

# **OPEN SOURCE INVESTIGATIONS**

*Handbook*

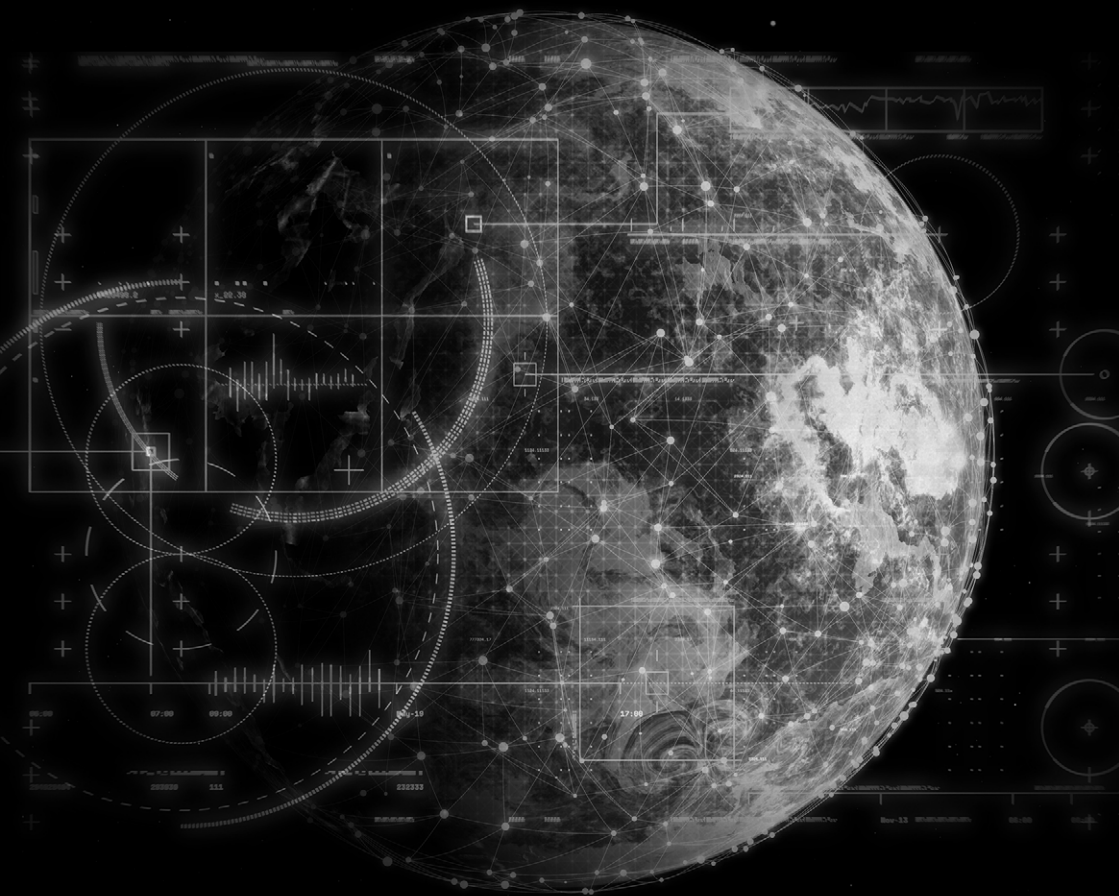
By:

**Sara Creta**

Edited by:

**Muhammad Al Khamaiseh**

**Nina Montagu-Smith**



# OPEN SOURCE INVESTIGATIONS

*Handbook*



ALJAZEERA  
MEDIA INSTITUTE

# Index

## Introduction

By Phil Rees,

Director of Investigative Journalism, Al Jazeera ..... 6

### *Chapter 1*

**What are Open Source Investigations?** ..... 10

### *Chapter 2*

**Planning and Carrying Out an Investigation** ..... 14

### *Chapter 3*

**Ethics and Safety** ..... 20

### *Chapter 4*

**Tracking Ships and Planes** ..... 24



*Chapter 5*

**How to Identify Weapons ..... 30**

*Chapter 6*

**Finding Out Who Owns  
a Corporation ..... 36**

*Chapter 7*

**Analysing Satellite Imagery ..... 40**

*Chapter 8*

**Tools and Networks ..... 49**

# Introduction

**Phil Rees**

Director of Investigative  
Journalism, Al Jazeera



Many of us know the scene in *All the President's Men*, Hollywood's interpretation of the Watergate scandal, when "Deep Throat" is standing in a car park basement in Washington DC. The man we now know to be the former Associate Director of the FBI, Mark Felt, was the secret informant who gave clues such as "follow the money" to the Washington Post journalist, Bob Woodward.


Finding the evidence that brought down US President Richard Nixon was a watershed event in investigative journalism and has rightly entered its folklore. To break investigative stories in the 1970s, you needed to develop sources. The skill sets needed for success were once described by the late Nick Tomalin as "ratlike cunning, a plausible manner, and a little literary ability". Tomalin was killed while reporting the Arab-Israeli war in 1973, a year after Woodward broke the Watergate story.

Investigative journalists were usually required to obtain confidential documents from people who did not hand them over without a great deal of persuasion. The investigator had to nurture sources over time. They'd met

in restaurants, coffee shops or late at night in bars. They persuaded whistle blowers to do the right thing; they gained their trust so that the identity of a source would not be revealed. Managing a source was a critical skill for an investigator. HUMINT - or human intelligence - was the cornerstone of investigative journalism and obtaining information that no one else had was essential for an exclusive.

A journalist usually carried only a notebook and a tape recorder. When I started in journalism, there were no mobile phones. There was little methodology to investigative journalism. Success depended on who you knew and how effectively you exploited them.

Then along came the computer and everything changed. Technology altered journalism, both its practice and its consumption.



The roots of OSINT lie in Computer Assisted Research. CAR began by exploring and analysing databases. In doing this we can discover patterns, trends and anomalies that may be useful in producing new information. The practical use of this methodology emerged with the Freedom of Information Act in the United States, which was introduced in the 1960s to open the workings of government to public scrutiny.

Philip Mayer, a pioneer of CAR, called it “precision journalism”. It was inspired by the methodology of social sciences where a journalist used evidence to prove his assertion.

A methodology was born that inspired a distinct storytelling style that distinguishes investigative from conventional journalism.

Conventional journalism is reactive and observational. It describes the world as it is seen,

usually after something has happened. Investigative journalism, by contrast, seeks to prove that some aspect of what the public thinks it knows about the world is wrong. Like a policeman or prosecutor, an investigator will discover a lead or obtain prima facie evidence that supports a hypothesis that “X is lying” or “X is corrupt”. The investigation will aim to prove this supposition. If it can’t, the investigation is dropped.

This investigative methodology, known as hypothesis-based narrative, replaced conventional character-based or travelogue storytelling. Evidence gathering became the glue that holds the narrative together.

Decades ago, Philip Meyer made the prophetic statement: “When information was scarce, most of our efforts were devoted to hunting and gathering. Now that information is abundant, processing is more important.”

In the last decade, open-source intelligence (OSINT) has emerged as a journalistic science, as the vast resource of data collected from social networks and internet-connected devices is mined for information beyond just databases.



The volume of data created, captured and consumed globally is projected to be around 200 billion gigabytes a year in 2025 (Up from 70 billion in 2020). Every minute on Facebook, around half a million comments are posted, and 150,000 photos are uploaded. More than four million hours of content is uploaded to YouTube every day. Add to that, 700 million tweets per day. Investigative journalism will increasingly rely on tapping these sources. We are not discovering truths that are strictly hidden from us - they are not confidential - but we are assembling information in a fashion that reveals new truths. We are unpicking the resources available online to tell the story behind the picture, the story that the metadata provides, or the story that shipping or flight data tells us about an event.

For filmmakers dealing with investigative content, there are new challenges. There will be more use of computer-generated imagery to tell the story and less use of video. There will be a need to harmonise different sources, such as vertical aspect ratio imagery, publicly generated and low-definition content with professional standards. Graphic designers, data scientists and filmmakers will need to work together in ways that presently rarely exist. The model of television production needs to adapt to a new method of storytelling.

With the amount of information in the public domain, investigative journalists of the future may be less con-

cerned with obtaining secret data than finding ways to make sense of public data, and tell stories based on that. More complex computer-based tools, such as data mining programmes, geographic information systems, demographic databases and so forth can be used to identify patterns, anomalies and discrepancies in data. Much of the new technology surrounding open source intelligence will involve machine learning, that is when a computer model is trained to analyse data much faster than a human being. In effect, you train a computer to do the hunting for you.

It means that investigative journalists no longer need to only learn how to write and turn on a tape recorder or camera. They will need to learn the tools of the internet. While computer scientists will write the programmes, journalists will need to understand the science of OSINT.

OSINT is not a substitute but a complement for HUMINT. For most investigations, journalists need to use human sources as well as data. Investigative journalists still need “rat-like cunning, a plausible manner, and a little literary ability”. But they also need to understand how to get value from the abundance of information on the Internet.

This handbook provides an invaluable guide to achieve this.



## Chapter 1

# WHAT ARE OPEN SOURCE INVESTIGATIONS?

An open source investigation (OSINT) uses intelligence gathering techniques and technologies including satellite imagery, social media posts and user-generated content to uncover the invisible. In recent years, open source investigations have become one of journalism's most valuable tools, largely due to its ability to tap into vast amounts of publicly available online information to reveal otherwise untold stories.

Collecting and analysing publicly available data and information from across the internet can include anything from analysing an IP address all the way through to interrogating public governmental records.

What is OSINT? Open source intelligence is the application of intelligence gathering techniques and technology to investigations that make use of open source data.

Security adjunct professor at Columbia University Mark M Lowenthal defines

OSINT as “any and all information that can be obtained from the overt collection: all media types, government reports, and other files, scientific research and reports, business information providers, the Internet, etc”.

The learning process of how to use open source tools is constantly evolving. This handbook provides core elements and tools for journalists who are interested in conducting open-source investigations. It introduces a framework and outlines ethical approaches, while examining case studies, to analyse the fundamentals of online search and research techniques for investigations.

Whether it involves using search engines to gather documentation, examines videos and satellite imagery to collect critical evidence, or evaluates data gathered from an online database, this handbook offers journalists the necessary skills to acquire and verify documentation.



From early conflict and environmental monitoring to high-profile investigations such as *Anatomy of a Killing*,<sup>1</sup> using advanced open source techniques has quickly developed to become a crucial practice for journalists in both long-form investigations and breaking news. Open source techniques involve researching, selecting, archiving and analysing information from publicly available sources.

An effective open source investigation begins by addressing these three questions: What do we need to know? Why do we need to know it? Who might have the information we need?

While various open source guidebooks available to investigative journalists differ on the exact process that should be followed in an open source investigation, they all agree on certain fundamentals.

First, you must have a clear strategy and framework in place for acquiring and using open source information. This involves identifying which investigation to pursue and how to transform your findings into an engaging story.

Second, you must identify a set of tools and techniques for collecting and processing open source informa-

tion without compromising the safety of your subject matter or those involved in investigating the story.

Third, you should develop the right strategies to validate your findings. Collaboration is an important consideration here.

Finally, as in many parts of the world information is heavily controlled, knowing how to preserve and archive data remains an important element that can affect accountability mechanisms.

---

<sup>1</sup> <https://www.youtube.com/watch?v=XbnLkc6r3yc>

## Benefits

- Wide array of information to collect
- No or low barriers to access
- Easy-to-locate publicly available data

## Risks

- Identity exposure
- Counterattacks from online adversaries
- Collection of misinformation

## Debunking Myths about OSINT

- Open Source investigation is just Googling
- Open Source investigation is only for cybersecurity professionals
- Open Source investigation is only for tech-savvy individuals or experts
- Open Source investigation is surveillance and violates privacy

## Tips for an open source investigator

### 1. YOUR SECURITY IS PARAMOUNT

– Don't forget to keep your identity hidden while searching.

**2. BE CAREFUL** – You really need a good eye for small details. A successful open source investigator has sharp observational skills to detect even the slightest bit of information that might contribute to the bigger picture.

**3. PERSEVERANCE** – To be a successful open source investigator, you need to stick out the seemingly never-ending process of compiling data and research.



## BOX #1

### Challenges for Open Source intelligence

#### CASE STUDY

#### Forced out - Measuring the scale of the conflict in South Sudan

In 2019, Al Jazeera's AJLabs data journalism team, in partnership with the Pulitzer Center, published an open source investigation to better understand the complexities and scale of displacement and land rights in South Sudan.

For this story, Carolyn Thompson and Kristen van Schie worked with land rights experts and statisticians to survey more than 35,000 random phone numbers across South Sudan in order to paint an accurate picture of displacement across the world's youngest nation which had descended into civil war.

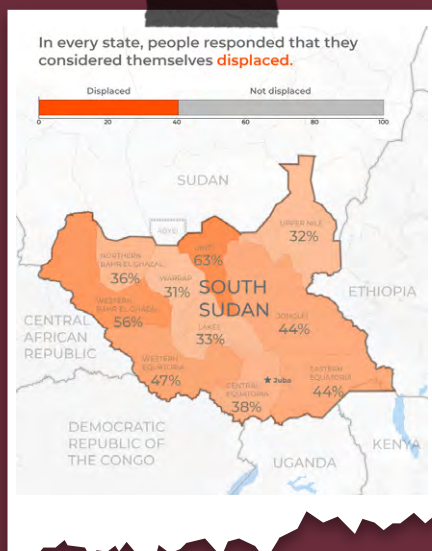
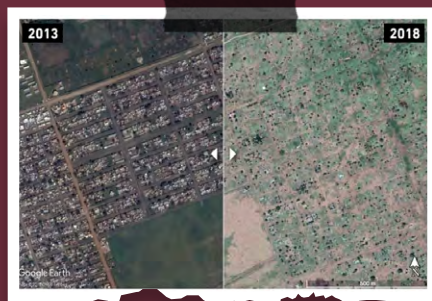
As of 2019, nearly 2.5 million refugees had fled to neighbouring countries.

As many journalists are denied access to or even barred from reporting within the country, Al Jazeera used a mo-

bile phone survey to gather information from those in places traditional journalism cannot reach. That data, which included questions on demographics, displacement, destruction, and plans to return home was then verified against other reporting tools, including satellite imagery, on-the-ground interviews, UN reports, public records, photos of the destruction and testimonies from internally displaced people and refugees.

The result was an interactive longform which included maps, videos, infographics and before-and-after sliders.

In 2020, the Investigative Reporters and Editors (IRE) awarded the story third place in the Philip Meyer Award for "an outstanding example of a determined group of reporters using social science methods to get to the root causes of a refugee crisis, even with severely limited press freedom, possible government interference, and a scared population."



## Chapter 2

# PLANNING AND CARRYING OUT AN INVESTIGATION

Journalists can follow the following four steps to start their open source investigations:

### Step one: Planning

Before diving into a story you should first determine if an investigation is possible or needed. To keep an investigative mindset it is important to always start with a series of questions. With your questions in mind, you can then formulate a clear strategy and choose the right tools to search for key information. When it comes to information gathering, journalists can decide either to make contact with the target during the investigation or to remain distant from the target and thus have a lower risk of being detected.

**Get started by answering the following questions:**

1: What has prompted the need for the investigation?

2: What are the key questions that need to be answered?

3: Which tools and platforms can help gather the required information?

### SEARCH TECHNIQUES

- Incorporate social media data to cross-reference your findings. Pay particular attention to who was the original source of this information, when this information was posted and where this information was posted from.
- Do a reverse image search using [TinEye](#) or [Google Images](#). A reverse image search allows you to upload an image and immediately see when and where this image was first used across the web.
- Use other platforms like [WeVerify](#), to fact-check videos and images online.

## Step two: Structure and secure information

Once you have a plan in place, you can now begin identifying the sources you will be using to collect and archive your data so that it remains secure. It is important not to lose sight of ethical, safety and legal considerations especially when dealing with personal data. Various data privacy laws including the General Data Protection Regulation (GDPR), the California Consumer Privacy Act (CCPA) and many others, exist to govern the collection, use and storage of personal data.

Always evaluate any potential data storage risks and keep evidence and documentation safe by using encrypted storage. Also, don't forget to take precautions to ensure your identity remains secure.

## ARCHIVING

Various groups including the investigative journalism group, Bellingcat; the Global Legal Action Network; and the Syrian Archive among others have created a standard process for archiving and investigating open-source evidence.

Collecting, preserving and building a body of evidence can serve as proof of power abuses and human rights violations.





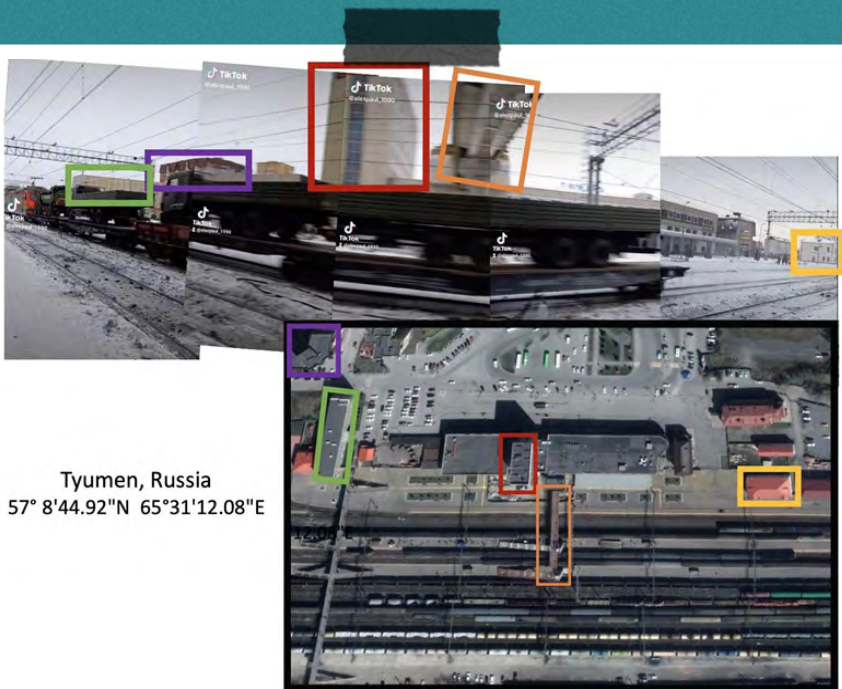
## Step three: Verifying your information

Raw information gathered must be analysed and processed before any useful or actionable conclusions can be drawn. This includes contacting people and verifying findings across multiple sources. Verification is an iterative process that involves three main phases:

*Verifying the source - Where did you get the information from?*

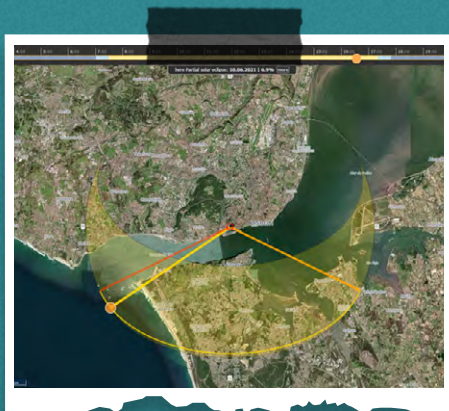
*Verifying the content - Is the information actually what it claims to be?*

*Verifying its relevance - Does this information fit into your investigation?*



## GEOLOCATION

Geolocation is the process of determining the geographic position of a particular event. This can be done by using tools such as Google Maps or Google Earth to match geographical features seen in the footage you are investigating. You can cross-reference stills from the footage to satellite imagery to confirm whether or not a video was indeed taken from a particular location. In some cases, it is possible to identify the approximate time the footage was captured by analysing the sunlight and shadows. Using [SunCalc](#) for example, it is possible to analyse the position of shadows and the sun at any given time and date, at any given location.



### Step four: Publishing your findings

Finally, journalists should publish their findings as well as show the process behind their investigations with the aim of ensuring transparency as well as building trust with the audience.

Ensure that your findings are presented across various digital platforms to ensure your story can have the widest available reach.

### BOX #2

#### Prepare, Don't Panic: Deepfakes and Synthetic Media

**Jacobo Castellanos, WITNESS**

Malicious deepfakes and synthetic media are - as yet - not widespread outside of non-consensual sexual imagery. However, with the rapid development of new technologies, it is expected that in the coming years these will be evermore photorealistic and pervasive, further blurring the lines between what is real and what is not.

For digital investigators and fact-checkers, the challenge of identifying synthetic media is growing. Already we are at a point where depending on our own eyes for detection is unreliable. There are some tips that can help spot them - for example, looking for visible glitches - but these are just current slips in the forgery process that will disappear over time (you can try to detect deepfakes yourself with this [MIT Media Lab test](#)).

The use of detection tools also provides no guarantees. If the technique used to generate the synthetic media is unknown, the results will tend to be unreliable as they would with low-resolution or compressed media generally found online. A recent experience of



a [suspected deepfake in Myanmar](#) shows that relying on publicly available detectors without further knowledge about how to interpret the results may lead to inaccurate assessments. What's more, recent attempts at developing deepfakes detection tools have not come up with models that were effective enough on known techniques or sufficiently applicable to new techniques.

Even if robust tools are developed, they may not be made available widely, particularly outside specific mainstream platforms and media companies. It is likely that media and civil society organisations in the Global South will be left out, and it is important to advocate [for mechanisms](#) that enable them to have greater access to detection facilities. WITNESS is arguing for increased equity in access to [detection tools](#), investment in the skills and capacity of global civil society and local newsrooms, and for the development of 'escalation mechanisms' that can provide timely analysis on critical suspected deepfakes.

As a way to tackle misinformation from AI-generated or manipulated media, there is a growing movement pointing towards the need for disclosure when synthetic media has been created or shared (see for example the [EU Code of Practice on Disinformation](#) or Partnership on AI's upcoming [Synthetic Media Code of Conduct](#)). 'Disclosure' can take the form of [labelling](#), or of other less visible techniques such as inserting forensic traces that are machine-readable, or metadata that contains information about its provenance.

Any one of these techniques could facilitate the process of identifying synthetic media, but without proper consideration they could





lead to further harm - for instance, labelling could lead to suppressing certain forms of free expression, particularly in art, parody or satire (see WITNESS's [Just Joking!](#) report for an analysis of these grey areas). Even well-intentioned efforts to provide tamper-evident provenance metadata for authentication, such as the work led by the [Content Authenticity Initiative](#) or the [Coalition for Content Provenance and Authenticity](#) (C2PA), could create risks of surveillance and exclusion for people who do not want to add extra data to their photos and videos, or cannot attribute the photos to themselves for fear of what governments and companies may do with this information (see the [WITNESS led Harms, Misuse and Abuse Assessment](#) of the C2PA).

Whether it be through detection tools, media literacy or disclosure mechanisms such as labelling, forensic traces or provenance-rich metadata, WITNESS is generally concerned that this work on 'solutions' does not adequately include the voices and needs of people harmed by existing problems of media manipulation, state violence, gender-based violence and misinformation/disinformation in the Global South and in marginalised communities in the Global North.

As these technologies evolve, the challenge for digital investigators and fact-checkers, as for journalists, human rights defenders and technology companies and synthetic media creators, will be to develop a better understanding of how to detect synthetic media and deepfakes in a way that is effective and accessible to those that need it most, while mindful of the unintended consequences, as well as potential misuses of these frameworks and tools.



## Chapter 3

# ETHICS AND SAFETY

Open Source Investigation carries important ethical concerns, as well as legal compliance. Information might be publicly available but personal data may be subject to data privacy regulations to varying degrees. Do not forget to consider the issues below when using open source investigative techniques:

**The origin and the intent of your sources:** Make sure that all your searches are targeted and that you are collecting only the information that is relevant to your investigation.

**Data is sensitive:** Make sure that you are collecting only public data and data that is freely available online. Make sure the data you collect is safely and securely stored so as to not breach data privacy rules.

**Use a VPN:** Do not forget to protect your identity. Using a Virtual Private Network or VPN can help mask your location and make your internet browsing more secure.

**Investigators can come into contact with a large amount of graphic footage. How to reduce the risk of secondary trauma?**

Secondary trauma refers to a range of trauma-related stress reactions and symptoms that may result from exposure to graphic details of another individual's traumatic experience.

As content from open source investigations is often very graphic, knowing yourself, and knowing what images affect you the most, is important to consider. Another factor in preventing secondary trauma is understanding your personal connection to the work you are investigating.

In 2020, a study<sup>2</sup> conducted at Berkeley, School of Law, USA identified six general practices as helping mitigate secondary trauma: processing graphic content, limiting exposure to graphic content, drawing boundaries between personal life and investigations, bringing positivity into investigations, learning from more experienced investigators and employing a combination of techniques.

---

<sup>2</sup> "Safer Viewing: A Study of Secondary Trauma Mitigation Techniques in Open Source Investigations" <https://www.hhrjournal.org/2020/05/safer-viewing-a-study-of-secondary-trauma-mitigation-techniques-in-open-source-investigations/>



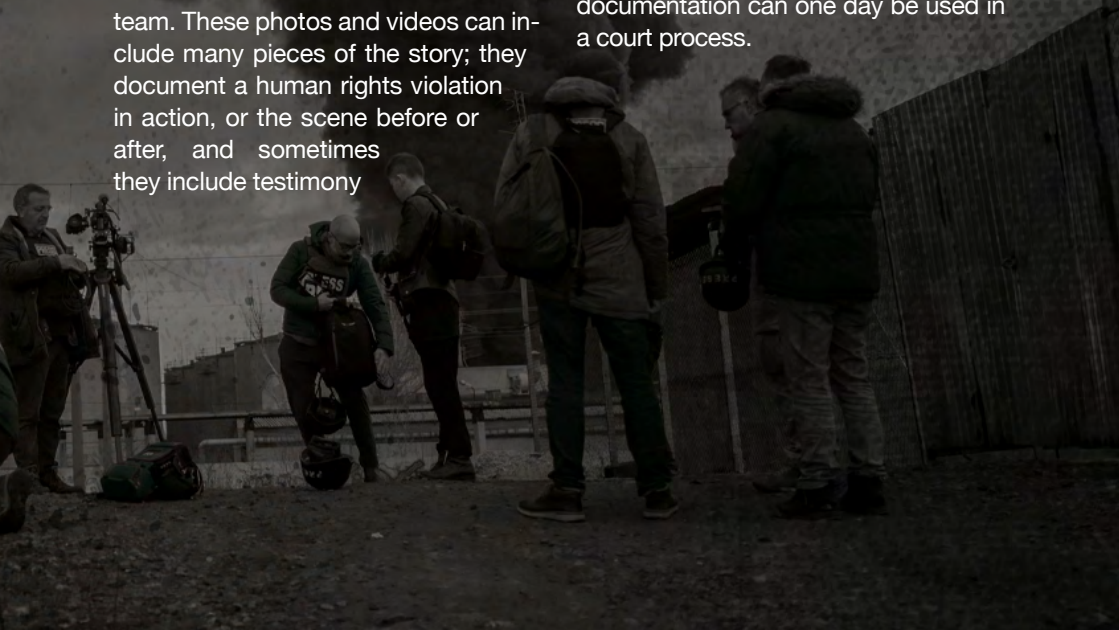
**BOX #3****Archiving for Accountability****Carolyn Thompson****THE SUDANESE ARCHIVE**

Since December 2018, the Sudanese Archive - a joint partnership run through Gisa and Mnemonic - has been gathering digital documentation, archiving it, and verifying it, with the goal of using it to contribute to investigations, court cases and other accountability mechanisms.

The project includes several components. First, a monitoring team collects material by scouring the Internet daily for evidence of human rights violations. This can include photos and videos filmed by documenters on the ground in Sudan and posted on Twitter, Facebook, Tiktok, and other open platforms. Also, materials are gathered from partners and contacts directly and shared with the Sudanese Archive team. These photos and videos can include many pieces of the story; they document a human rights violation in action, or the scene before or after, and sometimes they include testimony

from a victim of a crime. They can also include public statements or medical records, or other pieces of information that can help us understand what really happened.

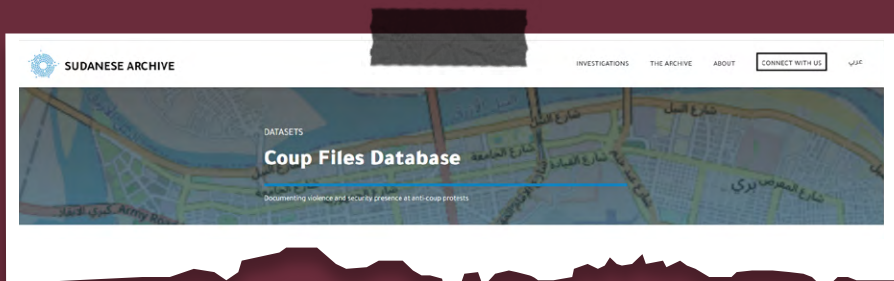
At first, we gathered links in a spreadsheet to track the protest violence, but we quickly realised many of those links would break when content was removed by the person who posted it for their safety, or by the platforms because of its graphic nature. All those important photos and videos proving what happened were getting lost. We tried downloading on our own computers, but there needed to be a centralised space to hold the content and keep it safe. That's why we began partnering with Mnemonic, which runs the Syrian and Yemeni Archives. Through Mnemonic's archiving process, all those pieces of digital documentation are permanently saved, and in a way that includes chain of custody components, such as time-stamping and hashing, to ensure the documentation can one day be used in a court process.



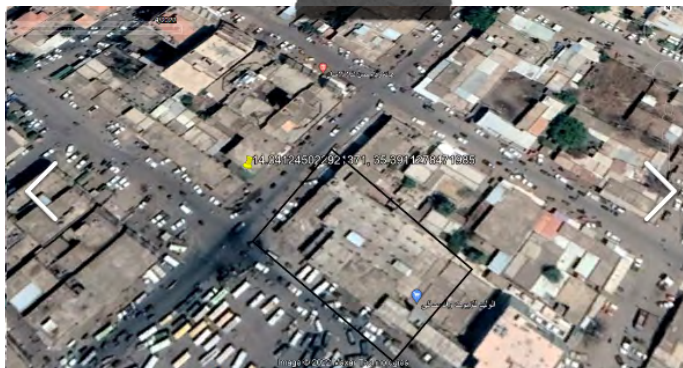


Once the material is archived, our investigative team sorts the content and identifies crucial pieces to be verified. The verification process involves determining the source of the video, the location where it was filmed, the time of day and date on which the incident happened, and any other relevant context.

Once many videos have been verified from the same event, our team can begin piecing together the truth of what occurred on that day. We use a standardised data tagging process to ensure every researcher is using the same tools and drawing the same conclusions, and we share those methods with our readers - an important piece of accountability is transparency in this process.



Our most recent investigation is a large dataset called the Coup Files, which aims to verify documentation of violent incidents at any protests that have occurred in opposition to the 2021 coup. In this dataset, our teams tag each investigated piece of documentation with identifiers that help us conclude who was the perpetrator of the violence. This includes tags focused on identifiable weapons, uniforms, vehicles and other indicators of those perpetrator groups. As well, we identify any protest characteristics that could help us prove there were indicators of excessive force or unlawful use of crowd control techniques.



That can be examples such as videos of tear gas canisters thrown directly into a dense crowd of people, or photos of live bullets at a protest involving the presence of students and children.

We publish incident reports focused on the protest days, grouping together violent incidents or the presence of security forces that we can confirm using this open source documentation. We also publish the data, set on a map, to help human rights advocates find the information they need - including by sorting for verified documentation of specific types of incidents or possible perpetrators.

- VIDEOS OF NUMEROUS HEADSHOTS



- VIDEOS OF PEOPLE BEATEN BY SECURITY FORCES



Already, our work has contributed to court cases within Sudan, and to international lawyers and sanctions teams. As well, numerous journalists have cited our investigations or worked with us to publish their own. While legal accountability processes are a significant part of our focus, we also prioritise the importance of ensuring we remain visible and consistent so that the perpetrators of these crimes know they are being watched, and those standing up for their rights know they are seen.



## Chapter 4

# Tracking Ships and Planes



Tracking the movement of ships and planes are increasingly valuable techniques in Open Source investigations. In the following chapter we present how these techniques can be used to investigate the movement of sanctioned goods, follow the travel paths of government officials and track illegal fishing or forced labour.

### Tracking Ships

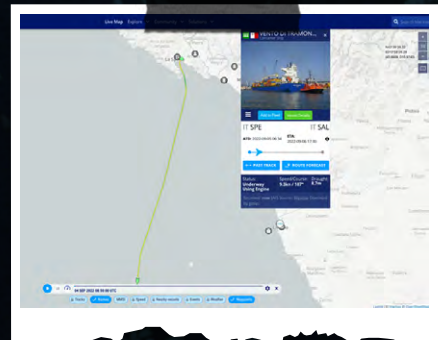
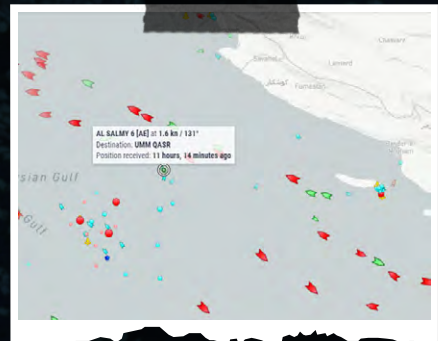
Most ships have an Automatic Identification System, or AIS, which transmits a vessel's position over time. By collecting historical AIS data, an investigative reporter can gain a better understanding of where a particular ship has been, measure how long that ship has been in a particular location and detect unusual travel behaviour.

### How to get started:

#### 1. Choose a ship-locating website.

Some go-to platforms for journalists looking for real-time shipping data include:

- [Marine-Traffic](#),
- [VesselFinder](#)
- [FleetMon](#)



## 2. Search for the name of the ship.

You can search for a vessel using its name. To ensure that the ship you're tracking is the correct one, compare the ship's unique IMO (International Maritime Organization) number and its MMSI (Maritime Mobile Service Identity) number.

*The **IMO number** consists of the three letters 'IMO' followed by a seven-digit number and is never re-assigned to another ship.*

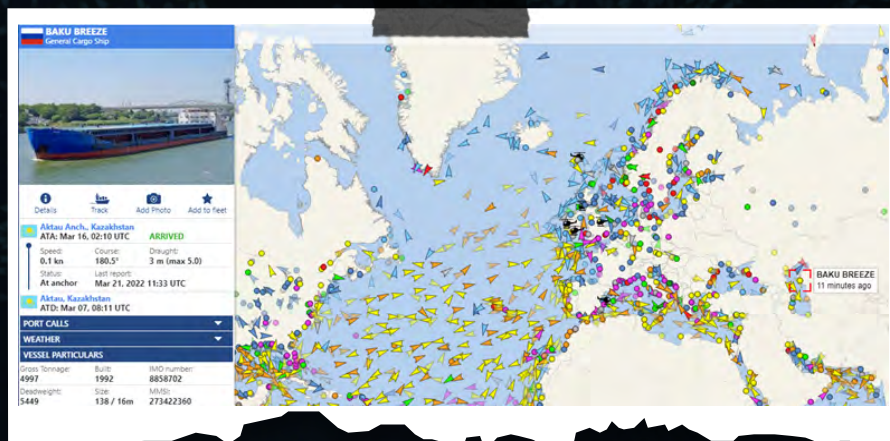
*The **MMSI number** is a unique nine-digit number for identifying a ship.*

## 3. Use the map to search for a ship in a specific location.

If you don't have the name of the particular ship you would like to track you can explore the map by zooming in or out of a particular location and then click on the ships around a particular port or shipping route.

## 4. Check with people on the ground or on the ship.

It is very helpful to reach out to crew members or other people working on the ground. You can try to find them using LinkedIn or other social media platforms.





## Other ship tracking tools

[Inmarsat Ships Directory](#) - Find the contact information of a vessel by searching for its name, number or call sign.

[Maritime Database](#) - Lists and details of shipping-related businesses and ports around the world.

[Global Fishing Watch interactive map](#) - Open-access online platform for visualisation and analysis of vessel-based human activity at sea.

## Don't forget

An alternative way to track ships is by using the Vessel Monitoring System (VMS), a satellite-based system that provides data to fisheries authorities on location. VMS is used to monitor the position, time, course and speed of fishing vessels. It is a key part of monitoring, control and surveillance programs at national and international levels.

Both AIS and VMS have limitations. If ships deliberately turn off their identification system, international, regional and national authorities, traffic management systems and surrounding ships are unable to identify or track vessels.

## Tracking Planes

Analysing an aircraft's flight pattern can help investigators track the movement of illicit commodities, scrutinise the movement of high-profile individuals and uncover the presence of surveillance aircraft.

*To get started with analysing a plane's movements, it is helpful to understand the following key terms:*

## Key terms

**Automatic Dependent Surveillance-Broadcast (ADS-B)** - A technology that broadcasts the position of an aircraft using satellite navigation or other sensors thus enabling open source investigators to track a plane's movements.

**The Call Sign** - The letters and numbers which identify an aircraft.

**Hex code** - A unique ICAO (International Civil Aviation Organization) 24-bit address, part of an aircraft's Certificate of Registration, used to identify an aircraft and broadcast by its Mode-S transponder. It allows for real-time and historical tracking.

**Registration number** - The number that appears on the tail of every plane. Looking at photos of an aircraft can help you determine the history of a plane. Two popular aviation image sites to search for visuals are [plane-spotters.net](#) and [jetphotos.com](#).

**Serial number** - Each aircraft is assigned a serial number by the manufacturer. This makes it useful for tracking a plane over time between owners, registrations and nations.

**ICAO airport code** - A four-character alpha-numeric code used to identify airports around the world.

**Aircraft Ownership:** Identifying the owner of an aircraft is theoretically possible, but practically difficult because most countries do not make their registries public. [AeroTransport](#), [CH Aviation](#), are good places to start looking, also see [Airframes](#), [RZJets](#) and [spotters](#).

The Isle of Man is one popular aircraft registration jurisdiction, providing a way to escape EU taxes, according to a report by the International Consortium of Investigative Journalists.<sup>3</sup> The ICAO code for the Isle of Man Airport is: EGNS

---

<sup>2</sup> <https://www.icij.org/investigations/paradise-papers/offshore-gurus-help-rich-avoid-taxes-jets-yachts/>



## Aircraft Tracking Websites

As long as the transponder of an aircraft is on, you should be able to use the following flight tracking services to track its movement:

### [ADS-B Exchange](#)

The world's largest source of unfiltered flight data. Does not filter out information about US aircraft that have requested anonymity.

### [FlightAware](#)

Allows guest users free tracking options, including alerts on planes of interest.

### [Flightradar24](#)

A commercial flight tracking service that permits free tracking of flights.

### [RadarBox24](#)

A flight tracker with live maps and search function.

### [Freedar](#)

A flight tracker that includes military aircraft. It also has monitoring of air traffic control audio.

### [OpenSky Network](#)

A non-profit association based in Switzerland that provides open access to flight tracking control data.

Anyone can deploy an ADS-B ground receiver that will triangulate satellite and aircraft transponder transmissions. If you are interested in helping increase ADS-B coverage, you can request a receiver from [flightradar24](#).

The Swedish aircraft tracking service will send you the ADS-B receiver sets (including receiver, antenna, and cables) free of charge that require a 10 to 20-minute setup, and once turned on will widen the coverage of ADS-B in your area.

>> Check the latest [GIJN guide to track aircraft](#) around the world

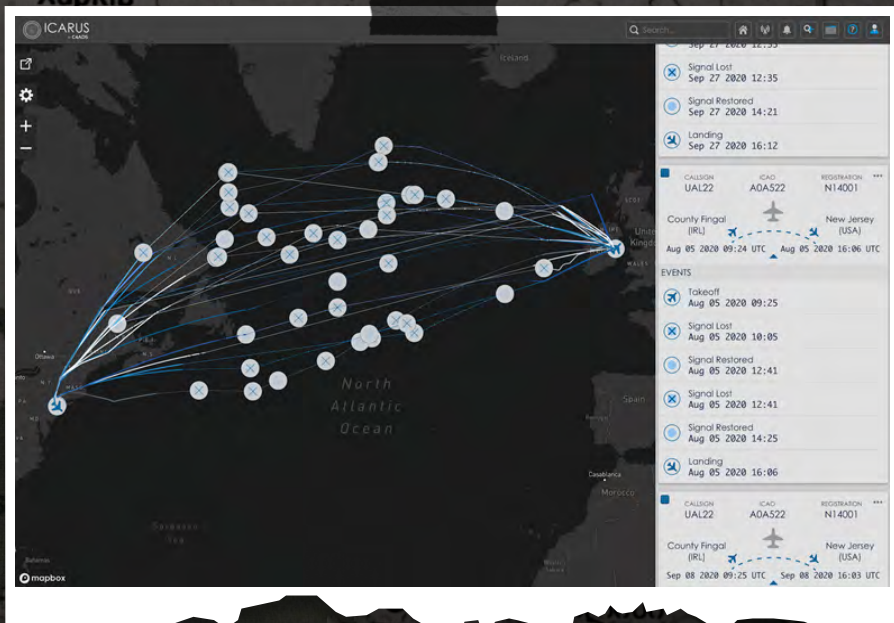


## Icarus Flights

Washington-based non-profit C4ADS has released Icarus Flights, a robust new system designed to help journalists trying to monitor illicit activity over a geographic area or during a certain period of time.

The Icarus tool kit includes transponder data, aircraft ownership records, and analytical tools. It provides location-based searches for investigators who want to study or document which aircraft have flown in a given area.

<https://icarus.flights/>



## Chapter 5

# How to Identify Weapons

Since the conflict in Yemen began in 2015, it has become harder for international rights organisations, UN bodies and journalists to document violations committed by all parties to the conflict.

Investigators have to work very hard to identify and verify the details of possible unlawful attacks, mainly using intelligence gathering techniques and technologies. One of these techniques involves analysing photos and videos to verify the types of weapons being used.

Investigators can study the shape of a crater left behind after a missile strike, watch footage of air raids to classify the types of missiles used, or analyse weapons trade data to understand ownerships of these munitions.

This documentation can provide essential evidence which can later be used to hold perpetrators of violations accountable. Documenting and archiving these findings paves the road to justice through using them in legal procedures.

### Here are a few steps to help you identify weapons

#### 1. Determine the weapon's class.

Broadly speaking, there are three main classes of weapons: small arms, light weapons, and heavy weapons.

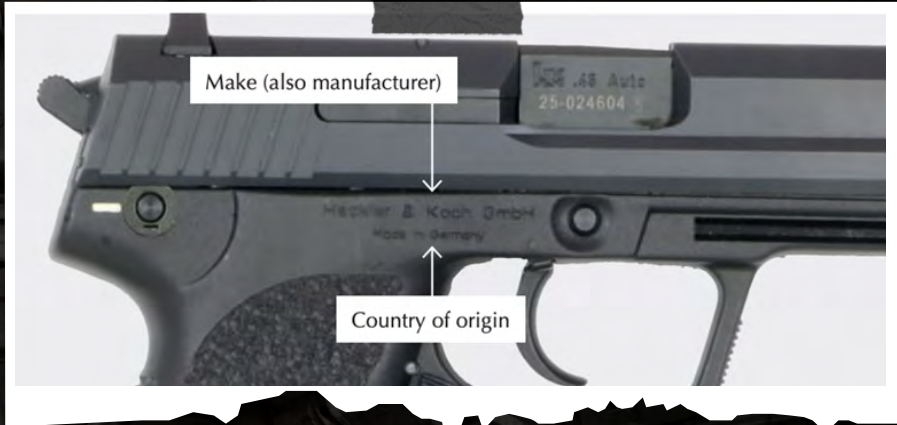
- *Small arms include pistols, rifles, light machine guns and other weapons that can be carried and operated by one person.*
- *Light weapons include larger machine guns, rocket-propelled grenades (RPGs), man-portable air-defence systems (MANPADS), mortars and other weapons that require a small crew to operate.*
- *Heavy weapons systems include tanks, helicopters, fighter planes, submarines and warships.*





## 2. Determine a weapon's make, manufacturer and country markings.

Weapons usually have markings that denote the make and/or manufacturer, country of origin, and, less frequently, the production facility and/or storage arsenal. This information can usually be found on the weapon itself as is highlighted in the image below.



## 3. Model and calibre designations.

The model refers to the make and design, while calibre refers to the diameter of the bullet - usually measured in millimetres or inches. One of the most common sizes is the 9mm calibre which is used in various handguns.

## 4. Find the serial number.

Serial numbers are useful for tracing weapons when they are recorded in documentation pertaining to manufacture, import, export, licensing, or in-country transfer.



**To track a weapon's origins, several databases are available for you to use:**

#### **Weapons Identification Database**

The Weapons Identification Database includes several small arms and light weapons. The database about these weapons includes information about the producer, type, calibre as well as photos to help you visually match a weapon.

#### **Arms Embargoes Database**

The Arms Embargoes Database aggregates data about all multilateral arms embargoes that have been adopted by the EU or the UN, or a group of nations.

#### **Arms Transfers Database**

The Arms Transfers Database traces suppliers and recipients of arms. It enables individual comparisons between countries with an option to select the range of years to cover and the weapon systems to include.

#### **Military Expenditure Database**

The Military Expenditure Database contains data about the military spending of 171 countries since 1988 as well as of NATO member states from 1949 or from their time of accession.

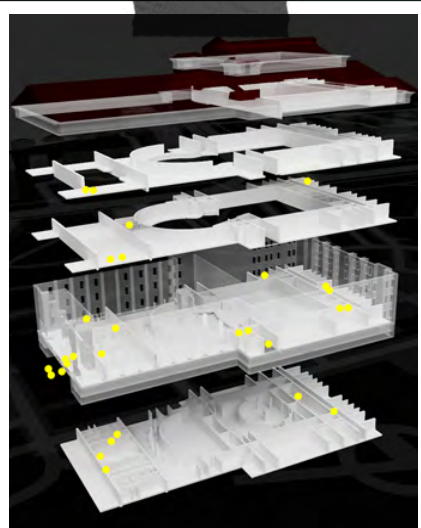


## BOX #4

**Detailed Investigation Into Russian Air Strikes on the Mariupol Theatre, Ukraine**

**Tom James, Sophie Dyer,  
Stella Cooper, Crisis Evidence  
Lab, Amnesty International**

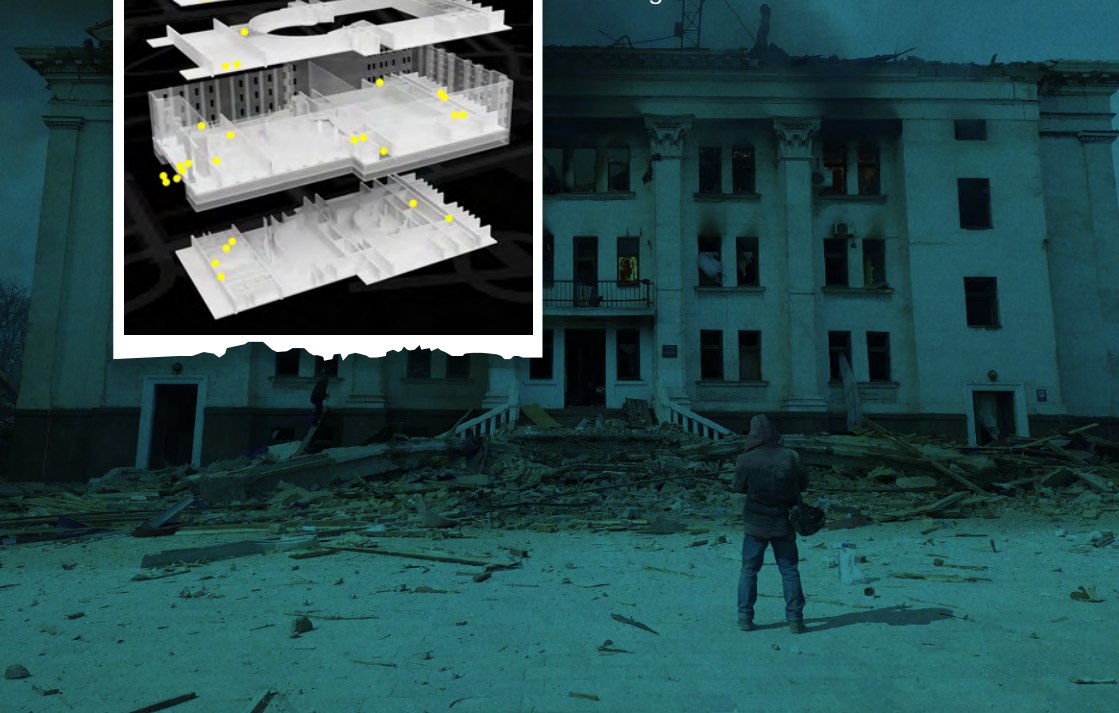
In the Crisis Evidence Lab at Amnesty International, we use digital 3D models to both generate new findings (evidentiary) and to communicate existing findings (demonstrative).



It is worth comparing them to their analogue counterparts, that is, physical architectural models. Architectural models can be quick sketches, conceived to check the viability of an idea or richly detailed presentations that take months to produce. They can be made to different scales: from a tiny detail study to a city-scale model that fills an exhibition room.

Digital models can be all these things too, either separately or all at once. They can contain many different layers of data that can be turned on and off or overlaid. They can be zoomed almost infinitely, enabling 3D and 2D elements to be viewed together at different scales.

In the work of the Crisis Evidence Lab, digital models become operational in multiple ways at different stages of an investigation.



They can be containers for organising evidence - including photos, videos, satellite imagery, drawings, witness testimony - in time and space. Through this process we gain new insights into the research materials, revealing details not readily apparent when analysing the media independently. Lastly, we use models as presentation devices: to publish often complex findings in print, video and interactive visualisations.

Our most recent model was built to accompany an [in-depth report](#) on the March 16, 2022 attack on the Donetsk Academic Regional Drama Theatre in Mariupol, Ukraine. Amnesty International interviewed more than 50 survivors and witnesses, and collected extensive digital evidence. The investigation concluded that Russian military forces likely deliberately targeted the theatre with air strikes, despite knowing hundreds of civilians were sheltering there on 16 March, making the attack a clear war crime.

A key secondary finding is that the damage to the theatre, although devastating, was relatively localised, and the death toll was probably lower than previous counts published by local authorities and international media. Witnesses described how the blasts caused severe damage and multiple fatalities in certain areas and left other areas protected from the explosion.

To better understand the impact of this attack, we built a 3D model which we overlaid with witness locations and testimonies about areas of greatest damage, photos and videos, as well

as a mathematical model of the blast wave. By doing this as a volumetric 3D model rather than 2D floor plans, we can visualise the layers in place, and explore the damage from multiple angles with more clarity.

We started by reconstructing in 3D the overall dimensions of the theatre using photos from before the attack. By teaching the 3D software which are the main features defining the perspective of the photo, it is possible to 'solve' the camera, and establish a good estimate for camera position and focal length. In doing so, we can reference the image to model exterior details in the correct proportions. The overall scale can then be corrected using satellite images, or referencing objects of known sizes in the photos, such as cars or street furniture.

Initially we only had access to hand-drawn plans of the interior of the theatre, produced by a witness who made measured drawings of the layout. Subsequently we were given access to full CAD (Computer Aided Design) plans, produced when the building was refurbished in 2018, which made all interior details clear. At this point it was straightforward to trace the plans and extrude the walls of each floor. We stacked them in 3D using clues from interior and exterior photos, as well as stair details, to estimate floor heights.

We used 3D software to geolocate the building accurately on satellite imagery, and then imported map data to generate the surrounding buildings, roads, and open spaces in that area of Mariupol.



This city district-level context was important for establishing the relative geographic isolation of the theatre to other buildings, when considering the intentionality of its targeting. It is often common to treat 3D models of buildings as ‘objects’, floating in space, and viewed from afar. By anchoring the building in its urban context, we establish its scale and relationship to a living city, as well as the significant portion of it that was underground.

We took the decision not to model any of the damage to the building. This can work for certain investigative or diagrammatic purposes, but here it was far better to use the model as the backdrop to overlay photos of the damage, which are far more revealing. In this way, the model becomes a representation of the building prior to the attack, and a container for the media recorded afterwards.

To locate images in the model, we used a similar method to the one described earlier, using perspective lines to calculate the camera position and focal length - or reading the focal length from the metadata if available - and then estimating the position manually. Moving between photos and videos in the model is a powerful way to map the spatial or geographic dimension of an investigation. Moving from image to image while using the model as a backdrop helps the viewer to visualise a sequence of events in time and space.

When it came time to present the model, rather than building an interactive platform we opted for a scripted ex-

plainer-style video, with a voiceover from a Ukrainian human rights activist that guides the viewer through events in lay terms, targeted at a general audience. This allowed us to incorporate not only the 3D model and overlaid media showing structural damage to the theatre, but also other elements including open- and closed-source footage, satellite imagery and an animated visual timeline of events. Additional outputs were a series of short clips and stills that were included in the report, shared across the Amnesty movement, and published on social media.

Building a 3D digital model gives us great flexibility in producing drawings and animated segments that evolve alongside the script as it is refined. Segments can be exported in low resolution to check timing and script beats, iterated upon and then rendered at full quality. By cutting into or ‘exploding’ parts of this model we were able to show how the ordnance penetrated the main theatre space with minimal resistance except for the roof, before exploding at stage level.

In this investigation, the 3D model served to synthesise the testimonies and other evidence into an overall view that convincingly corroborated the areas of damage against those with known casualties, and served as a narrative and illustrative device in support of the detailed report. Incorporating it into a six-minute video greatly enhanced the user experience and made the findings of the investigation more accessible to a far wider audience.

## Chapter 6

# Find Out Who Owns a Corporation

▼

If you would like to investigate the world's largest companies and reveal who owns offshore companies and trusts, free databases are your starting point. There are other ways to research companies; you can find official and court records, and search on subscription databases or corporate websites.

Whatever you are investigating on global money-laundering cases or bribery investigations, you can use [OpenCorporates](#) to try to identify who is who and who is transacting with whom. The database can provide the company's incorporation date, its registered addresses, and the names of directors and officers. You can search connections between companies, or work out which companies are run by the same CEO and even do more specific searches focusing on particular countries. Similarly, journalists trying to 'follow the money' across borders can use the [Investigative Dashboard](#), created by the [Organised Crime and Corruption Reporting Project \(OCCRP\)](#), to allow access to hundreds of databases that detail company records and online and offline court records from nearly every country in the world.



If you are trying to expose organised crime and corruption around the world, the [Offshore Leaks Database](#), developed by the International Consortium of Investigative Journalists (ICIJ), can help you to find information and documents on persons of interest and their business connections. The database contains leaked documents about nearly 785,000 offshore companies and trusts.

If you are interested in covering oil, gas and mining and you would like to discover the connection between the companies that own and operate oil rigs, and how they are incorporated as companies in, or working through, maritime tax havens; check the portal [Double Offshore](#) developed by Code for Africa. The same organisation developed the project the [Miners of Mozambique](#), to discover the individuals behind the mining industry in Mozambique and their connections.

#### MORE:

- [ResourceContracts](#): A portal that houses over a thousand mining and oil contracts.
- [Resourceprojects.org](#): A repository of extractives projects

**BOX #5****How did a complex network of shell companies trade Syrian phosphates despite sanctions?****Bashar Deeb**

investigator at Lighthouse Reports

Whether it's a warzone in the Middle East, Ukraine or Africa, or borders between Greece and Turkey, it's often very difficult to send journalists to inquire about things in such places. During a joint investigation between Lighthouse Reports, the Organised Crime and Corruption Reporting Project (OCCRP), and Syrian Investigative Reporting for Accountability Journalism (SIR-AJ), we were looking at the exports of [Syrian phosphates](#) to see where they were ending up. The exports started off from the Mediterranean port of Tartous, which is controlled by Russia. This port is not a place where journalists can visit and speak to people, so OSINT was critical to this work.

In theory, finding the ships that are loading the phosphates in the port was the main challenge to figure out where the phosphate was going because then we could track these ships on commercial AIS services to see their final destinations. During the investigation, my colleagues noticed that the port's FB page was regularly publishing a list of the working ships on a daily basis. The list also contained the type of cargo for each ship and the piers where each of them were docked.

But this didn't last long, the page deleted most of their recent lists and stopped publishing new ones after June 2020. But using google dorking techniques - a search string that uses advanced search queries to find information that are not easily available - we managed to find other pages which had copy-pasted these lists and reconstructed the timeline of the working ships in Tartous port. This allowed us to track the ones carrying phosphates to their European destinations. Of course, we did extra traditional verification work in some cases by asking for landing bills for these ships to make sure our analysis was correct, in some other cases we obtained custom records that verified to us that these ships were indeed moving phosphates.

But we still needed to find out where the new shipments were going. Here, the challenge was to figure out a way to identify which ships were moving phosphates after June 2020. Based on our observations of these lists but also on reading old online articles about the structure of the port, we noticed that all the phosphate-carrying vessels would dock in berths 18-19. This pier was built specifically to handle phosphates, with a crane operating between it and a dedicated phosphate storage area. We used satellite imagery to verify. Having this information allowed us to look through photos taken by port workers or photos from official visits to the port and identify a few ships that were docked in this pier.



Also by looking at visual material posted by workers at the port, we found a live Facebook video taken from inside the phosphate section which showed a truck driver emptying phosphate from a bigger truck. The truck had the logo of a security company which was born out of a militia. This company was reported as being the one providing security to the phosphates convoy coming from the mines to the port, but nevertheless we added visual evidence that confirmed this reporting.

Two examples of that are 1/ a series of photos posted from the port after the visit by the Syrian energy minister to the port in May 2021, where we identified two ships in the Phosphates pier; 2/ a selfie taken by a port worker which showed a ship in the background docked on the same pier. We tracked these ships and figured out that they ended up in Romania and Ukraine.





## Chapter 7

# Analysing Satellite Imagery



When you have a story, but still need to tie up loose ends to answer where or when a particular event occurred, analysing satellite imagery can point you in the right direction.

### So how do you get started?

Analysing satellite imagery can be useful in providing geographical context, reconstructing events, or even verifying if a particular event even happened at all.

The use of satellite imagery has become an indispensable tool for investigative journalists to report on conflicts, environmental destruction, developments in military infrastructure and natural disasters. Satellite imagery has also become a compelling centrepiece for visual storytelling, and a window into remote or restricted locations. Investigative journalists can use satellite imagery to make visible what governments or institutions want hidden out of sight.

## SATELLITE IMAGERY PROVIDERS

Over the past few years, several free and subscription-based earth imaging companies have emerged allowing anyone to access high-resolution satellite imagery from all over the world. Some of these services include:

### Free services

- [Google Earth](#)
- [NASA's Worldview](#)
- [The European Space Agency](#)
- [World Imagery Wayback Tool](#)
- [Zoom Earth](#)

### Subscription services

- [Maxar Technologies](#)
- [Planet Labs](#)
- [Sentinel Hub](#)
- [SI Imaging Services](#)
- [Spaceknow](#)

However, just having access to these services is not always enough. For satellite image analysis to be effective in your investigation you will need to ensure that the recency of the images as well as the satellite image resolution are adequate to match your needs.

Companies like Maxar Technologies and Planet Labs will often publish very high resolution, up-to-date satellite images on image wire services such as AP, AFP and Reuters. These companies also often provide image archives of big stories to the media.

Once you have identified your image provider, the next step is to make sense of the satellite imagery. Examining images can complement other research and provide corroborating evidence.

***To unlock the rich information in a satellite image, you should:***

1. Determine the image scale to help you determine the size of the area you are analysing
2. Look for patterns, shapes and geographical textures including natural and man-made landmarks.
3. Find where north is facing to help you determine the direction of movement of subjects of interest and/or shadows.

4. Analyse the direction of the shadows and colour of the terrain to help you determine the date and time a particular image was captured.

5. Consider your prior knowledge of a location to see if anything stands out in the environment

**Be careful**

With more people using satellite images you will also get more people trying to misuse satellite images to better align with their agenda.

***Here are some things you should do when you have doubts about the validity of a satellite image:***

- Verify that the image matches the original source of the satellite imaging provider
- Compare the satellite image with other sources
- Try to verify when an image was captured by using tools like [suncalc.org](https://suncalc.org/) to analyse the position of the sun and shadows
- Consult a remote sensing expert



## CASE STUDIES USING SATELLITE IMAGERY

*One of the most common uses of satellite imagery is to compare before and after images in a specific location.*

### BEFORE AND AFTER

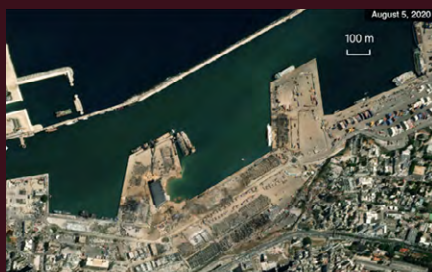
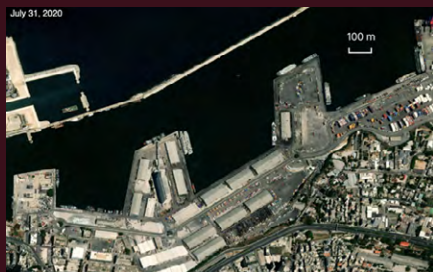
On August 4, 2020, a [massive explosion](#) in the Port of Beirut ripped through Lebanon's capital, killing 218 people, injuring 7,000 and leaving 300,000 displaced.

The blast, which is considered one of the biggest non-nuclear explosions to have been recorded, damaged 77,000 apartments and caused an estimated \$3.8-4.6bn in material damage.





Satellite images captured on August 5 highlighted the extent of the damage to the surrounding area.



## LEBANON

### How big was the Beirut explosion?

On August 4, 2020 a massive explosion in the Port of Beirut ripped through Lebanon's capital, killing 218 people, injuring 7,000 and leaving 300,000 displaced.

218 7,000 300,000  
Killed Injured Displaced

77,000  
Apartments damaged

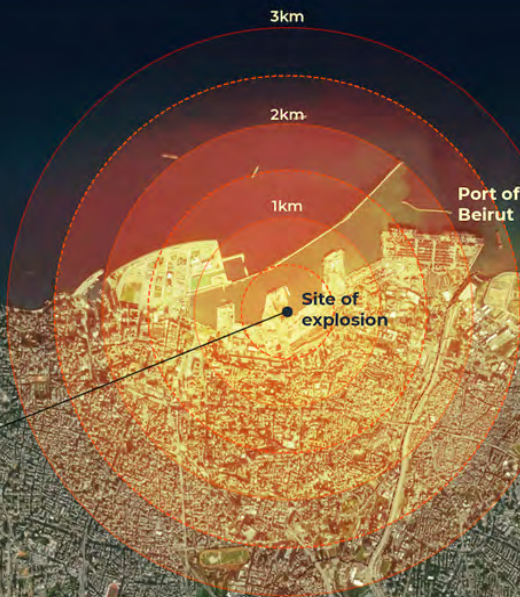
\$3.8bn-\$4.6bn  
Cost of material damage



Before the explosion



After the explosion



Source: HRW

## MAPPING ENVIRONMENTAL IMPACT

Journalists are increasingly using satellite imagery to conduct disaster damage assessment or carry out environmental monitoring. Free Satellite Data on Africa, a new tool by Digital Earth Africa, offers [free satellite data](#) on water resources and flood risks, agriculture and food security, urbanisation and more. [Smoke Screen project](#) used analysis of satellite data to prove deforestation by large private landowners in the Amazon.



## CASE STUDIES

### • Xinjiang detention camps

In 2018, [Shawn Zhang](#), a Chinese law student in Canada, began scouring Google Earth for evidence of detentions in Xinjiang, an official autonomous region in China.

Since then, several organisations including the Australian Strategic Policy Institute (ASPI) have used satellite imagery, witness accounts, media reports and official construction tender documents to classify the detention facilities into four tiers depending on the existence of security features such as high perimeter walls, watchtowers and internal fencing.

ASPI says they have identified more than 380 “suspected detention facilities” in the region, where the United Nations says more than one million Uighurs and other mostly Muslim Turkic-speaking residents have been held in recent years.

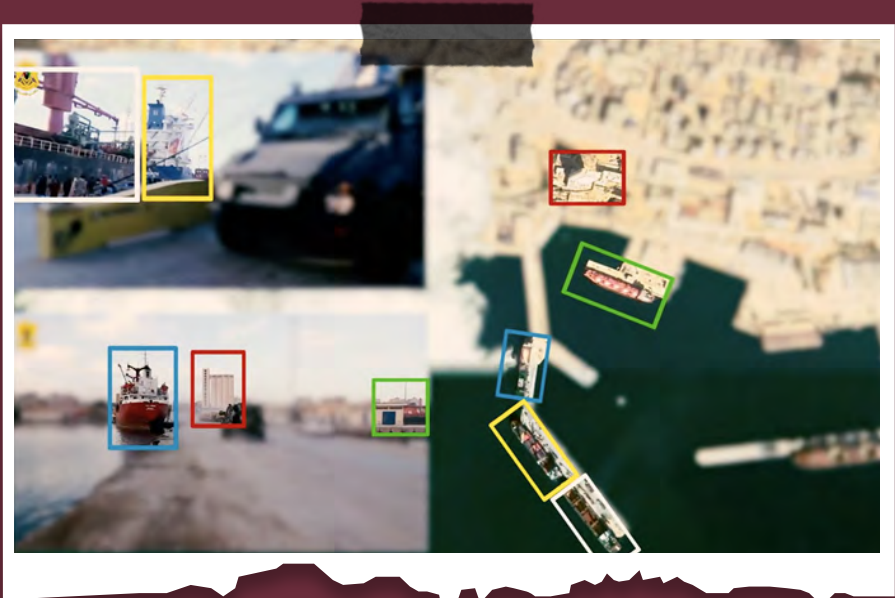




## • Weapons sales to Libya

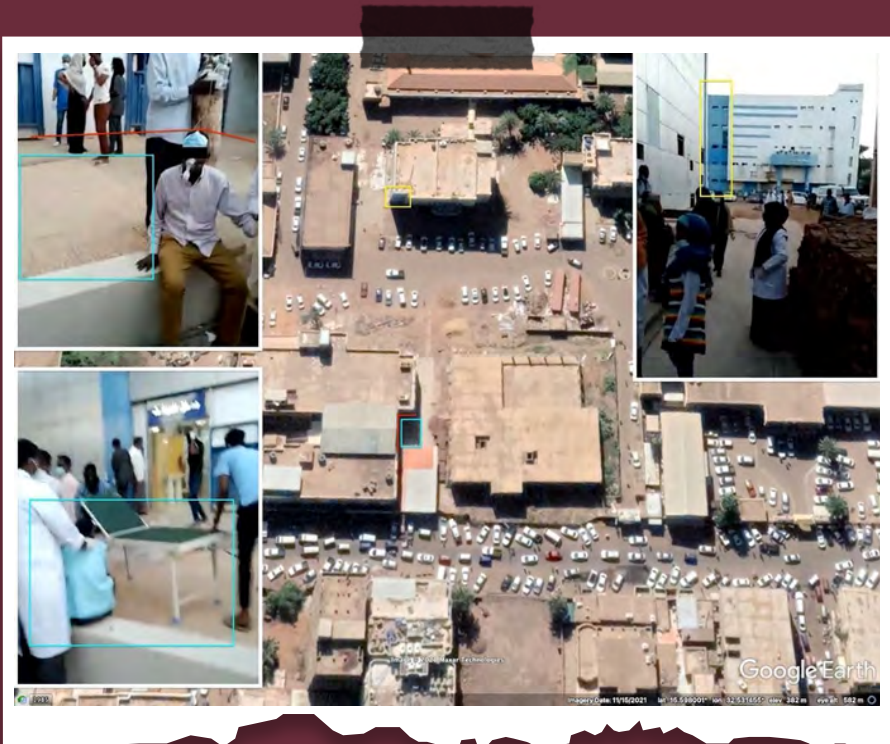
Using tracking tools and open source data, investigative journalist Mahmoud Al-Waqi and [team](#), revealed how weapons and armoured vehicles manufactured in the United Arab Emirates and Turkey were sold to warring factions in Libya, in violation of the UN Arms embargo on Libya.

This case study helped the team to claim that the United Arab Emirates and Turkey have been violating the UN arm embargo. As it was impossible to visit sites in Libya, satellite imagery helped verifying claims.



## • Attacks on Hospitals and Medical Staff in Sudan

An investigation by Benjamin Strick used open source [investigative](#) techniques to geolocate, chronolocate and analyse footage of two attacks by security forces on an emergency department in Sudan in December 2021 and January 2022 where staff and patients were tear-gassed while inside the hospital. As noted by the author of the investigation, the purpose of this work is to stimulate conversation, research and development in the open source investigations community, the human rights field and the events happening in Sudan, as well as to document wrongdoing and identify those responsible.



## • Massacre in Tigray

An investigation by [BBC Africa Eye](#) uncovered evidence that a massacre in northern Ethiopia was carried out by members of the Ethiopian military. It also revealed the precise location of the atrocity, in which at least 15 men were killed. This investigation reconstructs the exact place of the massacre, the period in which it took place and even the identity of the perpetrators, without leaving London, only with the help of open source tools and techniques.





## Chapter 8

# Tools and Networks

- Bellingcat's Online [Investigative Toolkit](#)
- First Steps to [Getting Started](#) in Open Source Research
- [OSINTcurio.us](#) features weekly podcasts, webcasts and “10 minute tips” on video covering many aspects of doing open source investigations. It's a community project begun in late 2018 by about 10 contributing experts
- The [Open Source Intelligence Framework](#) has a very detailed and ever-growing list of digital investigative tools
- [Exposing the Invisible](#) Kit by Tactical Teck
- [Open Source Intelligence Techniques](#) by Michael Bazzell
- [Online research tools](#) by Global Investigative Journalism Network
- [The OSINT Framework](#)

## NETWORKS

- [Global Investigative Journalism Network](#) (GIJN)
- [Organised Crime and Corruption Reporting Project](#) (OCCRP)
- [Arab Reporters for Investigative Journalism](#) (ARIJ)
- International Consortium of Investigative Journalists (ICIJ)
- [C4ADS](#) is a non-profit organisation dedicated to data-driven analysis and evidence-based reporting of conflict and security issues worldwide.





ALJAZEERA  
MEDIA INSTITUTE

# Open Source **Investigation** *Handbook*

By

**Sara Creta**

Edited by:

**Muhammad Al khamaiseh**  
**Nina Montagu-Smith**

Designed by:

**Ahmad Fattah**

With special thanks to:

**Mohammed El-Haddad**  
**Phil Rees**



**ALJAZEERA**  
**MEDIA INSTITUTE**



AJmediatraining



+974 44897666

training@aljazeera.net

<http://training.aljazeera.net>