Bellingcat's Yemen Project

This text is chapter two of ten in the publication *Investigative Methods: An NCRM Innovation Collection*.

How to cite this document

Waters, N. (2022) Bellingcat's Yemen Project. In: Mair, M., Meckin, R. and Elliot, M. (Eds) *Investigative Methods: An NCRM Innovation Collection*. Southampton: National Centre for Research Methods, pp. 19-31. DOI: 10.5258/NCRM/NCRM.00004545

2. Bellingcat's Yemen Project

Nick Waters (Bellingcat)

In this contribution, Waters outlines Bellingcat's use of Open Source Investigations to lift the veil on the abuse of power in the conflict in Yemen. Bellingcat's Yemen Project aimed not only to unearth evidence of incidents, but also to increase the quantity and quality of verifiable data being recorded in connection with the conflict. Waters delineates how this information was used in an effort to hold UK arms companies to account for the sale of weapons to Saudi Arabia for use in the conflict, as well as in a conceptual legal setting to test the admissibility of this kind of information as evidence in the courts of England and Wales.

Introduction

Six years into the conflict in Yemen and reliable information concerning violent incidents is scant. With the exception of several reputable NGOs, reporting on the Saudi-led military campaign in the country is often inaccurate or inaccessible. Lack of access to the scenes of attacks has undermined local and international reporting, shielding both state and non-state actors from accountability and allowing them to nurture an environment of misinformation by exercising airtight control on the conflict's narrative.

For reasons that will be discussed in this paper, Open Source Investigation (OSI)[1] is an untapped resource and a viable solution to many of the problems facing investigators in the context of conflicts such as Yemen's, where information is widely available online but is most useful when augmented with specific local knowledge.

Over the last seven years, Bellingcat has become adept at finding and using this kind of information to investigate events around the world. Initially focusing on the question of who shot down Malaysian Airlines flight MH17 over Ukraine in 2014, Bellingcat has used OSI to lift the veil on the abuse of power around the world. Perhaps most well-known for its work revealing the existence of a Russian state-sponsored assassination program, Bellingcat's work has included tracking police violence in the USA, examining environmental issues and revealing violence perpetrated against migrants and refugees on the border of the European Union.

With its emphasis on holding power to account, Bellingcat decided that the application of its techniques might prove useful in seeking accountability for events in Yemen. Bellingcat's Yemen Project, therefore, aimed to unearth not just evidence of the incidents, but also to increase the quantity and quality of verifiable data that is accessible to an online investigator. This information was then used in an effort to hold UK arms companies to account for the sale of weapons to Saudi Arabia for use in the conflict, as well as in a conceptual legal setting to test the admissibility of this kind of information as evidence in the courts of England and Wales.

The Yemen Project

The Yemen Project was an attempt to provide meaning to an otherwise unmanageable mass of open source information related to the conflict in Yemen. Open source information obtained from social media, local media and NGO reports provided an untapped reserve of information about the Saudi-led Coalition (SLC) air campaign. The project aimed to combine this information with local knowledge and assessments from both formally-qualified and informal subject matter experts in order to produce a series of reports examining the air campaign.

It has become commonplace for people to pull out their mobile phones and film noteworthy events as they encounter them. In the face of even extreme events, like an airstrike, some people's immediate reaction will be to record and then disseminate it amongst their personal networks, whether over social media or messaging apps. The drive to do this appears to be incredibly strong, with many risking their lives in order to capture these kinds of events on film. This isn't generally a voyeuristic action, rather it is a conscious decision to witness, and to inform others of the event and its consequences (Reading 2009).

This behaviour has manifested itself in the most extreme events and conflicts across the world, perhaps most notably in Syria, where the length of footage of the war exceeds the length of the war itself. Time and time again people subjected to the most terrible violence have recorded their experiences in a desperate attempt to show the world what is happening. This content has then been observed and further amplified by many organisations in an attempt to respond to these kinds of events.

In the context of the conflicts across the contemporary Middle East, notably Iraq and Syria, OSI has been instrumental in investigating targeted Russian airstrikes against hospitals, the civilian death toll from the Coalition air campaign and the chemical weapons campaign of the Syria government against its own people. The Yemen Project was intended as a way to formalise for the first time the collection, processing and amplification of this open source content through OSI related to the conflict in Yemen, which could then be used in advocating for justice and accountability efforts, particularly in the context of legal cases.

The Yemen Project initially began as a concept which was then put into practice in the form of a hackathon in London, followed by the writing of a series of reports examining airstrikes in Yemen. Alongside the advocacy element of these reports, which were designed to highlight issues of justice and accountability for airstrikes, they were also submitted to the UK Parliament as evidence, and formed the basis for a mock hearing to test their admissibility in court. In this case study, we will examine each of these elements of the project in turn, noting what succeeded and what did not.

Laying the Conceptual Basis for the Hackathon, Project 'Arim

In 2018, Rawan Shaif returned from Yemen frustrated at seeing first-hand the effects of the Saudi-led Coalition bombardment. She attended a workshop hosted by the Global Legal Action Network (GLAN) and realised that OSI could provide evidence in support of legal cases to prevent the sale of weapons to Saudi Arabia.

Along with Dearbhla Minogue at GLAN and Tara Vassefi, a human rights lawyer, Rawan Shaif at Bellingcat settled on the idea of a hackathon at which a large number of researchers would come together to use OSI to investigate SLC airstrikes.

GLAN surveyed the legal practices and principles relevant to determining the weight and admissibility of evidence and consulted with international and domestic legal practitioners. They then identified a set of basic standards which would address the core evidentiary priorities of demonstrating impartiality, chain-of-custody, record-keeping and the pursuit of all reasonable lines of inquiry. All of these factors serve to address potential legal concerns around the quality and provenance of evidence.

Drawing on GLAN's work in preparation for this event, we created a replicable operational methodology that was written according to evidentiary standards for potential use in impact-oriented investigative journalism, academia, and legal endeavours. We drew upon the knowledge

of multiple open source investigators, as well as the expertise of specialists in archiving, journalism and the law. With this methodology we aimed to establish a benchmark for the practicalities and efficacy of conducting remote open source investigations and publish a series of high-quality assessments.

These standards were specifically designed to be "light touch": that is, to be feasible for investigators to adhere to without unduly slowing down their progress. This was the first stage in an on-going process through which this ground-breaking methodology will be constantly refined. The methodology focused on four aspects: searching, preservation, verification and analysis. Each part of this methodology was geared to increase the likely value of this kind of open source information as evidence in court. Simple steps, such as recording search terms used, clearly outlining verification steps taken and the use of analytic language in written reports were key to this. Perhaps the most significant part of the methodology was the forensic preservation of content. Within the project, this function was carried out by the Yemeni Archive, part of Mnemonic (whose work is discussed further below). Unfortunately, as well as being the most significant step in the methodology, it is also the most difficult for external investigators to replicate, as it requires a partnership with an organisation which has the ability to forensically preserve digital media.

The exercise took its name from the 'Arim Dam in Marib, Yemen. Archaeologists suggest construction started around 2000 BC. An engineering marvel of the ancient world, it is the world's oldest known dam. The Sabaeans built the dam to store the periodic monsoon rains that fell on the nearby mountains, allowing the collected water to be redirected into an irrigation system that nurtured the kingdom's vast gardens and made agriculture possible. Despite several breaches over its long history, including an airstrike during the conflict, parts of the dam still stand today. In naming this exercise after the dam, we sought to signal our hope that the framework we sought to devise would have some degree of longevity and durability.

Phase 1: Project 'Arim – The Hackathon

In late January 2019, for four days, Bellingcat and GLAN brought together over 40 open-source intelligence (OSINT)-trained investigative journalists, technologists, and lawyers to combine their expertise and investigative capacity to discover, verify, analyse and preserve a dataset of 100 aerial bombardments allegedly carried out by the Saudi-led Coalition occurring in Yemen between the 25th of March 2015 – 31st of December 2018.

Hackathon attendees included formally qualified subject matter experts, such as ammunition specialists, linguists, and legal experts, but also subject matter experts without any formal qualifications who were recognised to be experts in their field. This included attendees who specialised in geolocation, chronolocation, or had other unusual but relevant skills or knowledge and who were recognised as experts within their peer group. The plan also included Yemeni journalists joining via video link, however, technological problems meant this was not possible. As such local contextual knowledge, which was recognised as being an important part of the investigative process, was limited for the hackathon itself.

The first day consisted of training in open source techniques, the methodology, as well as an introduction to the teams and subject matter. The remaining three days were spent investigating alleged airstrikes. Each team was assigned a number of incidents related to different categories derived from restrictions within International Humanitarian Law (IHL), the body of rules which sets out legal limits to the use of force in armed conflict:

1. Attacks on Objects Indispensable to the Survival of the Civilian Population
2. Attacks Causing Grave Civilian Harm
3. Attacks on Objects with Specific Protected Status
4. Destruction of Civilian Property
5. Attacks on Government Buildings Which Are Not Military Objectives

The attendees conducted open source investigation on incidents which appeared to fall under these categories, while subject matter experts provided expert opinion on matters such as geolocation or munition identification. Data points from these investigations, such as the time of day, the munition used and the target were collected and input into a data sheet.

The result was a huge collection of data gathered from social media, local news outlets, blogs and other online sources of information, comprising images, videos and texts regarding the incidents of interest. Some of this was structured, such as specific data points added to the data sheet, some was semi-structured in draft reports, and some was unstructured and listed as links within those draft reports. Initially the intent had been to complete full reports of all 100 strikes during this hackathon, however it was clear that this was not feasible due to time constraints and a lack of consistent reporting style across the large number of investigators.

A core purpose of the hackathon was the collection of open source information. Two of the major constraints on this were the need to archive open source information gathered as part of the exercise and the requirement to verify that each piece of information was actually related to the incident in question, which are discussed below. This not only formed the basis for assessing and linking the content, triangulating multiple pieces of data across each incident, but also allowed us to be confident about the individual data that was collected, and establish the provenance of the content, potentially an important question if the content was to be used in a legal context.

Archiving

One of the major issues which affects open source investigations is the availability and longevity of relevant information. Although content relating to airstrikes may be posted online, there is no guarantee that it will remain there. For a variety of reasons, content relating to extreme and violent events is often removed from platforms. The content may be identified by an automated algorithm as breaching the rules of the platform upon which it is hosted; it may be reported by bad-faith actors who disagree with the content; it may have hostile and fraudulent copyright claims made against it, or indeed the user themselves may choose to remove it from the platform (Banchik 2020).

To reduce the impact of this, the Yemen Project developed a protocol for preservation which covered both the content itself, and the method by which it was discovered. In order to preserve the content, we partnered with the Yemeni Archive, part of Mnemonic, a non-profit which specialises in the preservation of online content depicting conflict. The URL of any content investigators wished to preserve could be entered into a spreadsheet which would be subsequently preserved and digitally hashed to ensure its integrity.

However, the Yemeni Archive cannot itself track an investigator's journey. As such, a tool called "Hunchly" was used to do this, noting which websites were visited and what media was viewed, as well as creating a PDF of the page itself and a hash, a kind of digital fingerprint, of that visit (Roussev 2009).

The combination of Hunchly and archiving by the Yemeni Archive mitigated the majority of the issues resulting from the removal of content, as well as the chain of custody of this content, however gaps still remained. A minority of websites were not compatible with either Hunchly or the Yemeni Archive, resulting in either partial capture or no capture at all. Human error was also present, particularly at the hackathon stage, where some were not familiar with the software, or forgot to use it as directed. Considering human nature, as well as the multitude of differing formats websites take, these factors will continue to exist and must be recognised in order to mitigate it effectively.

Verification

The issue of erroneous images or videos being used to support news reports is one that is already well known. This kind of erroneous information most commonly tends to be content from a completely different incident being relabelled as being from the incident in question. Occasionally footage from films or video games may be passed off as real footage. It may even be possible for a malicious actor to try and create an entire video from scratch and claim it is from a particular incident, although such an attempt would be very complex and is certainly not common. Indeed, it has been argued that repeated claims of footage being manipulated may in fact be more damaging than the risk of these kinds of manipulated videos.

The vast majority of erroneous images or videos can be identified by the act of verification using geolocation and chronolocation: placing the content in time and space. If done correctly, these simple actions, although sometimes extremely complex in practice, will systematically remove the vast majority of erroneous content.

Additionally, contextual verification can help assess if a piece of content is genuine. For example, an airstrike in the middle of a city will result in a large number of people taking images and videos of the event. If a single Twitter user claimed that there was an airstrike, but no other users did, then it's unlikely to be a genuine claim. Similarly, inconsistencies in weapon effects, for instance, seen in content can act as an indicator that content may not be genuine.

For example, after the 2020 Beirut explosion, multiple videos appeared which purported to show missiles hitting the warehouse, causing the blast, but the size and nature of the explosion was inconsistent with any conventional weapon system. This immediately indicated that the videos had been altered to add the missiles. Once the footage was examined frame by frame it became clear the "missile" was in fact simply clip-art added to genuine footage.

This type of verification was carried out on every piece of content used in analysis by the Yemen Project. This not only added a layer of verification which systematically excluded erroneous information from analysis, but also produced accurate data about each incident. In multiple instances airstrikes had been reported as being in a slightly inaccurate location or time, even when investigated by teams on the ground. Through the verification process this could be identified and corrected, resulting in more accurate data.

Though there remains the possibility of actual digital manipulation of the footage itself, this is rare. It would require a digital forensic specialist to reach certainty about a single, isolated video, but steps can still be taken to reduce the risk. If multiple videos depict a single incident, then those videos can be cross referenced for inconsistencies. Additionally, if videos have been live streamed, or posted a short time after an incident, this either removes or significantly reduces the possibility of digital manipulation. Even with these mitigating steps, the opinion of a digital forensic specialist would be useful, especially with content that lacks corroboration.

Case Study – SLC Strike on the Office of the Presidency

On May 7th 2018 the SLC carried out a strike targeting the first and second ranking members of the Houthi leadership. Multiple news outlets reported that the Presidential Palace in Sana'a had been struck, yet satellite imagery revealed no new damage to the palace and no residents of Sana'a posted reports of that specific area being bombed.

However, multiple people reported on social media that another location had been struck. By examining photos and images of the airstrike posted by Sana'a residents and cross referencing multiple related images, it was possible to establish, using the process of geolocation, exactly which building was struck.

Example social media post claiming an airstrike had taken place.

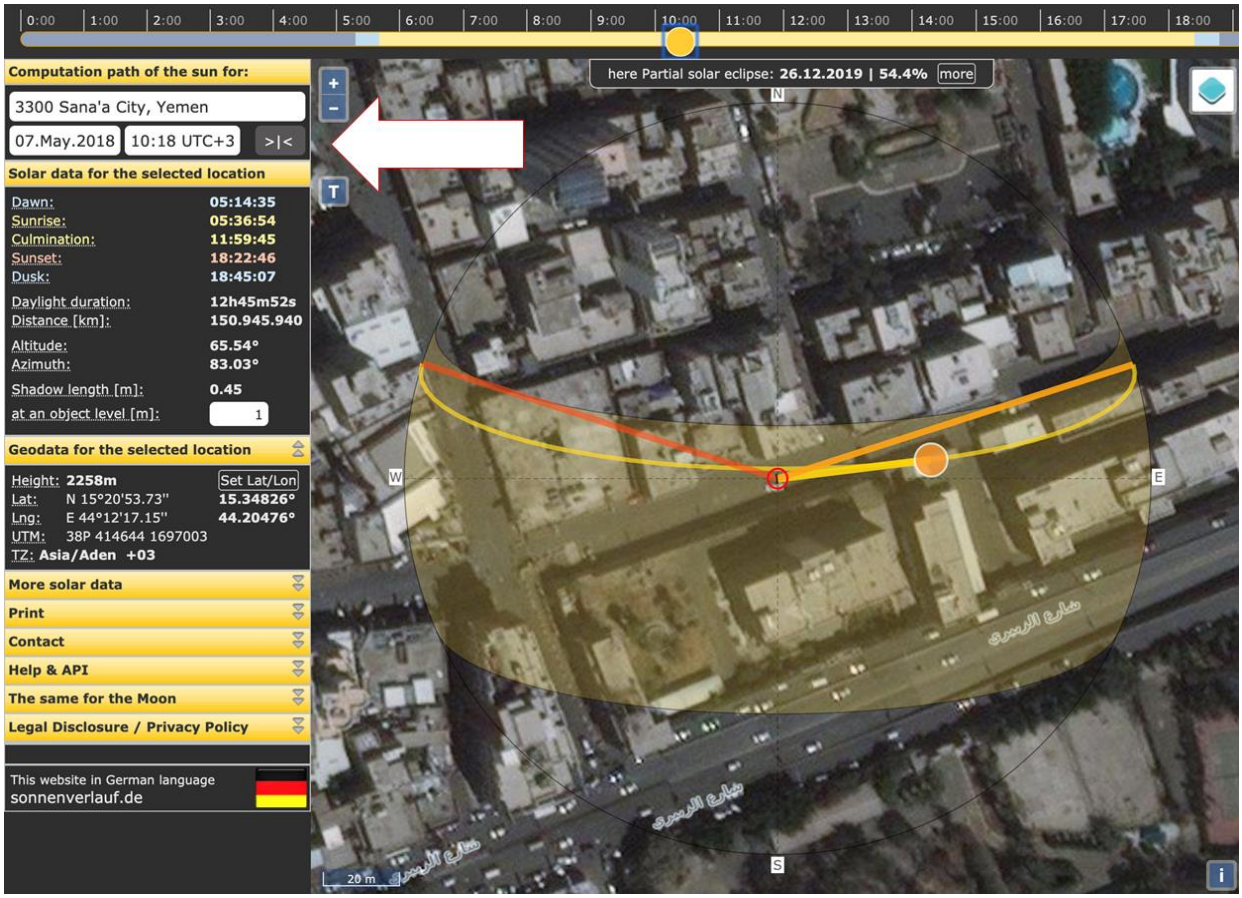
In the example below, we selected buildings visible in the photo which could be identified in satellite imagery. We could then align those buildings in the image and then on the satellite imagery to establish the precise line along which the airstrike had taken place.

One of the images of this strike that has been geolocated.

Once we had this line, we could then identify the exact location, confirming it by matching features seen in images and videos with satellite imagery.

Establishing a time can be achieved either by identifying the earliest social media post mentioning an airstrike at this location, or by examining indicators within images and videos, such as the length or angle of shadows. In the still below, we see a man standing on a road with his shadow clearly visible.

The proportional length of his shadow allows us to establish the angle of the sun in the sky and so therefore the time. This can be achieved quite easily simply by entering his height as one meter and his shadow length as 0.43 meters into SunCalc, an online tool that simulates shadows at particular times and dates.



An examination of a crowd-source mapping service, Wikimapia, showed that the building which had been bombed had been tagged as being associated with the office of the presidency or presidential administration since before the start of the conflict.

This example shows the power of these verification and investigative techniques. Despite widespread reporting that the "Presidential Palace" had been bombed, it was possible to establish using social media posts and mapping services that it was in fact an administrative building in a completely different location that had been bombed.

The content obtained as part of this investigation also served to highlight how much OSI can add to an analysis of International Humanitarian Law compliance. Finding the exact location on satellite imagery allowed legal assessors to appreciate how densely populated with civilian life this area was – and how obvious that fact would have been to the attacking party. User-generated content showing civilians attempting to rescue a young casualty as a second airstrike landed could help with a legal assessment of whether the attacker knew they were targeting a civilian area and could corroborate other evidence that civilians were harmed by the attack. The extent of the destruction seen in the online footage could aid a legal assessment of whether the SLC complied with the proportionality principle. Finally, analysis by investigators suggested that the purported military targets of the attack were later seen alive.

Phase 2: Report Writing

It quickly became evident that the length and level of detail in each draft report created as part of the hackathon varied significantly. As such a considerable amount of time was spent editing and further researching these reports in order to produce output which could then be published.

To create reports in a consistent format, a report template was created with headers for specific types of information, such as date, location, nature of weapon used and so on. This made the process of laying out the findings of an investigation easier, and made it easier to fill out the data sheet directly from a report. However, this template did mean the reports were written in a particular style which did not lend itself to storytelling.

In parallel with the creation of an incident template, a process flow was created which ensured that after the hackathon each report went through identical steps, ensuring that all content relevant to the report was archived, and that all relevant data was input into the data sheet in a uniform format.

During the hackathon many incidents had been investigated, creating reports with varying amounts of links to content, analysis, and verification steps. After the hackathon, a report would then be assigned to a first author, usually part-time or freelance staff, who would structure that information and perform additional analysis and verification as required. A second author, usually a staff member with more experience, would then examine the report, check its conclusions and edit it for style. At this point the report was then published on yemen.bellingcat.com. Finally, all the links would be archived, if they had not been already, and the data from the report was input into a data sheet. The progression of each report was tracked through a matrix on a spreadsheet as due to the complexity of tracking up to 20 reports being worked on at any one time.

Phase 3: Outputs

The primary output of this project was 21 in-depth open source reports on various airstrikes in Yemen. Although information about further incidents was collected as part of the hackathon, various factors, both operational and financial, meant that no further reports were published after January 2020. A major lesson from this project was that the collection, analysis and report writing of these kinds of events is a process which requires considerable investment of time. This becomes even more significant when the reports are being investigated and written to such a specific framework given the strict standards embedded within it.

Six of these reports formed the basis for a submission to the parliament of the United Kingdom to demonstrate the risk of British weapons being used in breaches of International Humanitarian Law (IHL). Despite this, and many other submissions, the UK government decided in July 2020 to continue the sale of weapons to Saudi Arabia. Although the government acknowledged that there were "credible incidents of concern" where breaches of IHL may have occurred, they ruled that these were "isolated incidents" and as such there was "not a clear risk" that UK arms might be used in breaches of IHL.

In February 2021 a mock hearing was organised by GLAN with Swansea University School of Law in which one of these reports was used as the basis for expert testimony. Undertaken with experienced QC's for both the prosecution and defence, with Judge Joanna Korner presiding over proceedings, the purpose of the hearing was to establish if this kind of open source information could be accepted as evidence by an English court.

The mock hearing was based on a fictionalised court case in which Bellingcat had submitted a video as evidence. Bellingcat investigator Nick Waters, acting as the independent and fictionalised witness "Frank Palmer," had produced a verification report on this video. The video and the expert report were scrutinised during this mock hearing to establish whether they could be accepted as evidence.

Judge Korner deliberated and ultimately decided to accept this kind of information as evidence, although she noted several caveats and aspects that stood against her decision. The most significant of these was the need for a digital forensic assessment of whether an image or video had been manipulated. Although an OSI analyst may be able to identify a poorly manipulated example, it is difficult to reach complete confidence without the opinion of a digital forensic investigator. She also accepted that "Frank Palmer" could indeed be an expert witness. This ruling, although made as part of a mock hearing, is potentially significant. Although it cannot be cited in court, it is a useful indicator of how an actual court in England of Wales may react to this kind of information being presented as evidence.

Conclusion

From its inception in 2018, the idea behind the Yemen Project was ambitious in both scope and intended outcome. Its aim was to plug the evidentiary holes that currently prevent justice and accountability areas of conflict around the world. In order to do this the project used the most up-to-date open source techniques available, collecting, verifying and preserving images and videos of SLC airstrikes across Yemen.

It was not, however, without its problems, and similar projects could certainly improve on it. The project was fantastically successful at bringing together a diverse group of people for the hackathon, producing a huge amount of data related to incidents in question. However, due to technical difficulties at the time, it was not possible to communicate with the Yemeni journalists, who should have been central to the event, and who were due to join. The scale of the information collected, and the detailed nature of the incident reports, meant that the ambitious scope of the project was never fully realised. The project did not achieve its target to publish 100 reports, managing only 21. It quickly became apparent that the level of resources required for these kinds of reports was significantly higher than was available at the time. Despite this, 21 high quality reports detailing strikes which caused gross civilian harm were published. Seven of these 21 reports formed the basis for a submission as evidence to the UK Parliament's Committees on Arms Export Controls. In summary, then, the novelty of this effort also meant that much of what was done was breaking new ground, sometimes effectively, sometimes less so.

In other respects, the project was certainly successful. It resulted in the production of a light-touch and easily replicable methodology for investigating the kinds of events that are most likely to feature on the open-source record. It resulted in the reports that were produced by this methodology being submitted and accepted as evidence by the Parliament of the United Kingdom, and it led to a mock hearing accepting the kind of information collected and verified in this way as evidence. Now that this information has been deployed in efforts to hold arms suppliers to account, only time will tell whether it will prove useful in a real prosecution.

These lessons have formed a basis of knowledge which will be applied to further iterations of this project. Ultimately, we hope that the successes and failures of this project so far can inform others who wish to carry out similar projects, and further the potential for the use of open source information as evidence to hold the powerful to account.

Notes

[1] The author uses the terms open source investigation (OSI) and open source intelligence (OSINT). These processes often make use of similar open source digital data are distinguishable mainly by output, where OSINT produces analyses that can be used for decision-making, particularly in security contexts, and OSI does not produce analyses for such purposes.

Bibliography

Banchik, A. V. 2020. "Disappearing Acts: Content Moderation and Emergent Practices to Preserve at-Risk Human Rights-Related Content." New Media & Society 23 (6): 1527-1544. doi: 10.1177/1461444820912724.

Reading, A. 2009. "Mobile Witnessing: Ethics and the Camera Phone in the 'War on Terror.'" Globalizations 6 (1): 61-76. doi: 10.1080/14747730802692435.

Roussev, V. 2009. "Hashing and Data Fingerprinting in Digital Forensics." IEEE Security & Privacy 7 (2): 49-55. doi: 10.1109/MSP.2009.40.