privacy and security of offenders	Participants mention feeling responsible towards the privacy and security situations of offenders (for example, proclaimers of hate speech)	P7: ‘’When a relatively unknown civilian Twitter user tweets hate speech, I generally do not expose their account. That's because I see journalistic relevance in reporting on assertions, not in reporting on a random person.’’ P2: ‘’You don’t want their [offenders’] identity to surface either, for they might harm or even kill themselves when that happens – something that I don’t want to contribute to.’’
Gray information	Participants mention their stances towards using gray information: supportive of; undecided towards, opposed to.	P4: ‘’I join certain Telegram groups for my investigations, which are invitation-only. So I joined with a fake identity. I have to be careful, but if we think it is in the interest of journalism we believe we are allowed to get information this way.’’ P6: ‘’I never use it, because to me it is not an option, and in The Netherlands, it’s also unnecessary. I always find legal ways to get the information I need.’’ P3: ‘’If we need leaked or hacked data because otherwise, the story cannot be told, we must choose between telling a story with illegal data or not telling it at all... We then must decide the importance of the story. How important is it that people know about this? Is this importance high enough to use illegal data? It’s a

		continuous discussion and struggle.''
Interorganizational collaboration	Participants mention their connections and involvement with colleagues, editors-in-chief and their organization's legal department.	P2: ''Having conversations, although sometimes they are time consuming, is hugely important. Because you have to be so careful about all of this. Sometimes it takes very long, but some conversations with colleagues take less than 10 minutes and then we have found a great solution already.'' P6: ''For every case it's a new conversation. I talk to the editor in chief and my colleagues all the time to make sure we are not throwing someone under the bus accidentally.''
Guidelines on privacy issues	Participants mention (the absence of) guidelines and rules on privacy-related issues.	P7: ''Open source is developing so quickly that news organizations haven't found the time to reflect and create rules.'' P1: ''Within [my organization], we have never found the time to draw up guidelines, since the open-source department was only established about a year ago.''
Education on open-source methods	Participants mention receiving or giving instructions about conducting open-source practices.	P1: ''I have never officially been educated in open-source methods, I have taught myself and been a computer geek for about 20 years now. Not everyone receives training, many just learn it because they love it.''

		P7: ‘’I have completed some open-source courses and workshops. Everybody always says it’s easy, everybody can do it. But it’s actually quite hard.’’