doi:10.1017/aju.2024.2

SYMPOSIUM ON DIGITAL EVIDENCE

ETHICAL CONSIDERATIONS FOR OPEN-SOURCE INVESTIGATIONS INTO INTERNATIONAL CRIMES

*Alexa Koenig**

Over the past decade, the field of digital open-source investigations has both expanded and matured. Open source investigations rely on information that anyone can access from the Internet, and have been increasingly deployed by legal investigators, human rights researchers, and journalists. The investigatory methods include Boolean searches; sourcing videos, photographs, and other data from social media; determining locations of events by comparing photos and videos to satellite images; mining the deep web for government records; determining time of day by analyzing shadows in photos; and more. In this essay, I argue that digital open-source investigators can (and should) rely on a three-step process when faced with an investigations-related dilemma: (1) identifying what the law says they can or cannot do according to their professional identity, intended purpose, and relevant jurisdiction; (2) soliciting guidance from their professional code of ethics (if they have one); and (3) weighing their options against the values of safety, accuracy, and dignity. Given space constraints and prior coverage of the first two steps, I focus here on step three. The values of safety, accuracy, and dignity arguably apply to all open-source investigations and investigators, regardless of jurisdiction or professional identity, and their application reflects a relatively novel attempt to help set ethics-based boundaries around investigation-related activities.

The Need for an Applied Ethical Framework Specific to Open-Source Investigations

While the early 2010s were marked by tremendous advances in the ability to find, preserve, analyze, and present information from online "open" spaces, the last several years have highlighted the many ways in which such investigations can also come with a cost—for survivors of atrocities, investigators, the public, and others.[1] These costs may include inaccuracies due to unchecked bias, inadequate verification, or insufficient research expertise; physical, digital, or psychological harm; or insults to human dignity, for example when sensitive information becomes public or is amplified because of an investigation.

Ethical guidelines are designed to offset such costs, advancing moral decision making in ways that minimize physical, psychological, reputational, or other harm to individuals, communities, organizations, and society at large.[2] They are crucial for outlining professional boundaries by helping to clarify the values generally adhered

* *JD, MA, PhD, Co-Faculty Director, Human Rights Center and Adjunct Professor, University of California, Berkeley School of Law, United States. The author thanks Maggie Andresen, MJ Durkee, Neha Jain, and Tamar Megiddo for their feedback on earlier drafts.*

[1] *See, e.g.*, Giancarlo Fiorella, *Notes from the Digital Field: Ethical Dilemmas in Open Source Research*, Bellingcat (Sept. 18, 2023); Ed Millett, *Deploying OSINT in Armed Conflict Settings: Law, Ethics, and the Need for a New Theory of Harm*, Humanitarian L. & Pol'y (Dec. 5, 2023).

[2] *See* Peter Singer, *Ethics*, Britannica (last updated 2023).

to within a profession as well as providing guidance on how to best address practical normative challenges and resolve ethical dilemmas.[3]

Ethical dilemmas arise when, no matter which action out of many is chosen, at least one ethical principle will be compromised.[4] In the open-source investigations context, such conflicts are frequent. One common example is when fidelity to methodological transparency conflicts with privacy or security interests, such as when an organization posts the geo-coordinates of where someone was likely standing when they shot a video that was crucial to an investigation. While this transparency can build confidence in findings, the coordinates may also pinpoint someone's location, allowing them to become a target for retaliation.

The failure of legal and ethical codes to keep pace with the rapid development of technology creates a gap between what investigators *can* do and what they *should* do. When the law fails to provide adequate guidance, professional ethical codes ideally step in to indicate a path forward. But when professional codes are silent or irrelevant, what then?

A significant obstacle stands in the way of developing a comprehensive ethics framework for digital open-source investigations: the multidisciplinary origins of the underlying methods. Relevant ethical considerations may come from the fields of intelligence, journalism, law, and the social sciences. While related ethical principles may overlap, they do not converge. For example, social scientists in the United States are bound to do no harm, as well as to maximize benefits and minimize potential harms of their research,[5] while journalists are charged with finding and reporting the "truth," ensuring accuracy, and safeguarding objectivity. Lawyers' duties vary, but most jurisdictions' ethics codes prioritize the interests of clients and the legal profession. These differences can generate confusion and even contention among open-source investigators, who may come from different professional backgrounds. For example, a quintessential tension is whether to disguise one's identity when conducting online research. Journalists are often prohibited from misrepresenting who they are, so may not disguise their identities. By contrast, human rights researchers may adopt inauthentic persona (or "sock puppets") in order to protect themselves, their investigation, and others.

Despite these differences, cross-professional commonalities have been clarified with the recent publication of relevant guidelines. This includes publication of the Berkeley Protocol on Digital Open-Source Investigations,[6] which aims to foster methodological consistency, facilitate communication across disciplines and institutions, professionalize practice, strengthen safety and security, and ensure high-quality work.

Consistent with the Berkeley Protocol, I argue that investigators should follow a three step-approach when facing an ethical dilemma. The first is to ask, "what does the law say I can or cannot do?" The investigator must know the legal context in which they are working,[7] as the rules may differ significantly based on the jurisdiction and/or the identity of the investigator. For example, law enforcement officers face different legal limitations than lay investigators—the former may need a warrant to monitor a social media page, while the latter does not. But even for the latter, there may be downstream effects if a lay person's investigation produces information that has potential evidentiary value for courts. Ultimately, if an investigator ignores the law, they should do so knowingly, with an awareness of how their actions may affect their reputation, their investigation, and those impacted or otherwise implicated by their research.

[3] *Applied Ethics*, Oxford Bibliographies.

[4] *See, e.g.*, Karen Allen, *What Is an Ethical Dilemma?*, New Social Worker Online (2018).

[5] *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*, 81 J. Am. Coll. Dent. 4 (2014).

[6] Berkeley Protocol on Digital Open-Source Investigations: A Practical Guide on the Effective Use of Digital Open-Source Information in Investigating Violations of International Criminal, Human Rights and Humanitarian Law (United Nations & University of California, Berkeley eds., 2022).

[7] *Id.* at 19.

https://doi.org/10.1017/aju.2024.2 Published online by Cambridge University Press

Second, the investigator should ask "am I subject to a professional code of ethics, and if so, what does it say I should do?" A journalist may face different ethical constraints than a social scientist or lawyer. For example, ethical codes may dictate whether sources should be named (preferred in journalism in the interests of transparency) versus kept anonymous or pseudonymous (often preferred in social science for privacy reasons).

Third, if one's professional code of ethics is silent, then the investigator should look to emerging norms based on shared values. There are three values that most open-source investigators, regardless of profession, seem to agree are central to the responsible use of open-source information: safety, accuracy, and dignity.[8] Each is discussed briefly below.

Safety

First, when facing an ethical dilemma, investigators should prioritize the safety of all stakeholders. A non-exhaustive list of stakeholders includes the investigator, their team, perpetrators, bystanders, victims, family members, their communities, and the audience for any end product, whether journalistic, legal, or advocacy oriented.

Such safety considerations should be holistic, focusing on the physical, digital, and psychosocial well-being of all potentially impacted individuals.[9] Security risks and mitigating measures should be identified during investigation planning, prioritized from the outset, and adjusted throughout an investigation, as needed.

Often overlooked yet critical is planning for the psychosocial risks that frequently attend digital open-source investigations. Such risks can be acute for all stakeholders, including investigators. For example, the assumption that remote investigations are "safer" for investigators than those *in situ* is increasingly understood to be false; instead, the risks just shift from the physical toward the psychological. Indeed, open-source content—such as user-generated videos posted to social media—can offer an intensely raw and intimate view of violence that can be difficult to process, especially when combined with the relative isolation of online investigators and the volume of online content. Thus, digital open-source investigative teams should have a plan to foster resiliency and mitigate psychosocial harm.[10] This is critical for ethical reasons but it also protects individual researchers, including career longevity and the quality of their work.

Resiliency plans should, at a minimum, include three categories of information: (1) awareness strategies; (2) psychosocial support strategies; and (3) technical strategies—all informed by research that outlines the potential impact of graphic digital content.[11] Training should encourage awareness of the particular risks of analyzing user-generated content and the nature of secondary or vicarious trauma, including changes in behavior that could signal distress, as well as options for support following any such signals.

In addition to increasing awareness of these risks, investigators should implement strategies to safeguard against them. Common practices include turning audio off or down when viewing graphic content (since so much emotive content is embedded in sound); obscuring graphic material if not needed for analysis; minimizing screen size; avoiding working in isolation, especially late at night and/or in one's bed so that environment remains a "safe

[8] For information on the workshop that led to these insights, see Sam Dubberley & Gabriela Ivens, *Outlining a Human-Rights Based Approach to Digital Open Source Investigations: A Guide for Human Rights Organisations and Open Source Researchers* (Engine Room and Human Rights, Big Data and Technology Project, 2022).

[9] Tactical Technology, Holistic Security Manual (2016).

[10] Berkeley Protocol, *supra* note 6.

[11] *See, e.g.*, Sam Dubberley, Margaret Satterthwaite, Sarah Knuckey & Adam Brown, *Digital Human Rights Investigations: Vicarious Trauma, PTSD and Tactics for Resilience*, *in* Digital Witness: Using Open Source Information for Human Rights Research, Advocacy and Accountability (Sam Dubberley, Alexa Koenig & Daragh Murray eds., 2020); Alexa Koenig & Andrea Lampros, Graphic: Trauma and Meaning in Our Online Lives (2023).

https://doi.org/10.1017/aju.2024.2 Published online by Cambridge University Press

space" disassociated from traumatic content; taking frequent screen breaks; and safeguarding time for activities unrelated to investigatory work. Each investigative unit should have a written resiliency plan that includes these strategies, and all members of a team should be regularly trained on how to put that plan into practice.[12]

Finally, anonymization and pseudonymization may be used to protect the identities and thus the well-being of victims, witnesses, and other sources. Online investigators must, however, stay cognizant of the mosaic effect: the ways in which even small fragments of online information can be combined in ways that may be revealing of protected identities.[13] For those skilled in online research, a fact provided in a report or story, even if seemingly innocuous, may be the final piece in the puzzle needed to identify someone.

Accuracy

Accurate investigation is critical for effectuating justice and for strengthening the legitimacy of digital open-source investigations. The Berkeley Protocol details principles designed to safeguard the quality of an investigation. These principles are clustered into "professional" (the skills ideally possessed by an investigator), "methodological" (focused on how to do the work), and "ethical" (overarching commitments).

The Berkeley Protocol also emphasizes the importance of investigation planning to ensure that research is thorough and includes strategies for offsetting technical and human biases.[14] A central component of investigation planning is the creation of a digital landscape analysis—an assessment of who is communicating in online spaces, as well as how and where, so that critical perspectives are not overlooked. Also relevant is the composition of the investigation team: for example, is the team diverse with respect to gender, ethnicity, language, or other expertise? Diversity promotes the identification and correct interpretation of information.

Additional quality safeguards include testing multiple working hypotheses or proving the "null hypothesis," that is, reporting against your own investigation to find holes in your research or to try to disprove your current understanding of the facts. Journalists might focus on "reporting against their story"; lawyers on testing multiple theories of a case. For all professions, peer review can provide yet another check on accuracy.

Verification of digital information is especially critical, given the risk of photos, videos, or other online information being miscontextualized, digitally modified, or fully generated with the help of artificial intelligence. The Berkeley Protocol recommends three verification activities. The first is technical analysis. This may include examining any still-attached metadata and/or reverse image searching a photograph to see if it has previously appeared online in a different context to assess reliability and authenticity. The second is content analysis: assessing whether visual cues in an image are consistent with what the source has said that image depicts. The third is source analysis —for example, finding the original post and analyzing the reliability of the person who posted the item for the information that was shared.

Many digital open-source investigators showcase the steps in their verification process to generate trust in their analysis. However, methodological transparency can raise its own ethical dilemmas, especially when results are shared publicly, and especially when producing information about an investigation in real time, before thorough vetting. First, it is critical not to cross the line into doxing—the "compiling and releasing [of] a dossier of personal information on someone."[15] It can be tempting for some civil society investigators, for example, to reveal

[12] Sample resiliency plan on file with the author.

[13] *See, e.g.*, Jill Capotosto, *The Mosaic Effect: The Revelation Risks of Combining Humanitarian and Social Protection Data*, HUMANITARIAN L. & POL'Y (Feb. 9, 2021).

[14] For more on biases that commonly arise in digital open-source investigations, see Yvonne McDermott Rees, Alexa Koenig & Daragh Murray, *Open Source Information's Blind Spot: Human and Machine Bias in International Criminal Investigations*, 19 J. INT'L CRIM. JUST. 85 (2021).

[15] Mat Honan, *What Is Doxing?*, WIRED (Mar. 6, 2014).

https://doi.org/10.1017/aju.2024.2 Published online by Cambridge University Press

information on social media about those they have been investigating to encourage crowdsourcing of additional information on public platforms like Twitter/X. Notoriously, such practices have resulted in identifying innocent people as possible perpetrators (as happened on Reddit, when "citizen sleuths" misidentified one of the Boston Bombers as someone who had recently gone missing, causing significant distress for his family).[16] Similarly, when the U.S. Capitol riots of January 6, 2021 were investigated online, several civil society investigators shared personal information about alleged rioters on social media in ways that raised significant risks of retaliation for those persons, as well as their family members, bystanders, and others.

Dignity

The third value emerging as common to open-source investigators is dignity—a concept that is central to the international human rights framework.[17] Dignity requires that people not be viewed as means to an end, but as agents with inherent value. The benefit of adopting the human rights framework as a guide to ethical decision making is that the framework is internationally recognized, meaning that much of the logistical work of figuring out which interests should be protected and the political work of generating buy-in could be reduced relative to developing a framework from scratch.

Dignity may become relevant in an open-source investigation in several ways. Respect for dignity may include not taking credit for others' work; crediting all who participated in an investigation, paying special attention to what may be "invisible" labor; ensuring that victim-survivors have a role in setting investigative priorities and/or benefit from the investigation; being thoughtful and strategic about who participates in the investigation and how;[18] seeking consent to use data, especially sensitive data that has been posted to social media;[19] and ensuring that a range of perspectives, voices, and experiences are incorporated into an investigation, minimizing bias and maximizing representativeness.

Investigations should also be sensitive to underlying power dynamics, which might be abused deliberately or inadvertently. Avoiding "gamification"—treating online investigations like a game and losing the perspective that the investigation concerns real people with very human needs and interests—is especially important for data involving harm to individuals. For example, investigators should never unnecessarily share videos of sexual violence, a precaution that safeguards the dignity of those who experienced the violation and colleagues who may be negatively impacted by that unnecessary exposure.

Conclusion

Associate Justice Potter Stewart of the United States Supreme Court once declared that "ethics is knowing the difference between what you have the right to do and what is right to do."[20] Open-source investigators—regardless of their underlying profession—have the legal right to do quite a bit with information available in online public spaces. But what is *right* to do is a far more nuanced consideration.

[16] *See, e.g.*, Traci G. Lee, *The Real Story of Sunil Tripathi, the Boston Bomber Who Wasn't*, NBC News (June 22, 2015).

[17] Dubberley & Ivens, *supra* note 8; CELE, *Beyond Ethics: Why a Rights-Based Framework Is Essential for AI Governance*.

[18] *See* Zara Rahman & Gabrielle Ivens, *Ethics in Open Source Investigations*, *in* Digital Witness, *supra* note 11; Sylvanna Falcon, Alexa Koenig, Sofia Kooner & Jess Peake, *Symposium on Fairness, Equality and Diversity in Open Source Investigations: Democratizing OSINT–University-Based Lessons on Diversity and Inclusion*, Opinio Juris (July 2, 2023).

[19] *See* Alexa Koenig, Simone Lieban Levine, Anthony Ghaly & Hayley Durudogan, *Confronting Power and Privilege: The Importance of Consent in Digital Open-Source Investigations of Conflict Related Sexual Violence*, 22 J. Int'l Crim. Just. _ (forthcoming 2024).

[20] *See, e.g.*, Rahman & Ivens, *supra* note 18.

https://doi.org/10.1017/aju.2024.2 Published online by Cambridge University Press

This essay argues that digital open-source investigators should weigh methodological options against their legal obligations and any relevant professional code of ethics, as well as against the values of safety, accuracy, and dignity. These values are emerging as common to digital open-source investigators regardless of their profession and they may help to illuminate the best path forward with some consistency when the stakes are high, clarity is low, and the surrounding terrain uncharted.

https://doi.org/10.1017/aju.2024.2 Published online by Cambridge University Press