THE ENGINE ROOM
Accelerating Social Change

The Human Rights, Big Data and Technology Project

Outlining a Human-Rights Based Approach to Digital Open Source Investigations:

A guide for human rights organisations and open source researchers

by

Sam Dubberley and Gabriela Ivens

March 2022

UKRI Economic and Social Research Council

Human Rights Centre | University of Essex

Table of Contents

I. Introduction

Open source information is “information that any member of the public can observe, purchase or request, without requiring special legal status or unauthorized access”.[1] It has always played a key role in researching human rights violations. Christof Koettl, Daragh Murray and Sam Dubberley note that the human rights NGO Amnesty International was founded on open source information. They also reference how the International Criminal Tribunal for the former Yugoslavia exhibited videos to reconstruct scenes and events near Srebrenica in 1995.[2]

Digitalisation has, in recent years, led to an increase in the volume of open source information – this has been through the availability of high quality, cheap photo sensors in mobile telephones, the global spread of high speed internet connectivity and the popularity of social media platforms such as Facebook, YouTube and TikTok. Social media users film and post details of their lives to the internet. These include visits to restaurants or family celebrations. Social media users also document crackdowns on protests or the misuse of tear gas.

We define this digital open source information as “publicly available information in digital format, which is generally acquired from the Internet”.[3] The availability of this information has grown as access to the sites of human rights abuses has become harder, either logistically, for instance through the denial of a travel visa, or for safety reasons. It can present compelling evidence of violations of international human rights law, international criminal law and international humanitarian law. Emma Irving points to the arrest warrant issued in 2017 against Mahmoud Mustafa Busayf Al-Werfalli in Libya by the prosecutor of the International Criminal Court as a watershed moment for digital open source information.[4] Werfalli was charged with murder as a war crime under Article 8(2)(c)(i) of the Rome Statute based on seven separate

[1] ‘Civil Liberties and Privacy Guidance for Intelligence Community Professionals’ (Office of the Director of National Intelligence 2011) DNI Pre-Pub 20140708.

[2] Christoph Koettl, Daragh Murray and Sam Dubberley, ‘The History of the Use of Open Source Investigation for Human Rights Reporting’ in Sam Dubberley, Alexa Koenig and Daragh Murray (eds), *Digital Witness: Using Open Source Information for Human Rights Investigation, Documentation, and Accountability* (Oxford University Press 2020).

[3] Christoph Koettl, ‘Citizen Media Research and Verification: An Analytical Framework for Human Rights Practitioners’ [2016] Human Rights in the Digital Age: CGHR Practitioner Papers <https://www.cghr.polis.cam.ac.uk/publications/cghr-practitioner-papers-series/paper-1> accessed 5 December 2020.

[4] Emma Irving, ‘“The Role of Social Media Is Significant”: Facebook and the Fact Finding Mission on Myanmar’ (*Opinio Juris*, 7 September 2018) <http://opiniojuris.org/2018/09/07/the-role-of-social-media-is-significant-facebook-and-the-fact-finding-mission-on-myanmar/> accessed 5 December 2020.

incidents captured in videos posted to Facebook.[5] Videos from popular protests in Hong Kong posted across social media in 2019 allowed journalists from the Washington Post, in collaboration with Amnesty International, to show how the police ignored their own internal operational guidelines in their crackdown on protesters.[6] Videos of attacks on civilians in the conflicts in Yemen and Syria have been archived and analysed by a variety of human rights groups to present evidence of attacks on medical facilities or the use of prohibited chemical weapons.[7] The United Nations Office of the High Commissioner for Human Rights now integrates open source investigations into its Commissions of Inquiry, and Fact-finding Missions.[8] Digital open source investigations have become such an inescapable feature of contemporary human rights investigation and reporting that it led to the publication of Berkeley Protocol on Digital Open Source Investigations ("The Berkeley Protocol") in 2020.[9] This document outlines "the professional standards that should be applied in the identification, collection, preservation, analysis and presentation of digital open source information and its use in international criminal and human rights investigations".[10]

Open source information is never the only source for human rights research. However, analysed alongside interviews, expert testimonials, or other methods that reconstruct a more complete, accurate picture of violations, it helps convince the public, policymakers, or judicial bodies to hold perpetrators accountable for crimes

[5] *The Prosecutor v Mahmoud Mustafa Busayf Al-Werfalli Situation: Situation in Libya* International Criminal Court ICC-01/11-01/17-2.

[6] 'Hong Kong: Leaked Police Manuals Show Officers Often Ignored Guidelines in Protest Crackdown - Washington Post' <https://www.washingtonpost.com/graphics/2019/world/hong-kong-protests-excessive-force/> accessed 5 December 2020.

[7] 'Medical Facilities Under Fire in Yemen' (*Medical Facilities Under Fire in Yemen*) <https://medical-facilities.yemeniarchive.org/> accessed 16 October 2021; 'Syria: A Year On, Chemical Weapons Attacks Persist' (*Human Rights Watch*, 4 April 2018) <https://www.hrw.org/news/2018/04/04/syria-year-chemical-weapons-attacks-persist> accessed 16 October 2021.

[8] 'OHCHR | Venezuela: UN Report Urges Accountability for Crimes against Humanity' <https://www.ohchr.org/EN/HRBodies/HRC/Pages/NewsDetail.aspx?NewsID=26247&LangID=E> accessed 5 December 2020; 'OHCHR | Yemen: Collective Failure, Collective Responsibility – UN Expert Report' <https://www.ohchr.org/EN/HRBodies/HRC/Pages/NewsDetail.aspx?NewsID=24937&LangID=E> accessed 5 December 2020.

[9] Lindsay Freeman, Alexa Koenig and Eric Stover, 'Berkeley Protocol on Digital Open Source Investigations' (United Nations 2020) <https://www.ohchr.org/Documents/Publications/OHCHR_BerkeleyProtocol.pdf> accessed 5 December 2020.

[10] ibid.

covered by international law, ranging from criminal processes against individuals to proceedings against the state.[11]

While we have reason to appreciate the increased visibility of serious international crimes that warrant investigation, these probes come with difficult decisions and raise human rights and ethical challenges. The appeal of open source investigations in allowing investigators to map violations across time and space pull these concerns to the fore. Open source research can, for example, uncover the identity of witnesses, victims or perpetrators and the location of a crime in near real time, potentially placing these people at risk. Open source research relies on the collection and analysis of large data sets that create and expose patterns in data. Human rights investigators and their organisations not only need to be aware of the ethical challenges this form of research presents; they also must consider and integrate their responses to them into as they plan, execute and make public their research.

This paper shows that, while local laws, such as the General Data Protection Regulation (GDPR) in the European Economic Area, which place duties on institutions and organisations' operations, need to be followed, this may not be enough to prevent open source investigation methodologies still interfering with the full enjoyment of human rights of those who capture, appear in or share digital open source information - including the rights to life, private and family life, home, health, and freedom of expression. This paper discusses these challenges to guide human rights organisations and their research teams in tackling ethical and human rights challenges when using digital open source investigations techniques. The goal is to provide points not only for organisations experienced in open source techniques to consider and implement, but also for those beginning to integrate these techniques into their work. As Zara Rahman and Gabriela Ivens highlight, “the end mission of defending human rights and revealing rights violations means that investigators should be particularly cautious about their actions and understand the responsibility they carry. In essence: human rights should not be violated during the process of a human rights investigation”.[12]

This paper aims to define both a Human Rights-Based Approach (HRBA) to digital open source investigations and to explain the ways this approach ensures that human rights organisations do not adversely affect the enjoyment of the human rights they seek to protect. Section II, “What is a HRBA?” defines a HRBA and its operational features. Section III defines a digital open source investigation workflow.

[11] Fred Abrahams and Daragh Murray, ‘Open Source Information: Part of the Puzzle’ in Sam Dubberley, Alexa Koenig and Daragh Murray (eds), *Digital Witness: Using Open Source Information for Human Rights Investigation, Documentation, and Accountability* (Oxford University Press 2020).a

[12] Zara Rahman and Gabriela Ivens, ‘Ethics in Open Source Investigations’ in Sam Dubberley, Alexa Koenig and Daragh Murray (eds), *Digital Witness: Using Open Source Information for Human Rights Investigation, Documentation, and Accountability* (Oxford University Press 2020).

Section IV, “Rights at risk during an open source investigation”, produces a non-exhaustive list of human rights risks that digital open source investigations present to human rights NGOs or investigators. In addition, it gives examples of these challenges. Section V, “Mitigating human rights risks throughout the open source investigation cycle”, identifies how human rights can be adversely impacted when refusing to take this approach. It also suggests ways to avoid or mitigate these effects at each stage of the investigation. The case studies presented in section VI present several ethical scenarios for the digital open source researcher to consider.

Two workshops with digital open source investigators inform this report: a workshop held by the Human Rights, Big Data and Technology Project at the University of Essex in 2019 and a workshop at Human Rights Center at the University of California, Berkeley in 2020. In addition, it incorporates 14 interviews with both human rights open source investigators and technologists who build tools to support these investigations. These interviews were conducted with the agreement of anonymity in order to protect the safety of those being interviewed. This anonymity also allowed participants to speak more freely and provide a more detailed and authentic account of their actions. Also integrated in this paper is the extensive feedback from a wide-range of academic and civil society experts in human rights open source research. This has enabled us to provide first-hand accounts by researchers about the quandaries encountered and contribute to a better understanding of the tools serving the open source research community.

Our desired outcome is to outline a Human Rights-Based Approach to digital open source investigations that is hands-on, practical, and accessible. We hope it offers material to open source human rights investigators and other organisations applying such an approach to their research, with discussion points that ensure that this rapidly-developing field always remains focused on the victim.

II. What is a HBRA?

A. Introduction

Human rights organisations position themselves as positive influencers on human rights. Yet, if careful attention is not paid, human rights organisations can also adversely impact human rights through their research methodologies - including open source investigation methodologies. For example, they might risk the rights of individuals or groups, or mistreat employees or collaborators. For instance, human rights organisations may analyse and publish videos or images that identify participants in protests, risking their rights to privacy, liberty or freedom of assembly. Amplifying videos of individuals tortured or humiliated by security forces could lead to re-traumatisation or re-victimisation, compromising rights to health or dignity.[13] The ill-considered publication of open source videos or photographs may also undermine the accused's rights to a fair trial and the victim's right to justice.

Therefore, this report advocates for the adoption and implementation of a Human Rights-Based Approach (HRBA) by human rights organisations conducting research in open source investigations. Often referred to as the rights approach or rights perspective, this framework is based on international human rights standards.[14]

B. Defining a HRBA

Under international law, state actors and institutions must respect, protect and fulfil human rights. However, unlike national laws, such as those that incorporate GDPR, that establish legal duties, international law imposes no direct requirements on human rights organisations. Consequently, by integrating a HRBA framework into all their organisational processes and procedures, NGOs and human rights organisations can limit harming human rights in digital open source investigations or other activities, avoiding jeopardising not only their policies but also their reputations.

While no universal HRBA exists, human rights-based approaches in general come from the system of rights and corresponding obligations established by international law and human rights standards and principles stemming from the Universal Declaration of Human Rights and other human rights instruments. These include all civil, cultural, economic, political and social rights, as well as the right to development. As such, a HRBA requires that the human rights principles of universality, indivisibility, equality and non-discrimination, participation and accountability guide actions with potentially adverse effects on human rights.[15]

[13] Mitchell Paquette and Ariela Levy, 'How OSINT Helps Us Hold Governments to Account during the COVID-19 Pandemic' (*Citizen Evidence Lab*, 1 May 2020) <https://citizenevidence.org/2020/05/01/osint-covid-19-pandemic/> accessed 5 December 2020.

[14] 'UNSDG | Human Rights-Based Approach' <https://unsdg.un.org/2030-agenda/universal-values/human-rights-based-approach, https://unsdg.un.org/2030-agenda/universal-values/human-rights-based-approach> accessed 5 June 2021.

[15] 'UNSDG | The Human Rights Based Approach to Development Cooperation Towards a Common Understanding Among UN Agencies' <https://unsdg.un.org/resources/human-rights-based-approach-development-cooperation-towards-common-understanding-among-un, https://unsdg.un.org/resources/human-rights-based-approach-development-cooperation-towards-common-understanding-among-un> accessed 16 October 2021.

C. Due diligence assessment in four steps

Human rights organisations should act in a way to foresee and mitigate any negative human rights impacts of their work. One way to do this is by adopting the framework provided by a human rights-based approach and conducting a human rights impact assessment before, during and after an open-source research project. The UN Guiding Principles on Business and Human Rights suggest four steps for a due diligence framework.[16] While these are not directly applicable to human rights organisations, the framework provided by these steps is useful conceptually. These steps are:

i. **Identification**: Identification should be split into the identification of relevant human rights and relevant groups potentially affected by rights abuses.

a. Identifying relevant human rights under a HRBA requires a holistic understanding of human rights protected under international law and their relevance to operational or strategic activities in human rights organisations.

b. Human rights organisations engage with diverse groups, including citizen journalists, victims of human rights violations, groups susceptible to human rights violations – like children, seniors or people with disabilities – and human rights organisations' employees. Different risks affect each group's rights.

ii. **Prevention**: once rights and groups at risk are identified, steps should be taken to prevent rights abuses. This means considering how human rights organisations can prevent or minimise any harm while conducting digital open source investigations.

iii. **Mitigation**: A HRBA's effectiveness in an open-source investigation depends on developing a customised plan for each scenario throughout each stage to mitigate potential abuse of human rights law. The process of conducting a due diligence risk assessment, therefore, involves not just stating the risks in the investigation, but also articulating the steps to mitigate these risks. These actions both stop the risks from occurring and provide action plans in case of an identified risk.

iv. **Accountability**: Human rights organisations need to establish transparency and accountability procedures for their open source work. This includes showing research and verification processes and allowing individuals who are harmed in an open source

[16] 'UNSDG | Human Rights-Based Approach' (n 14).

investigation to have access to a remedy. An effective remedy should ensure three key elements: prevention, redress and non-recurrence.[17]

D. The risk of amplification

For a digital open source investigation, it is key to consider to ways the investigation results are amplified. As discussed in further detail in section III, a digital open source investigation can reveal private information about individuals or groups. It may also bring underexposed public information to the fore. If discovered, a human rights organisation could be tempted to amplify this information to either raise awareness of human rights abuses that have happened, or to prevent further or future abuses to groups and individuals. The results of open source investigations also appeal to human rights organisations not only because they provide strong evidence, but also because investigation results can be included in advocacy campaigns – particularly in effective audiovisual campaigns.18 Before amplifying the findings of an investigation, however, the human rights organization must consider if amplification will incur further risks for individuals or groups. This can include, for instance, revealing their location, or revealing their identity. An investigator explained to us how their organisation's amplification of a video put groups at risk: "We had published a video gathered from YouTube depicting certain people in a geographic area at a time that was opposition controlled, that area is now government controlled and having that video online puts people at risk and for various reasons the video could not be removed from YouTube. We got contacted by the group who asked us to take the video down from being publicly visible".[19] This example also highlights the importance of the iterative process and being prepared to evaluate risk at every stage of the research process. When this video was published, there was limited risk. When the situation changed, risks emerged and action was taken to make the video less publicly discoverable. Another interviewee told us how they "used dating apps to gather more information about a person. Having knowledge of a person's sexual orientation and whether it was legitimate and whether it was ethical to draw from those sources was a big question".[20] They realised that including this information in their investigation "would have exposed someone's private life",[21] and ultimately chose not to use it.

[17] 'The Universal Declaration of Human Rights at 70: Putting Human Rights at the Heart of the Design, Development and Deployment of Artificial Intelligence' (*HRBDT*, 20 December 2018) <https://www.hrbdt.ac.uk/the-universal-declaration-of-human-rights-at-70-putting-human-rights-at-the-heart-of-the-design-development-and-deployment-of-artificial-intelligence/> accessed 16 October 2021.

[18] Abrahams and Murray (n 11).a

[19] 'Interview with Interviewee A2, May 2019'.

[20] 'Interview with Interviewee A8, June 2019'.

[21] ibid 8.

Even if an organisation chooses not to amplify a video or photograph in a press release or campaign, they still can send relevant information to authorities, UN rapporteurs or litigators who may amplify evidence for their own purposes. Recognising, therefore, that risks are not just associated with an investigation but also with what happens after its completion is important in ensuring human rights compliance in human rights open source investigation.

III. The digital open source investigation workflow

In this section we outline a typical open source investigation workflow to help open source investigators operationalise a HRBA.

The availability of digital open source information shared on social media platforms has grown in the past decade. Consequently, as Lindsay Freeman notes, "social media is becoming more and more important in international criminal and human rights investigations".[22] Today, investigators can initially react to crises by scouring the internet for evidence of crimes. Of course, this is not always possible. The infiltration of smartphones and mobile internet connectivity is not universal, and governments can resort to internet shutdowns to block content sharing.[23] However, it is still routine practice today for human rights organisations to use open source research methodologies, or at least to explore the possibility of adopting them when starting a broader investigation into human rights abuses.

Here we discuss the workflow proposed by the Berkeley Protocol on Digital Open Source Investigations.[24] It outlines six steps in the digital open source investigation cycle, followed by discussing reports on findings. While we apply the Berkeley Protocol workflow, other human rights organisations have published their own outlines of similar workflows, including Mnemonic,[25] Amnesty International's Crisis

[22] Lindsay Freeman, 'Prosecuting Atrocity Crimes with Open Source Evidence' in Sam Dubberley, Alexa Koenig and Daragh Murray (eds), *Digital Witness: Using Open Source Information for Human Rights Investigation, Documentation, and Accountability* (Oxford University Press 2020).

[23] '#KeepItOn: Fighting Internet Shutdowns around the World' (*Access Now*) <https://www.accessnow.org/keepiton/> accessed 20 May 2021.

[24] Freeman, Koenig and Stover (n 9).

[25] 'Syrian Archive: Methods and Tools' <https://syrianarchive.org/en/about/methods-tools> accessed 5 December 2020.

Evidence Lab,[26] the Digital Verification Unit at the University of Essex's Human Rights Centre,[27] and the Global Legal Action Network.[28]

A. Online inquiry

The process starts with an online inquiry, the step when investigators begin searching for content with digital open source techniques. This may entail searching on social media platforms, finding the coordinates to review satellite imagery, or looking for large data sets on government databases. Some frequently refer to this step as the discovery phase. The investigation moves on to the next step of preliminary analysis if potential relevant content is found at this stage. If investigators find no evidence in the inquiry stage, they can stop.

B. Preliminary analysis

With results from the online inquiry stage, the investigation can transition into the preliminary analysis stage, where the investigators conduct the initial examination of the results of their online inquiry. Several questions typically asked at this stage include: Does the satellite imagery indicate a human rights violation? Do online videos depicting crimes and perpetrators merit further investigation or an attempt to verify them? Can the dataset available be analysed? Can the investigation yield results? This moment is key in deciding whether an investigation is worth pursuing.

All research at this stage of the open source investigation must be considered as biased. The bias has several forms. While Scott Edwards argues that overt human rights abuses are easier to document through open source investigation at the expense of investigating less overt abuses,[29] Yvonne McDermott, Alexa Koenig and Daragh Murray note that researchers should be aware of technical and cognitive

[26] Mitchell Paquette and Sam Dubberley, 'An Open Source Methodology for Mapping Tear Gas Misuse (and Other Human Rights Abuses)' (*Citizen Evidence Lab*, 12 June 2020) <https://citizenevidence.org/2020/06/12/dvc-methodology/> accessed 5 December 2020.

[27] Frederik Aahsberg and others, 'Introductory Guide to Open Source Intelligence and Digital Verification' (University of Essex Human Rights Centre Clinic 2018) <https://www1.essex.ac.uk/hrc/documents/Introductory_Guide_to_Open_Source_Inteligence_and_Digitial%20Verification.pdf>.

[28] 'Bellingcat Yemen Project Methodology and Workflow' <https://yemen.bellingcat.com/methodology/workflow> accessed 5 December 2020.

[29] Scott Edwards, 'Open Source Investigations for Human Rights: Current and Future Challenges' in Sam Dubberley, Alexa Koenig and Daragh Murray (eds), *Digital Witness: Using open source information for human rights investigation, documentation and accountability* (Oxford University Press 2020).

biases in an investigation.[30] This is why investigators or researchers must understand the implications of integrating bias into their investigation from the beginning. We discuss challenges presented by these biases in section V.B in more detail.

C. Collection stage

If researchers decide to pursue the investigation, they usually proceed by collecting links from social media posts or other data. Then, they enter these into a documentation platform. The means of collection differ depending on why the information is gathered. This impacts the way information is stored. Sometimes, it is stored in a spreadsheet. But researchers may also employ specialised software. If the investigator aims to go beyond advocacy or public reporting to contribute their research to legal proceedings, they must follow stringent requirements in the collection stage.[31] Through the process of structuring collected data, the researcher starts to create and analyse new data. This means they spot patterns in the data indicating human rights violations, which allows them to generate links between individual identities and events.

D. Preservation stage

The preservation stage may occur simultaneously with or after the collection stage. At this stage, the researcher stores and preserves the collected data so it can be retrieved later. The investigator will automatically or manually scrape and save digital items like videos, photographs, audio files or social media posts in accessible local storage. Due to the ephemeral quality of digital open source information, the preservation stage is important. Data disappears from the digital open source space for a host of reasons, including removal by social media platforms.[32] As such, the researcher must confront risks that emerge when they scrape data and move it from a public space into their private space. These risks include the challenge of securely hosting sensitive content on servers or removing the content creator's right to delete traces of content they created.

[30] Yvonne McDermott, Alexa Koenig and Daragh Murray, 'Open Source Information's Blind Spots: Human and Machine Bias in International Criminal Investigations' [2021] Journal of International Criminal Justice.

[31] Alexa Koenig and Lindsay Freeman, 'Open Source Investigations for Legal Accountability' in Sam Dubberley, Alexa Koenig and Daragh Murray (eds), *Digital Witness: Using Open Source Information for Human Rights Investigation, Documentation, and Accountability* (Oxford University Press 2020).

[32] Belkis Wille, 'Video Unavailable: Social Media Platforms Remove Evidence of War Crimes' (Human Rights Watch 2020) <https://www.hrw.org/report/2020/09/10/video-unavailable/social-media-platforms-remove-evidence-war-crimes> accessed 3 February 2021.

E. Verification stage

Once researchers have collected and preserved the digital items, they should conduct a rigorous verification process to check the information's veracity and reliability. They can proceed in three different ways, depending on the open source information analysed: source evaluation, content analysis and technical analysis.

- Source analysis looks at the account posting content to a social media platform by generating questions about the source that include: Does the account appear to belong to a user, or is it a bot or troll account? Does the account appear reliable? Is the account holder a partisan in a conflict?

- Content analysis attempts to identify a video or photograph's capture location and capture time. This stage in the workflow also seeks details such as military uniforms and ranks, types of weapons used or the identity of people in the images. Consulting experts – for example, military experts or medical experts – may participate to confirm the content's finer details.

- Technical analysis involves analysing video or image files to pinpoint evidence of tampering or to uncover metadata contained in the media file that will assist in the verification.

The verification stage usually consumes the most time and can vary in length depending on the information's size or complexity. Steps needed to complete and document the verification process concretise the crime but may also violate the right to the content creator's privacy when determining whether they are victims or perpetrators. Such risks increase if organizations or researchers do not use robust data management techniques that also consider laws such as the GDPR.

F. Investigative analysis

The workflow cycle ends with investigative analysis, which determines arguments verified and supported by open source information. It also establishes the ways these arguments fit into a wider human rights investigation. Open source investigations can corroborate event details, and the researcher must decide if their open source content analysis supports allegations of violations of international human rights, humanitarian or criminal law. For example, investigators may need to determine if shadow-analysis (i.e. analysing shadows cast by objects in an image or video as if they are a sun dial) techniques can corroborate the time of an airstrike on a medical facility given by an eyewitness in an interview.[33] They also might have to confirm that remote sensing has found large groups of migrants at a border.[34] They might need to authenticate videos of security forces torturing detainees.[35] Using information requires discussion at the organisational level about verification accuracy, existing gaps in analyses requiring transparency and ways the open source investigation supports other elements of the research conducted.

G. Reporting on findings

Once the researcher has completed their investigation, they should decide if and how to publish their enquiry results. Publication takes many forms. Some digital open source information may be published on its own while some will combine "with field research and other methods to help reconstruct a complete accurate account of violations, which may convince the public, policymakers and, if relevant, judicial bodies to hold perpetrators to account".[36]

Facilitating the continuous and iterative cycle in the open source investigation workflow is crucial to maintain the respect and protection of human rights. At every stage of the workflow in the due diligence process, the researcher must reassess their approach. By adjusting the workflow when needed, they create opportunities for responding to new human rights challenges and threats.

IV. Rights at risk during an open source investigation

A. Introduction

As noted in Section II C, identifying relevant the human rights and relevant groups potentially affected by rights abuses is necessary to conduct a due diligence

[33] Aric Toler, 'How to Verify and Authenticate User-Generated Content' in Sam Dubberley, Alexa Koenig and Daragh Murray (eds), *Digital Witness: Using open source information for human rights investigation, documentation and accountability* (Oxford University Press 2020).

[34] 'Thousands of Ethiopians Are Detained in Nightmare Conditions in Saudi Arabia' (*Amnesty International*) <https://www.amnesty.org/en/latest/news/2020/10/ethiopian-migrants-hellish-detention-in-saudi-arabia/> accessed 5 December 2020.

[35] 'Mozambique: Torture by Security Forces in Gruesome Videos Must Be Investigated' (*Amnesty International*, 9 September 2020) <https://www.amnesty.org/en/latest/news/2020/09/mozambique-torture-by-security-forces-in-gruesome-videos-must-be-investigated/> accessed 5 December 2020.

[36] Abrahams and Murray (n 11).

assessment. A digital open source investigation can affect the rights of a broad range of the following stakeholders:

i. Subjects of data. For example, if caught on camera, they can be identified in a dataset. If an investigator uses the social media posts of a family member to track a person of interest, the subject's family is also at risk. Other members of a group not in an image could also be put at risk if the location of a group is identified.

ii. Producers or sharers of data. This might involve someone posting a video to a social media account.

iii. Receivers or users of data. For instance, someone carries video files on their phone to share on social media later.

iv. Reactors to data. Posting comments or 'likes' on social media, for example.

Less recognised are the risk to an investigator's wellbeing, which include:

i. Digital security risks.

ii. Targeting through trolling.

iii. Consumption of visual, written or audio content depicting extreme violence or its effects may lead to vicarious traumatisation.[37]

In this section we look first at rights at risk during an investigation in general and then we will consider the specific individuals who can be affected.

B. Examples of rights at risk in the context of open source investigations

1. Right to privacy

The right to privacy enjoys protection under UN and regional human rights treaties, as well as under the constitutions of a majority of states around the globe. Article 17 of the International Covenant on Civil and Political Rights states, that "No one shall be subjected to arbitrary interference with their

[37] Sam Dubberley and others, 'Digital Human Rights Investigations: Vicarious Trauma, PTSD, and Tactics for Resilience' in Sam Dubberley, Alexa Koenig and Daragh Murray (eds), *Digital Witness: Using Open Source Information for Human Rights Investigation, Documentation, and Accountability* (Oxford University Press 2020).

privacy, family, home or correspondence, nor to attacks upon their honour and reputation. Everyone has the right to the protection of the law against such interference or attacks”.[38] Privacy concerns prevail through all stages of the open source investigation: from the discovery and collection of data, to its publication, it is key to protect this right because privacy abuses pave the way for abuses of other rights. For example, identifying a protester can threaten their right to life or security. It is vital for any open source investigator or human rights NGO conducting an open source investigation to understand the severe implications of privacy abuses.

Investigators must recognise the privacy abuses that can occur on camera and that videos of human rights violations capture different categories of people, including:

i. The perpetrators of human rights violations
ii. Colleagues of the perpetrators of human rights violations
iii. The victims of human rights violations
iv. Bystanders

People caught on camera can rarely stop a video’s publication. For victims or bystanders, this unsolicited public exposure may make them vulnerable to different forms of retribution, ranging from public targeting by authorities or non-state actors to judicial harassment or punishment.[39] Publication – and, therefore, the non-consensual loss of privacy – triggers a host of other risks to violations of individual rights, depending on their identity, location and the reactions of state or non-state actors when confronted with public exposure of their practices. As noted above, revealing a person’s identity opens the path to further abuses such as: unlawful killings, torture, inhuman or degrading treatment, deprivation of liberty arrest or detention, dismissal from employment or the social stigmatisation of the person or their family members.

Also, risks are not limited to appearances in videos or photographs. With technological advances in geospatial observation, near real-time satellite imagery covers most of the populated world. For example, the satellite

[38] ‘Universal Declaration of Human Rights’ (*United Nations*, 6 October 2015) <https://www.un.org/en/universal-declaration-human-rights/> accessed 5 December 2020; ‘International Covenant on Civil and Political Rights’ (*United Nations Human Rights Office of the High Commissioner*) <https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx> accessed 5 December 2020.

[39] Sam Dubberley, ‘Protecting the Victim’s Identity. Should We Do More to Protect the Identity of Victims Featured in Eyewitness Media?’ (*First Draft News Footnotes*, 12 March 2015) <https://medium.com/1st-draft/protecting-the-victim-s-identity-3b7df432ec09> accessed 5 December 2020.

company Planet Labs claims its “constellation of satellites orbit the poles every 90 minutes, capturing the entire earth's landmass every day”.[40] While it is not possible to see individuals through commercially available satellite imagery, large groups of people are visible, such as refugees or internally displaced people massed at borders fleeing conflict. While human rights organisations may wish to publicise their analysis and bring attention to such movements, they should consider the risk to those groups first – such as parties to the conflict targeting these groups because their location has been revealed.

Risks to privacy rights go beyond being captured on camera. Similar implications exist for those capturing and sharing the content. Could an image shared with an open source researcher contain metadata identifying the phone model that captured the image? Can the Twitter or YouTube account that share videos of abuse be traced back to an individual?

Evaluating risks to the right to privacy must begin at the earliest stages of digital open source investigations and be implemented at every subsequent stage of the investigation. Content assessments made public through the media or in closed briefings with officials may entail decisions to dissimulate or blur faces, or to not show faces at all.

2. Right to life and freedom from torture, inhuman or degrading treatment

The right to privacy is the right that an open source investigation is the most probable to violate. In a worse scenario, a human rights organisation’s decision to publish an image or video can put the right to life or freedom from torture, inhuman or degrading treatment at risk.

For example, in 2019, a young protester in Venezuela filmed a demonstration against the government and then posted the video to his social media platforms. He was identified by the security forces, who then killed him. His family blamed his killing on the video posted two days before.[41] Similarly, in 2018 in Nicaragua, Valeska Alemán, a young protester, live-streamed a video of herself and other protesters as they were fired at by police in anti-

[40] ‘Satellite Imagery and Archive’ (*Planet*, 11 May 2021) <https://planet.com/products/planet-imagery/> accessed 16 May 2021.

[41] Fabiola Sanchez, ‘Deadly Crackdown Stokes Fear among Protesters in Venezuela’ *AP NEWS* (20 February 2019) <https://apnews.com/article/4ff1b4d39b7e409da3e496a8745394fd> accessed 7 December 2020.

government protests that had swept the country. The video went viral. Alemán was then detained twice and tortured before fleeing the country.[42]

Integrating images of torture or other inhuman or degrading treatment into published research risks subjecting the victims to the treatment human rights organisations are trying to prevent. While this has always been an issue for human rights organisations, the unprecedented volume of content now available from crises magnifies the issue. Publishing images of torture victims in detention can increase their risk of further harm in detention. Publishing images after their release might cause them to be rearrested or retraumatised. To mitigate against such risks, for example, Amnesty International decided not to highlight harassment targeted towards transgender people by security forces implementing curfews in several countries at the beginning of the COVID-19 pandemic because they identified risks of retraumatisation through amplification.[43] This exemplifies the benefits of due diligence by regularly updating impact assessments at every research stage.

3. Right to liberty and security

The right to liberty and security of individuals depicted in open source audiovisual content should also be considered. When videos or photographs of protestors circulate and are amplified by human rights organisations, steps must be taken to mitigate the risk of retribution by state security forces or non-state actors. An example comes from Iran, where women filmed protesting against the obligation to wear headscarves in public were reportedly arrested after they shared their videos on social media.[44]

Individuals have had their rights abused because they have shared images and videos to personal social media accounts. Human rights organisations must understand the ways in which collecting, analysing and amplifying data increases the risk to the rights of individuals.

The list of risks to individual rights outlined here is far from exhaustive. It intends to provide examples for discussion by open source investigators and their organisations. It also highlights the importance of integrating a due

[42] Kevin Sieff, ‘Trump Pandemic Border Policy Sends Asylum Seekers Back to Ortega’s Nicaragua’ *The Washington Post* (28 August 2020) <https://www.washingtonpost.com/world/the_americas/nicaragua-asylum-us-border/2020/08/27/9aaba414-e561-11ea-970a-64c73a1c2392_story.html> accessed 7 December 2020.

[43] Paquette and Levy (n 13).

[44] Eliza Mackintosh, ‘Iranian Police Arrest 29 Women over Hijab Protest - CNN’ *CNN* (3 February 2018) <https://edition.cnn.com/2018/02/02/middleeast/iran-arrests-29-women-after-hijab-protest-intl/index.html> accessed 7 December 2020.

diligence process by conducting a risk assessment before the research process begins, a process that will be evaluated at each stage of the workflow.

4. Risks to individuals off camera

Individuals captured on camera are not the only people at risk because their information has been shared online; in addition, risks exist for other groups in an investigation. In this section, we consider specific risks to the data creator and to the open source investigator.

a) The data creator

Individuals who create content depicting human rights violations and then make it publicly available are vulnerable to human rights abuses and retribution by state or non-state actors. For example, reports from the Syria conflict show activists killed for filming and sharing abuses committed by the Islamic State.[45] In Hong Kong, people who used cameras during anti-government protests in 2019 were detained and, in some cases, allegedly tortured in custody.[46] Steps in the open-source workflow, as outlined above, call for inquiry, collection, preservation, and verification. In addition, these steps involve collating and analysing already-existing data points to create new ones. This information may be amplified during final analysis or if included in reports and advocacy campaigns. The collected and analysed data require particular care and diligence. However, this risk also must be balanced with the data creator's freedom of expression. Despite the risk of further human rights abuses committed by governments or non-state actors, activists or interested citizens still film and share content to publicise their story to the world, and often tag videos and photographs on social media which identify usernames of large human rights organisations, prominent human rights defenders or politicians. When open source researchers gain consent and ask questions while conducting due diligence, they must not remove agency from someone who has willingly shared evidence of potential human rights abuses; it will harm their freedom of expression.

[45] 'Islamic State Conflict: Raqqa Activist Killed in Syria' *BBC* (17 December 2015) <https://www.bbc.com/news/world-middle-east-35122224> accessed 7 December 2020.

[46] Trey Smith, 'In Hong Kong, Protesters Fight to Stay Anonymous' (*The Verge*, 22 October 2019) <https://www.theverge.com/2019/10/22/20926585/hong-kong-china-protest-mask-umbrella-anonymous-surveillance> accessed 7 December 2020.

b) The open source human rights investigator

The digital open source researcher takes risks when conducting an investigation. They can, for example, be identified, targeted or trolled by those who support the abusers' cause.[47] This situation arises not only when the investigation results are published but also through the misuse of open source tools. An author of this report made a mistake made when investigating a person of interest through LinkedIn. The problem was that this researcher was logged into their personal account, and the person of interest viewed their profile.

The digital devices of local human rights defenders also run an increased risk of hacking.[48] This is particularly challenging not only because techniques for hacking devices grow in sophistication all the time, but also because the defenders can be careless, inattentive or lacking proper security protocols. In most cases, the researcher or human rights defender is unaware their device has been compromised. No matter how hacks occur, the researcher's personal information becomes vulnerable to doxing, which can include details about their contacts.

In addition to the risks human rights investigators face from abusive regimes, a digital open source investigation can, if not properly supervised by the human rights organization, also undermine the investigator's mental and physical health. Investigations expose the researcher to disturbing or traumatising data, which includes images of traumatic death or torture, accounts of sexual violence or satellite imagery showing destruction caused by long-term bombing campaigns. As such, the digital investigator risks exposure to vicarious traumatisation which can lead to post-traumatic stress disorder.[49] Much research and progress has been made regarding trauma in recent years, and some larger human rights organisations are starting to offer training and counselling to support the traumatic nature of the work. However, this requires serious engagement from managers and organisations, who must view this as essential instead of a 'nice to have'.

The open source investigator should always think through their risks when starting the research process. This entails interacting with grassroots human rights defenders or other contacts and integrating their physical, psychosocial

[47] 'Social Media Trolls Threaten Activists' (*Human Rights Watch*, 16 March 2017) <https://www.hrw.org/news/2017/03/16/social-media-trolls-threaten-activists> accessed 18 May 2021.

[48] 'Massive Data Leak Reveals Israeli NSO Spyware Used to Target Activists, Journalists, and Political Leaders' (*Amnesty International*, 18 July 2021) <https://www.amnesty.org/en/latest/press-release/2021/07/the-pegasus-project/> accessed 16 October 2021.

[49] Dubberley and others (n 37).

and digital security needs into each step of the research process starting from the beginning. They must also assess the subsequent risks when contacting sources on the ground or when publishing results that may adversely impact the investigator if trolled.

5. Risks to children

Protecting the rights of children in an open source investigation requires extra due diligence, when both conducting research and deciding to depict them. Some organisations establish a standard practice of not publishing images of identifiable children. If the violation or crime is so egregious that the human rights organization decides that images must be published to demonstrate the gravity of the violation, such as displaying bodies of children killed in airstrikes, then the identifying features of the children should be edited out of published content or blurred to remove any identifying features. Children's rights are equally important when conducting research through social media. For instance, if a person of interest is identified or located in an Instagram photo of a family wedding that has been posted by a young friend or relative, questions should be asked about the details of the person posting the image in case they are at risk. One interviewee explained their policy: "when we are trying to find out more information around a politically exposed person, we go through their affiliated friends and family on Facebook. If we go for a politically exposed person through family and friends, we would never publish that publicly, with the understanding that it is private not for publication by us".[50]

Digital open source investigations may risk the enjoyment of a wide range of human rights for several players: the subjects of the data, the person recording or sharing the data, investigators themselves and vulnerable groups such as children. It is through understanding the risks to different groups at each stage of the workflow that the investigator can implement a human-rights based approach to mitigate these risks and ensure that they do not create harm.

V. Mitigating Human Rights Risks throughout the Open Source Investigation Cycle

This section focuses on ways open source investigators can mitigate risks outlined in the preceding section in their work. We discuss specific issues from the investigation workflow and offer questions for the researcher to consider in a risk assessment plan. The accompanying workbooks[51] provides further guidelines about questions to

[50] 'Interview with Interviewee A5, June 2019'.

[51] See https://engn.it/hrosint1 and https://engn.it/hrosint2

raise during risk assessments. As noted previously, the risk assessment plan must be iterative and therefore re-evaluated throughout the digital open source investigation. The following mitigation steps intend to help when implementing the risk assessment plan.

A. Online Inquiry

As described in Section III A, during the online inquiry phase, the open source investigator conducts online searches to discover digital open source information. Depending on the investigation, the investigator's inquiry parameters include searching social media platforms like Facebook, YouTube, Twitter, or LinkedIn for information about a person of interest or videos or photographs depicting an event,[52] or analysing satellite imagery with tools like Google Earth Pro to compare geographical features of a city or village before and after airstrikes to gather evidence of burnings of villages or forced evictions.[53]

1. Setting up the researcher's desktop

Mitigating risks in an open source investigation begins with setting up a secure computer environment because both the researcher and the organisation will be exposed to emerging digital threats. Following recommendations by credible organisations devoted to online security ensure that the open source researcher remains aware of the fast-paced updates in digital security.[54] Choosing and assessing tools is an important step in starting a digital open source investigation. These tools appeal to researchers because they are free or low-cost; however, potential security issues can emerge. In our interviews, we asked open source researchers about how they chose their tools. We asked them to consider:

- Who built the tool?

[52] Paul Myers, 'How to Conduct Discovery Using Open Source Methods' in Sam Dubberley, Alexa Koenig and Daragh Murray (eds), *Digital Witness: Using Open Source Information for Human Rights Investigation, Documentation, and Accountability* (Oxford University Press 2020).

[53] Micah Farfour, 'The Role and Use of Satellite Imagery in Open Source Investigations' in Sam Dubberley, Alexa Koenig and Daragh Murray (eds), *Digital Witness: Using Open Source Information for Human Rights Investigation, Documentation, and Accountability* (Oxford University Press 2020).

[54] An unexhaustive list of such platforms include, at the time of writing, 'Consumer Reports Security Planner' (*Consumer Reports Security Planner*) <https://securityplanner.consumerreports.org/our-mission> accessed 7 December 2020; Tactical Tech, 'Security In-a-Box' <https://tacticaltech.org/projects/security-in-a-box> accessed 7 December 2020.

- Who funds the tool?
- Who else uses the tool?
- Who is the tool's target audience?

As one academic researcher said: "If it was a new tool, we would then vet it with other trusted people, and we would talk to the designers of the technology ourselves. We would never put anything very sensitive in some of the more public archiving tools. We have a tiering process, where we assess appropriateness and complexity, we try to look at investigations holistically".[55] To ensure employee privacy and security, organisations undertaking open source research must train their staff in basic online security and, if possible, appoint an on-site security officer to offer advice, vet new tools or provide additional training.

The following list presents questions human rights organisations must ask about security in an open source investigation**:**

- Can the open source researcher receive training in digital on-line security? Have they completed a risk assessment of the digital environment?
- Is the code used to build the tool open or closed source? Can the code be audited to identify possible security risks?
- How do tools deal with data regarding storage, third-party access and portability. Do they meet the data handling requirements in legislation such as GDPR?

2. Entering the Closed Source World

The investigation's inquiry phase may require a switch from open source inquiry into closed source inquiry. This means that the investigator will use information that is neither easily accessible nor inexpensive. They may need to acquire further information about an uploader or an event. They may additionally ask follow-up questions to assist with verification. Entering closed, members-only groups on social media platforms such as Facebook or messaging services such as WhatsApp or Signal might also be required. These actions turn open source investigations into closed source investigations. While the framework laid out in this report is not designed for closed source, it is still important for the open source investigator to consider the risks of making this move.

[55] 'Interview with Interviewee A5, June 2019' (n 50).

3. Exposure to trauma

Investigators risk exposure to violent, graphic and traumatic digital content during research which impacts their mental wellbeing.[56] This places the investigator's health at risk, and potentially threatens the investigation if they cannot work at full capacity. It protects the individual investigator, the investigation and all involved from harm. If an investigator's wellbeing is compromised, contacts and sources are at risk, and compromise the overall quality of the work.

Questions to consider about trauma exposure:

- Has the investigation team, including the line manager, conducted a trauma exposure risk assessment before starting digital inquiries?[58]
- Are there structures, policies and resources in place within the organisation that allow for mental health concerns to be addressed?

B. Preliminary Analysis

1. What is being investigated and what is not.

It is essential to assess the human resources available, languages spoken and the viability of gathering information through source techniques, at the preliminary analysis stage to ascertain the information biases in open source investigation. Scott Edwards notes that the cause of such biases are "both a function of the information ecosystem, as well as the cognitive biases of the investigator".[57] Yvonne McDermott, Daragh Murray and Alexa Koenig echo Edwards by categorising these biases into "two overarching categories: (a) technical biases, which are biases inherent to decisions made by computer systems, and (b) cognitive biases, which are systematic errors in thinking or reasoning that impact upon human decision-making".[58] For example, with technical biases, researchers must understand the ways algorithms steer searches for content depending on factors such as location, search history, the news cycle and other categories. Cognitive biases affect what the researcher brings individually to the investigation and includes biases formed by people capturing and uploading information. As one researcher interviewed for this paper noted, "A lot of people that are currently doing open source investigation work are English speakers, so this determines what gets

[56] Dubberley and others (n 37).

[57] Edwards (n 29).

[58] McDermott, Koenig and Murray (n 30).

attention, whose voices get heard, who is analysing the social media content".[59] Organisations conducting open source research should be aware that lack of diversity in research teams exacerbates this problem. Alexa Koenig and Ulic Egan observe that open source information about cases of sexual and gender-based violence often goes unmonitored or is missed because investigators do not know where to look, cannot recognize important details or lack the knowledge to interpret it.[60] Ulic Egan reports that open source data collection "encompasses an inherent danger by focusing global attention to armed conflicts occurring in areas with better access to technology, which creates an atrocity bias".[61] Indeed, while digital open source investigations present new, exciting and fashionable methodologies for human rights research, the researcher must remember, as outlined at the start of this paper, that they are complementary, and should be used alongside time-tested research methodologies such as interviewing eye-witnesses. Ignoring other types of methodologies risks missing abuses because open source methodologies appear more attractive.

Questions to consider about biases:

- Has the investigator received training on technical biases inherent in computer algorithms and their implication for the research process?
- How do the preferences or cognitive biases of the researcher or organisation affect ways they collect open source data? Can the investigator make working hypotheses of various outcomes? Is the work peer-reviewed to ensure results and methods are as accurate or transparent as possible to produce high quality investigations?
- Does the allure of open source techniques prioritize an investigation's capacity to see a certain kind of abuse at the expense of not seeing less visible, but perhaps equally important incidents?

C. Data Collection

In the data collection phase, digital open source information is manually saved or automatically scraped from the public internet 'through a screenshot, conversion to

[59] 'Interview with Interviewee A5, June 2019' (n 50).

[60] Alexa Koenig and Ulic Egan, 'Power and Privilege: Investigating Sexual Violence with Digital Open Source Information' (2021) 19 Journal of International Criminal Justice 55.

[61] Ulic Egan, 'Digital Accountability Symposium: Intersectionality and International Criminal Investigations in a Digital Age - Opinio Juris' (*Opinio Juris*, 19 December 2019) <http://opiniojuris.org/2019/12/19/digital-accountability-symposium-intersectionality-and-international-criminal-investigations-in-a-digital-age/> accessed 7 December 2020.

PDF, forensic download, or other form of capture'.[62] Several risks to human rights surface at this stage: the right to privacy, the right to life and the right to freedom of expression.

1. Gaining consent from those generating the data being used.

Data such as photos or tweets updating events posted online through social media platforms is generally considered to be "fair game" by investigators and they do not tend to gain explicit consent for its use from its creators. This problem also lies within the infrastructure of the social media platforms, permission settings and conditions of use terms. Aside from individual sharing preferences that define who can view the post, social media platforms provide no outlet for users to indicate the ways they would like others to be able to use their data. This inability to contribute actively to investigations means that investigators cannot assume they have acquired consent for the use of open source information.

To respect the right to privacy, the human rights investigator's default questions in the data collection phase should consider if steps to receive permission to use open source content in their research are possible without violating rights. This means the investigator must obtain informed consent. Zara Rahman and Gabriela Ivens describe a four-step process to gain informed consent:

i. **Notice or Disclosure** - the consent process must ensure that the person who captured the data is informed of the nature and purpose of the investigation, the foreseeable risks, the ways their information might be reused or shared, the choice they abstain from the investigation and the procedures for confidentiality and anonymity.

ii. **Capacity or Understanding** - if the content creator cannot understand information given to them by the investigator because it employs a specific style or context such as technical jargon**,** then the investigator should provide opportunities for the content creator to not only ask questions, but also to receive comprehensible answers to them.

iii. **Voluntariness** - the consent to participate must be voluntary, free of coercion or inflated promises, and, if possible, not involve people who have power over the participants.

iv. **Competence:** The participant must be competent to give consent.[63]

[62] Freeman, Koenig and Stover (n 9).

[63] Rahman and Ivens (n 12).

As described in the risk assessment outlined previously, the consent process must be flexible and iterative. If consent for a certain type of output is received through a report or press releases, and its focus subsequently changes – for instance to bring a prosecution – then the investigator must seek consent again. Consent procedures must also follow local laws. For instance, GDPR allows anyone to withdraw consent at any time. Therefore, an NGO must ensure that it has consent withdrawal procedures in place if based in a jurisdiction covered by GDPR.

If a large, well-known organization leads the investigation, they should proceed by not raising expectations for the content creator by promising legal remedies for alleged violations. This might leave content creators with the false belief that allowing a large human rights organization to use their video will improve the situation.

Despite best efforts, obtaining informed consent is not always possible. It must be balanced with other concerns and risks to human rights. As an investigator explained: “we have thousands of sources, and in some cases it is difficult to get permission and in some cases impossible. […] We definitely don’t ask in all cases”.[64] The researcher needs also to accept the fact that the content uploader decided to place their content in the public space when they shared it on a social media platform. When ensuring the privacy of those depicted in the content and those who captured the content the following issues must be respected:

i. The safety of the uploader or the content capturer
ii. The risk local activists take to capture and share the possible human rights abuses

In a HRBA, these considerations include risks to their right to life, freedom from torture or degrading treatment or right to liberty and security.

Questions to consider about uploader consent:

- Can fully-informed consent be obtained from an uploader? Or can the act of uploading content be seen as implicit consent for future human rights uses?
- Is it acceptable and safe for the organisation to collect and store an uploader’s data without permission if the research is at the preliminary analysis stage?

[64] ‘Interview with Interviewee A2, May 2019’ (n 19).

D. Preservation

In recent years, data disappearing from social media platforms has presented open source investigators with their greatest challenge. As an open source researcher investigating a large conflict told us: “Content is uploaded and is taken down immediately, if no one is archiving this content – in three years you might have no idea why it was captured or why it mattered [...] people need to know that this occurred”.[65] Preserving digital information by downloading it from a social media company’s platform and saving it locally on an organisation’s servers so the photograph, video or associated metadata cannot be altered solves this problem.[66] Two main approaches have been taken to doing this. One approach has been to ‘preserve everything’ automatically. A second approach is to preserve content on a ‘case-by-case basis’ manually.

1. The mass preservation approach

Mnemonic, a civil society organization focused on the preservation of online content related to suspected human rights violations, implements the ‘preserve everything’ method.[67] This entails, for example, automatically downloading and archiving all digital content matching keywords on social media platforms. This approach has the advantage that it captures content as it is uploaded, minimising the risk that content may be taken down before it is preserved. However, this also means that all content is captured regardless of what it contains. First, this means that the organisation may be storing content the originator may have removed for good reason. It also means that the organisation or research team is potentially archiving large volumes of content that will never be used because the volume is too great to curate or verify, bringing with it infrastructure costs and technical know-how that may not be sustainable for a small organisation and questions around principles of data minimisation.

[65] ‘Interview with Interviewee A8, June 2019’ (n 20) 8.

[66] Wille (n 32).

[67] Jeff Deutch and Niko Para, ‘Targeted Mass Archiving of Open Source Information: A Case Study’ in Sam Dubberley, Alexa Koenig and Daragh Murray (eds), *Digital Witness: Using Open Source Information for Human Rights Investigation, Documentation, and Accountability* (Oxford University Press 2020).

2. The case-by-case approach

The case-by-case method, adopted by organisations such as WITNESS, an organisation that uses video to protect human rights,[68] ensures the possibility to evaluate and consider content before it is stored. However, this approach is risky in a world where take-down algorithms are only improving in speed and efficiency, and where states and other actors use content-flagging to remove critical content before it is preserved. In addition, selective archiving risks creating bias in the evidence. For instance, a researcher may choose to archive video evidence of seemingly excessive use of force by security forces but omit the video taken just before the incident where protesters put the life of a security force member at risk.

Questions to consider in preservation

- Is the infrastructure – digitally, financially and skill-wise – of the organisation sufficiently strong and resilient to host an archive over time without putting human rights at risk?
- If content is removed from a social media platform, is it possible to establish who removed it and for what reason? If the data was removed by the uploader instead of the social media company, different considerations may apply. In this case, the uploader's decision to remove the content should be respected and removed from the archive.
- In addition to the adherence to local data protection laws, what other mitigation plans exist in case the archive is attacked, risking the rights of those whose data is stored?

E. Verification

The investigator most closely interrogates and analyses the digital open source information collected in the verification stage. Additional data not present in the original posting is also generated during this phase. As Aric Toler outlines, this requires several steps, including checking sources and content, as well as using geolocation. Both bring together different data elements not only to create a larger picture of the event but also to determine whether a piece of data is important.[69]

[68] Yvonne Ng, 'How to Effectively Preserve Open Source Information' in Sam Dubberley, Alexa Koenig and Daragh Murray (eds), *Digital Witness: Using Open Source Information for Human Rights Investigation, Documentation, and Accountability* (Oxford University Press 2020).

[69] Toler (n 33).

1. Checking sources

Anonymous accounts share a lot of content on the internet via social media platforms. On Twitter or on Instagram, for instance, an account name can, in theory, seem hard to link to the physical identity of the uploader. However, techniques used by open source investigators can lead to unveiling account identities, or showing the ways Twitter accounts are linked to Facebook accounts and so on. For example, a journalist uncovered that seemingly anonymous Twitter account belonged to former CIA director James Comey.[70] This highlights the risks associated with combining seemingly disparate data points to build a larger picture of an uploader. The potential human rights risks associated with unveiling an identity intensify when verifying accounts of human rights activists in conflict zones or crisis situations. Considering the ways in which this information is safely stored, handled and integrated into research so it does not harm the uploader becomes crucial for the investigation's integrity.

2. Location identification

Unveiling an event's location has the same impact on the integrity of digital open source investigation as revealing an individual's identity. Indeed, part of the digital open source verification process involves geolocation, which is defined as: "the determination of the exact location where a photograph or video was recorded".[71] Frequently, investigators conduct this work thousands of miles away from where the human rights violation occurred. Yet it can be very accurate, and can be conducted by open source investigators who often possess no local knowledge. Potential risks include, for example, exposure of location and possibly the identity of local human rights defenders or persons fleeing violations. If open source investigators many thousands of miles away can identify their location, then local security forces may be able to, too. One investigator interviewed for this research spoke of two cases where they had concerns over revealing the precise location of possible violations. They were also worried about revealing not only the event's location, but also the location from which the video was filmed. This put several people at risk: the individuals filmed in the video, the person filming the content and, potentially, other people disconnected from the event who lived in the same building. An author of this report researching on a situation in a conflict in which their research showed exactly where the event being researched happened and the apartment block from where it was filmed. Because of knowledge of the

[70] Ashley Feinberg, 'This Is Almost Certainly James Comey's Twitter Account' *Gizmodo* (30 March 2017) <https://gizmodo.com/this-is-almost-certainly-james-comey-s-twitter-account-1793843641> accessed 7 December 2020.

[71] Toler (n 33).

regime and concerns about retribution, the digital open source research was not published. This shows an example of appropriate due diligence where the risk of further rights abuse meant the information gathered through digital open source research was not amplified.

Questions to consider about location identification:

- Is data collected and aggregated in the verification process stored safely, and does the organisation have the right infrastructure in place to maintain the data's security?
- Is the investigator prepared to stop an open source investigation if the information uncovered can lead to harm greater than the abuse being documented? For instance, is there risk of reprisal to a documenter if evidence of an abuse is published?

F. Investigative Analysis and Publication

Many questions arise concerning the use and publication of the open source content's final analysis. If the verification process leads to an analysis revealing information otherwise not public, attention should be paid to the cost of reporting this publicly. For instance, in satellite analysis, a recently captured image could show the location of people hiding during an evolving situation. The geolocation of a video can do the same. This information, if published or amplified, could put people – even those not connected to the events documented – at risk of further harm or retribution, thus potentially compromising their right to life or liberty in real time.[72]
If the research includes graphic videos or photographs that show minors, dead people, and others who cannot give their consent, a human rights organisation may also decide not to publish identifying features like faces, or blur images of individuals.

Questions to consider about investigative analysis and publication:

- Who is in a position to make the decision regarding how open source information is shared, used and viewed?
- Should keeping the identities of people depicted in videos anonymous be the rule?. Should revealing the identity be the exception? Has a risk assessment been conducted on the possible impacts of revealing the identity or location of a group?

[72] Nathaniel Raymond and others, 'While We Watched: Assessing the Impact of the Satellite Sentinel Project by Nathaniel A. Raymond et Al.' (*Georgetown Journal of International Affairs*, 25 July 2013) <https://www.georgetownjournalofinternationalaffairs.org/online-edition/while-we-watched-assessing-the-impact-of-the-satellite-sentinel-project-by-nathaniel-a-raymond-et-al> accessed 11 December 2020.

- Is the open source investigator part of their organisation’s quality control process and therefore able to review drafts of reports, advocacy campaigns, or legal submissions before making them public to ensure consideration for human rights risks and harms?

VI. Case Studies

When asking the questions posed above, many case studies show the answers are frequently context specific. To help explain these questions more concretely, we have detailed some real-life situations experienced by digital open source investigators based on interviews we conducted. We are sharing these because our goal is to assist organisations in constructing a HRBA approach for open source investigations thereby generating subsequent due diligence in their work. Many of the scenarios here deal with the impact recorded when amplification alters the visibility of open source information. In order to think critically about these cases, we look at human rights principles and stay devoted to the idea and principle of minimising harm.

1. Scenario 1: Satellite imagery

- **Situation**: There is an uptick in violence in a particular country. Satellite imagery shows that people are sheltering in courtyards of churches and mobs are looking for them. This image has been received from a commercial satellite service. It is not available on open source satellite services such as Google Earth Pro.

- **Issue**: Conflict monitoring is being carried out using real time satellite imagery, and if the detailed satellite imagery is released, intelligence may be able to be provided to a particular actor. This image in question is less than 24 hours old and the people sheltering might still be there.

- **Action**: The satellite image is not released publicly, although other advocacy avenues can be pursued, including reporting that the imagery exists, but not making it public. If it is deemed critical to release the image, then time delays may mitigate the risk for people in the image.

2. Scenario 2: Publishing videos

- **Situation**: A human rights organisation has published a video gathered from a social media platform.

- **Issue**: The people depicted in the video have now become at risk and have contacted the organisation to remove the video.

- **Action**: The organisation withdraws this video from any publications and it adds metadata to the video in its database so it is not used by mistake in future research or advocacy. If the content is used elsewhere by other organisations, the organisation asks them to remove the content.

3. Scenario 3: Publishing videos

- **Situation**: A human rights organisation has a collection of videos it wants to publish of a protest.

- **Issue**: The organisation is concerned that the people documented at the protest will be identified through the videos it wants to publish. It cannot get consent from those depicted in the videos, and is are unsure about their security concerns.

- **Action**: The organisaton determines that the risk to protesters is too great, so a report including a video that does not reveal or blurs faces is published. The organisation notes that it has archived the original content.

4. Scenario 4: Publishing locations

- **Situation**: A human rights organisation is ready to publish a report on an ongoing conflict in which a party to the conflict is targeting civilian objects like medical facilities. The organisation usually publishes location coordinates to add credibility to its reports.

- **Issue**: These coordinates could be used in ongoing targeting of such facilities by a party to the conflict.
- **Action**: It is decided not to publish the coordinates and instead anonymise the data as these locations are still being targeted. The organisation reports that the imagery exists but does not make it public.

5. Scenario 5: Attribution

- **Situation**: A human rights organisation wants to attribute and credit those who both worked on the investigation and produced the data.

- **Issue**: It may be risky to name sources and might give away anonymity in doing so. Doing so may also threaten the personal safety of those involved in the investigation, but this needs to be weighed with the consequences of crediting the work of others.

- **Action**: The organisation conducts a thorough risk assessment that highlights concerns regarding credits given to individual investigators. The pros and cons are put to the collaborating investigator thereby allowing them to decide whether they want to be credited. They might prefer to be credited with a pseudonym instead. To credit the people who produced the data, a similar approach is implemented as to when seeking consent about preserving content from citizen journalists; a risk assessment is conducted about whether it is harmful to contact collaborators and then ask them if they wish to be credited. They are contacted through a secure method.

VII. Conclusion

Digital open source investigations, when intrusive, can lead to the identification of individuals, their families, fellow group members and their home's location. Without careful consideration, these investigations can infringe on a range of human rights: from the right to privacy to the right to health or, in the worst-case scenario, the right to life. With open source investigations in human rights research, a clear need has been established to abide by boundaries corresponding to human rights principles. Investigations aiming to defend human rights must support these rights throughout the entire work cycle. As such, investigators maintain the legitimacy and credibility of the human rights endeavour. This approach, however, relies upon a clear understanding and agreement about the definition of "human rights". The individual researcher and the human rights organisation must work together to align their conception of rights.

Using human rights principles as a framework for open source investigations to monitor abuses can guide decisions on investigation subjects, data preservation and data publication. Without clear standards and rules, human rights organisations conducting digital open source investigations expose themselves to being criticised for using the same techniques as those used by bad actors who abuse human rights. This report, and its associated workbook aim to help the researcher avoid this scenario by setting out a human-rights based approach to open source investigations, based on the principle of human rights due diligence.

Credibility matters as open source investigations push boundaries of the conventional practices employing established methods of human rights research. By taking a human rights-based approach and conducting appropriate due diligence before, during and after an investigation, investigators are encouraged to understand the risks involved not only for themselves but for all people involved in the process.

About the Authors

Sam Dubberley is a research consultant on the Human Rights, Big Data and Technology Project at the University of Essex and managing director of the Digital Investigations Lab at Human Rights Watch. He headed Amnesty International's Crisis Evidence Lab, after joining the organization in 2016 to set up and manage the Digital Verification Corps. At Amnesty, he conducted and led on a wide range of open source research, including the 2021 Webby Award-winning platform "Teargas: An Investigation". He is the co-editor with Alexa Koenig and Daragh Murray of *Digital Witness: Using Open Source Information for Human Rights Investigation, Documentation, and Accountability* (Oxford University Press 2020).

Gabriela Ivens is the Head of Open Source Research and part of the Digital Investigations Lab at Human Rights Watch. She integrates public sources of data and open source research techniques into human rights investigations and works to build capacity within the organisation by training staff, developing processes, and managing the technical development of tools. Previously, Gabriela was a Ford-Mozilla Fellow hosted at WITNESS, and prior to that led the investigative portal Exposing the Invisible. Gabriela holds a masters in Human Rights from University College London.

This work was supported by the Economic and Social Research Council [grant number ES/M010236/1].

The Human Rights, Big Data and Technology Project

Human Rights Centre,
University of Essex,
Colchester CO4 3SQ
+44 (0)1206 872877

@HRBDTNews
www.hrbdt.ac.uk