NAVIGATING ETHICAL DILEMMAS IN OPEN SOURCE INFORMATION ANALYSIS

S. LADERMAN, T. XU, M. FOWLER
Open Nuclear Network, a programme of One Earth Future
Vienna, Austria
Emails: sjladerman@oneearthfuture.org, txu@oneearthfuture.org, mfowler@oneearthfuture.org

B. LOEHRKE, L. KENAUSIS
Stanley Center for Peace and Security
Muscatine, Iowa
Emails: benl@stanleycenter.org, lkenausis@stanleycenter.org

Abstract

When facing ethical dilemmas in using open source information, one needs tools and methodologies that can assist in making ethically sound decisions. Towards that, this paper provides an overview of a framework to help open source information analysts and practitioners make ethical decisions, and proposes a code of ethics that can assist both non-governmental organizations (NGOs) and verification organizations (VOs) working with open source information.

1. INTRODUCTION

An increasing number of organizations, including news media employing investigative journalists, VOs, and human rights watchdogs, are relying more heavily on open source information[1] to facilitate their research. This global community of open source analysts now have access to a greater number of information sources, can break news, and add evidence to reporting, thus making an impact on public opinion and policy decisions.

Given the breadth and variety of the information currently available in open sources, ethical dilemmas are common [1]. Some open source products may also have unintended consequences as serious as compromising the privacy and liberty of individuals, endangering third parties and, in some cases, complicating diplomatic efforts and international relations [2].

2. ETHICAL DILEMMAS

Open source analysts are confronted by many of the same ethical issues common to all researchers and journalists, which have become even more salient in the age of internet-based open source investigations. For the more prevalent issues, such as plagiarism and opacity of methodology, there are clear solutions (proper attribution, citation, and transparency). There are, however, areas where competing priorities make the ethical course of action less clear to the analyst — this is the ethical dilemma.

2.1. Navigating Ethical Dilemmas

While every situation will certainly be unique, given the potential implications of the ethical dilemmas open source analysts may encounter, it is important that organizations develop guidance so that analysts do not have to navigate these delicate situations on a completely ad-hoc basis. Open source specialists at several organizations have been considering potential ethical dilemmas and how analysts might weigh various considerations when debating how to proceed with an open source investigation. Most prominently, the Stanley Center for Peace and Security organized an ethics workshop in late May 2022 for open source analysts, which addressed many of these issues, and encouraged analysts to consider potential dilemmas and how one might weigh various considerations when debating how to proceed with an open source investigation.

[1] Open source information can be defined as “information that any member of the public can observe, purchase or request, without requiring special legal status or unauthorized access”, although other definitions and interpretations exist. This definition is from The Office of the United Nations High Commissioner for Human Rights and the Human Rights Center at the University of California, Berkeley, School of Law, Berkeley Protocol on Digital Open Source Investigations (2022), www.ohchr.org/en/publications/policy-and-methodological-publications/berkeley-protocol-digital-open-source

For the workshop, participants used the following five ethical approaches, created by the Markkula Center for Applied Ethics, to weigh the competing ethical considerations in open source investigations [3, 4]:

— *Common Good:* this approach "seeks to put the benefit of the community over the individual."
— *Justice:* the origin of this approach is the notion that "we should treat each other equally, though it has evolved to recognize that 'equality' is not always 'fair.'"
— *Rights:* this approach focuses on the fact that "all humans have innate dignity and rights. Humans have the right to choose what they do with their lives freely without harm or hindrance."
— *Utilitarian:* this approach seeks to "find the best possible balance of good over harm. When applying it, consider who or what will benefit and who or what will be harmed."
— *Virtue:* this approach links ethics with virtues such as honesty, courage and compassion.[2]

Application of each of these approaches may lead to the identification of a different best course of action. For example, the ethical choice favoured by the rights approach may not be the ethical choice when considering the approach focusing on the common good. However, when the approaches do not coincide, and even when the approaches seem to contradict each other, the exercise will still be valuable, as each of the approaches "gives us important insights in the process of deciding what is ethical in a particular circumstance" [3]. Therefore, these approaches should not be taken as perfect tools that provide the one right ethical answer in every case, but rather as guides to help open source specialists understand the possible courses of action and to give them sufficient information and understanding to recommend or make the best possible ethical decisions.

2.2. Application of Ethical Approaches

2.2.1. Example One: Privileged Information

NGOs and VOs may encounter cases where information is made available either via third parties or inadvertently via other distribution mechanisms (e.g. Wikileaks, Panama Papers). In these cases, the analyst will need to think through a number of important questions regarding potential ethical implications of the use of such data, including the following:

— Could large-scale publication of privileged information lead to compromising national and international security interests?
— Could individual privacy or safety be compromised if the source of information is revealed?
— Could the rightful owners of the privileged data, potentially including Member States of VOs, object to the use of the data that may not be considered legal under their jurisdiction, keeping in mind that legal and ethical considerations are not necessarily the same in a given situation?

Given these questions, it is clear that individuals and organizations must think through the ethical implications of utilizing privileged information in their analyses and potentially publishing it to a wider audience. To illustrate the application of the ethical approaches, in a case in which the privacy of a source may be compromised by publication of some privileged information, the approaches can help to identify the ethical implications of the potential alternatives for action:

— *Common Good:* If the benefits of wide release of this information to the public may outweigh the harm to the individual information source, then large-scale publication should be considered.
— *Utilitarian:* In cases where the information source could be harmed by wide release publication, a more limited release — only to decision makers — may be considered.
— *Virtue:* Third parties would need to be kept informed in order to enable them to protect themselves and to make any necessary plans for their own privacy and safety.

As is evident from the application of the approaches to this example, the favoured actions of the different approaches can contradict each other. In such cases, the analyst or organization would need to utilise the understanding developed by considering the different approaches, and select the action based on the overall priorities of the organization. To allow for such decisions to be made in the best interest of the organization's mission, it may be helpful for the organization to have an established code of ethics that describes what considerations should be made and how such ethical decisions should be weighed.

[2] There is a sixth lens now, although it was not discussed at the workshop. This lens is the Care Ethics lens: "Care ethics is rooted in relationships and in the need to listen and respond to individuals in their specific circumstances, rather than to merely follow rules or calculate utility" [3].

2.2.2. Example Two: Security

Another ethical dilemma that could affect NGOs and VOs utilizing open source information is in the use of sources that require the registration of user information or creation of an account, often referred to as being behind a registration wall (shortened to regwall or regiwall). In such situations, the ethical dilemma is about transparency of intentions and identity of the organization, and can be in conflict with the need for privacy and safety of the analyst. A source that an analyst needs to access, yet it is behind a regwall, will give rise to questions such as the following:

- Should analysts reveal their identity or employer to the information source?
- Would revealing this information cause harm by disclosing sensitive or compromising information about the analyst to the information source?
- Would analysts lose access to important information if other website users knew their identity or that of their employer?
- Would creating a covert account to hide identity and protect compromising information be against the terms and conditions of the information source, or against the reasonable assumption of privacy of the information source?

In this situation, again, some of the same approaches could be used to make a decision on a case-by-case basis. For example, where an open source analyst would require access to information posted by an extremist group on a private forum in a social media website, the application of the aforementioned approaches could lead to the following considerations:

- *Justice:* Concealing one's identity to access information may be seen as unjust and, additionally, could break the legal terms of use of the information system hosting the information.
- *Rights:* Would concealing one's identity to access the private forum represent an unacceptable harm to the fundamental human right to privacy?
- *Utilitarian:* The harm of concealing one's identity in order to gain access to information one might not otherwise have access to would need to be weighed against that group's expectation of privacy, depending on the utility and importance of that information and the investigation.

In this illustrative example of an ethical dilemma, depending on the specifics of the case, the ethical approaches could favour either registering to access the information with full identity transparency, thereby potentially compromising the analyst's privacy and jeopardizing access to the data; or registering with a fake identity, thereby risking legal or other action by the information owner but protecting data access and their own identity.

In sum, while the exercise of considering all approaches may not give a single clear answer as to which action to take, it can help to expose the issues that must be assessed and weighted. Ultimately, this insight can be used to decide the best, most appropriate, and most objectively beneficial course of action. However, without overarching organizational guidelines, the analyst might make a decision that contradicts their organization's mandate. Therefore, any organization that may face such ethical dilemmas are strongly advised to have some guidelines or framework in place that can help analysts when facing such decisions.

3. AN OPEN SOURCE ANALYSIS CODE OF ETHICS

Given the variety of ethical dilemmas and potential outcomes outlined and described above, it is imperative that organizations think through these issues and offer guidance to their staff on how to navigate the ethical dilemmas they face. To help work through these large, philosophical problems, a code of ethics could be created and promulgated within organizations to ensure that ethical principles are applied in a fair, thorough, and consistent manner. Although many organizations have published general codes of ethics that outline generic principles that could be applied to open source analysis [5, 6], very few organizations have codified such guides explicitly for ethical decision-making in the use of open source information.

Open Nuclear Network (ONN), an NGO promoting dialogue and engagement for the purpose of nuclear risk reduction, published its own code of ethics in June 2020 in order to guide the work of its own open source analysts and provide an example for others [1]. By explicitly publishing its open source code of ethics, ONN has increased its own accountability in ensuring its analytical products are accurate, fair, objective, independent, and transparent.

To more easily allow other NGOs and VOs to adapt ONN's code of ethics to their unique contexts, needs, and responsibilities, ONN has created the following generic ethical framework for open source research and analyses. While VOs — such as the International Atomic Energy Agency (IAEA), Comprehensive Test Ban

Treaty Organization (CTBTO) or the Organisation for the Prohibition of Chemical Weapons (OPCW) — and NGOs share significant overlap in the type of ethical issues that could be faced, some important differences are implied in the example framework below.

Example Code of Ethics Framework for NGOs and VOs

General Ethical Guidelines

- All staff are obliged to abide by the laws and regulations of the jurisdiction in which they are located, the laws of other relevant jurisdictions and, to the extent relevant, the extraterritorial laws that may apply to them by virtue of their citizenship.
- Conflicts of interest, real or perceived, should be avoided, and unavoidable conflicts should be disclosed to the organization's leadership.
- Staff may not engage in professional misconduct, such as plagiarism, distortion of facts, slander, libel, defamation, or making unfounded assertions.
- The organization is responsible for ensuring that its work is accurate and fair. Factual information shall at all times be clearly distinguished from commentary, criticism, and advocacy.

Ethics of Open Source Information Collection

When collecting open source information, staff are committed to:

- always respecting individuals' reasonable expectation of privacy while anonymizing data that might reveal personal information, but keeping original files secure and intact should the validity of the information be challenged.
- assessing source material to remove undue bias based on gender, nationality, race or ethnicity, religion, political views, age, disability, or sexual orientation. Should quantitative data contain biased or non-randomized factors, analysts should publicly discuss these limitations and possible resulting distortions, and disclose any gaps in their own technical understanding of a quantitative analysis.
- never soliciting information that may jeopardize the safety or security of a person or an organization.
- soliciting or collecting only open source information.
- using primary sources and original data wherever possible. Citing those who have first-hand knowledge, such as local experts and authors.

Ethics of Open Source Analysis

When analysing open source information, staff are committed to:

- avoiding the language of causality or deterministic predictions when not supported by quantitative analysis. Instead, using more conservative and probabilistic language when framing quantitative results. (Correlation does not equal causation)
- consulting a native speaker when working with a source in an unfamiliar language, when possible, and not relying solely on the technical capabilities of translation tools.
- developing and improving the organization's methodology for assessing the reliability of sources.
- identifying subjective assumptions and possible personal biases that may affect the analysis.
- maintaining objectivity, for example, never assuming one side is always correct in a conflict situation.
- never plagiarizing; always attributing and identifying sources clearly to allow for replication of analysis.
- never spreading misinformation or disinformation, and refuting misinformation and disinformation when necessary.
- portraying data honestly, for example, avoid reporting mean values when skewness distorts the central tendency and instead report the median. Likewise, avoid crafting data visualizations that present an incomplete or distorted picture of reality.
- providing context that does not misrepresent or oversimplify developments, never stereotyping.
- treating all data sources critically, never assuming that information is error- or bias-free.
- using techniques to structure hypotheses and test them.
- when weighting data, declaring why the relative weights are assigned.

Ethics of open source publication[3]

When publishing open source information and related analyses, staff are committed to:

— acknowledging and promptly/prominently correcting mistakes; corrections and clarifications should be carefully and clearly explained publicly.
— checking sources are accurate, timely, and reflect the views of those it is intended to represent.
— disposing of sensitive data upon the completion of a project so that personal information will not be leaked or used for malicious purposes.
— ensuring speed should not take precedence over accuracy.
— explaining ethical choices and processes to audiences; dialogue with the public about best practices should be encouraged.
— forming teams to test hypotheses, where appropriate.
— keeping copies, when necessary, of all video material, images, databases, and other material used in the publication.
— properly attributing labour and contributions, no matter how small.
— vigorously peer reviewing analyses internally before publication.

The organization will commit to serving the public good and will balance the need for publishing information or analysis against potential harm in the following ways, remembering that legal access to information does not translate into an ethical justification to publish. A publication should not:

— compromise the source's safety or security.
— compromise the use of similar open source information collection methods in the future.
— contribute to discrimination on such bases as geographic, social or ethnic origin, political views, age, race, gender, sexual orientation, language, religion, or disability.
— contribute to escalation of conflict, hatred, or disinformation.
— contribute to proliferation or improvement of weapon designs.

The organization, by considering the long-term implications of publication of information or analyses, will not exercise undue influence — unintentionally, out of inexperience or with malicious intent — in ways that risk ethical harm. Consultations with peers or leadership to determine the best course of action will be conducted when necessary.

REFERENCES

[1] Open Nuclear Network, Code of Ethics (2020), available at: https://opennuclear.org/code-ethics.
[2] LOEHRKE, B., et al., Stanley Center for Peace and Security, Feeling the Burden: Ethical Challenges and Practices in Open Source Analysis and Journalism (2022), available at: https://stanleycenter.org/publications/ethics-osint-analysis-journalism/.
[3] Markkula Center for Applied Ethics, A Framework for Ethical Decision Making (2021), available at: www.scu.edu/ethics/ethics-resources/a-framework-for-ethical-decision-making/.
[4] Stanley Center for Peace and Security, Ethical Challenges and Trust When Using OSINT in Reporting Workshop (2022), unpublished.
[5] Society of Professional Journalists, SPJ Code of Ethics (2014), available at: www.spj.org/ethicscode.asp.
[6] International Federation of Journalists, IFJ Global Charter of Ethics for Journalists (2022), available at: www.ifj.org/who/rules-and-policy/global-charter-of-ethics-for-journalists.html.

[3] VOs should consider everything in this section as not just pertaining to publication, but also to utilization of open source information. This is especially important to VOs as a lot of their work is not publicly distributed, but used internally and subsequently for reporting to Member States or leadership.