Royal United Services Institute
for Defence and Security Studies

Centre for Emerging Technology and Security

The Alan Turing Institute

Occasional Paper

The Future of Open Source Intelligence for UK National Security

Ardi Janjeva, Alexander Harris and Joe Byrne

The Future of Open Source Intelligence for UK National Security

Ardi Janjeva, Alexander Harris and Joe Byrne

RUSI Occasional Paper, June 2022

Royal United Services Institute
for Defence and Security Studies

191 years of independent thinking on defence and security

The Royal United Services Institute (RUSI) is the world's oldest and the UK's leading defence and security think tank. Its mission is to inform, influence and enhance public debate on a safer and more stable world. RUSI is a research-led institute, producing independent, practical and innovative analysis to address today's complex challenges.

Since its foundation in 1831, RUSI has relied on its members to support its activities. Together with revenue from research, publications and conferences, RUSI has sustained its political independence for 191 years.

The views expressed in this publication are those of the author(s), and do not reflect the views of RUSI or any other institution to which the authors are or were affiliated.

Published in 2022 by the Royal United Services Institute for Defence and Security Studies.

RUSI Occasional Paper, June 2022. ISSN 2397-0286 (Online).

Royal United Services Institute
for Defence and Security Studies
Whitehall
London SW1A 2ET
United Kingdom
+44 (0)20 7747 2600
www.rusi.org
RUSI is a registered charity (No. 210639)

Contents

About RUSI and CETaS

RUSI is the world's oldest and the UK's leading defence and security think tank. Its mission is to inform, influence and enhance public debate on a safer and more stable world. RUSI is a research-led institute, producing independent, practical and innovative analysis to address today's complex challenges.

The Centre for Emerging Technology and Security (CETaS) is a policy research centre based at The Alan Turing Institute, the UK's national institute for data science and AI. The Centre's mission is to inform UK security policy through evidence-based, interdisciplinary research on emerging technology issues. CETaS conducts policy-focused research and analysis on topics related to data science, AI, cyber security and privacy technology, engaging with a diverse range of stakeholders across government, academia and the private sector. Connect with CETaS at cetas.turing.ac.uk.

Executive Summary

THIS JOINT PAPER from RUSI and the Alan Turing Institute's Centre for Emerging Technology and Security aims to establish an independent evidence base to inform future government policy development regarding the use of publicly available information (PAI) and open source intelligence (OSINT) for national security purposes. The findings are based on in-depth consultations with stakeholders from across academia, civil society, commercial organisations, law enforcement and the UK government.

The paper explores the extent to which the increasing proliferation of PAI – and wider accessibility of tools leveraging PAI for OSINT – is changing perceptions of modern intelligence. From this foundation, it asks what the commercial, cultural, policy and technological implications are for UK national security stakeholders.

The writing of this paper coincided with Russia's invasion of Ukraine in 2022. This has emphasised the use of PAI and OSINT in the conflict environment. Government social media accounts have been publishing regular updates of the situation in Ukraine with open source corroboration,[1] while satellite imagery has featured consistently in traditional media columns.[2] A significant precedent has been set for the use of PAI to make decisive connections and provide critical insights in public. With the public able to 'see' more than ever before through high-quality open source reporting, the government's risk calculation for declassifying and sharing more secret intelligence has also changed. The justification required to keep intelligence secret (termed here the 'reasonable threshold of secrecy') has been markedly raised and it is unlikely there will be appetite to see that reversed. These developments mean that the findings and recommendations of this research are even more timely.

Government stakeholders who have embraced the opportunities provided by PAI and OSINT must now address the challenges required to sustain short-term successes over a longer-term period. Investment in PAI must keep pace with its increasing value to maintain an effective national security posture.

The research identified five central factors driving the importance of PAI in today's intelligence environment:

1. The **accelerating data curve**: An exponential increase in data poses acute challenges to the security community's approach to managing risk and proactively responding

[1] Ministry of Defence (@DefenceHQ), <https://twitter.com/DefenceHQ>, accessed 11 May 2022.

[2] *BBC News*, 'Ukraine War: Large Russian Convoy Redeploys Near Kyiv – Satellite Images', 11 March 2022; Chris Baraniuk, 'How Access to Satellite Images Shifts the View of War', *BBC News*, 22 March 2022.

to emerging security threats. Concurrently, this increase enables the development of unique solutions to investigative problems.

2. The **contributions of non-government actors to the OSINT space**: These stakeholders are conducting investigations and reaching conclusions which before were confined to government buildings and distributed in a limited capacity. This open playing field is spurring rapid developments in open source tradecraft and creating positive externalities for government stakeholders.
3. A favourable **cost equation** where PAI is becoming cheaper to collect and analyse while the cost of secret collection, validation and dissemination is increasing.
4. The **reciprocal investigative relationship between PAI and classified information**: The former might provide the first layer of intelligence knowledge with the latter enhancing this foundational intelligence, filling in gaps, and answering the 'so what' question of that information. Conversely, PAI might provide the final layer of verification via corroboration with other sources. The accelerating data curve is increasing the number of routes through which this investigative relationship can be reinforced.
5. The **use of PAI in parallel evidencing**: Legal considerations permitting, PAI can be useful in substantiating and triangulating insights gathered from secret sources, increasing levels of detail that can be provided to allies and the broader public alike, while protecting covert assets and methodologies. As the availability and value of insight from PAI increases, the value of insight into genuine secrets also increases as secret capabilities are targeted more efficiently.

A range of data sources can be used to leverage PAI for OSINT purposes, such as news, social media and other user-generated content, remote sensing data, commercial and third-party data sources, and grey literature (such as academic articles and books). Freely available software tools and the use of data science and machine learning techniques to augment analytical capabilities are also key to leveraging PAI. To use these sources and techniques effectively, the development of certain roles and skillsets is essential. The 'ideal' open source investigative unit would include a combination of information professionals, data scientists, analysts and subject-matter experts.

Despite the potential of those techniques and skillsets to revolutionise the way PAI is leveraged in the national security sphere, three main barriers to progress have been identified in this research:

1. **Tradecraft barriers**: Having rigorous processes in place to evaluate the veracity of any information output is vital. Despite some shared principles between verification for PAI and classified information respectively, this research found that the structures in place to analyse, curate and disseminate secret information have been developed extensively over time and seeped into the institutional memory of the UK security community. The same cannot yet be said for OSINT – filling this gap in tradecraft should be a policy priority.
2. **Cultural barriers**: Biases favouring classified information and internal datasets stem from institutional factors such as inertia and infrastructural factors limiting access to non-classified sources, bolstered further by the perceived power of the 'secret' label.

Meanwhile, current approaches to training and upskilling the security community to better leverage PAI are inadequate for the scale of the challenge and opportunity. The fragmentation of open source activity across government has also contributed to varying definitions, interpretations and perceptions of PAI and OSINT, making cohesion and standardisation more difficult.

3. **Resourcing and partnership barriers**: Underinvestment in digital tools and data resources is a risk to UK technological competitiveness. In some cases, individuals within government need more empowerment to build the best technical solutions. In others, a dynamic approach to procurement of OSINT tools will be necessary to take advantage of market innovation. While collaboration between government stakeholders and civil society stakeholders is challenging due to reputational and legal considerations, targeted collaboration across government itself must be supported.

This research has also highlighted a series of observations in the legal and ethical context. First, there is a disconnect between legislation pertaining to PAI, interpretations of that legislation in the form of internal policy and further interpretations of that policy which are baked into organisational culture. Second, there are notable differences in legal thresholds across sectors, contexts and departments, which can deter potentially valuable interactions with PAI in some instances and encourage it in others. There are also continuing changes to privacy regulation which may alter the ‘art of the possible’ in open source data accessibility for citizen investigators. Finally, the principles of necessity and proportionality that underpin the legitimacy of the national security community maintain relevance in the PAI and OSINT sphere – if information can be found via a publicly accessible search engine instead of more intrusive covert techniques, then there is a legal duty to use the less intrusive option.

Related considerations centre on the risk-versus-reward ratio when open source investigators reveal methodologies in granular detail. This granularity may be viewed by one party as an ethical issue – risking future investigations in the short term and by extension public safety – but by another as a pragmatic one, furthering innovation in open source techniques and thereby benefiting public safety in the long run. In most contexts, the net advantage of publicising tools and information so they can be used for public good is likely to outweigh the risk of adversaries better understanding the techniques used to investigate them.

With the adoption of this paper’s policy recommendations, there is a greater chance of overcoming these barriers to progress. This would realise the ideal scenario where OSINT is viewed firmly alongside other types of intelligence and integrated cohesively into existing analytical approaches across the national security community.

Introduction

THE ALAN TURING Institute's Centre for Emerging Technology and Security partnered with RUSI's Technology and National Security programme to conduct an independent research study into the future of open source intelligence (OSINT) for UK national security. The increasing proliferation of publicly available information (PAI) and wider accessibility of tools transforming PAI into OSINT are changing perceptions of modern intelligence. This paper explores the commercial, cultural, policy and technological implications of these developments for UK national security stakeholders, and establishes an independent evidence base to inform future policy development.

Methodology and Structure

The findings presented in this paper are based on semi-structured interviews and focus groups conducted virtually between January and March 2022 with 34 participants across academia, civil society, commercial organisations, law enforcement and the UK government,[1] alongside a targeted literature review focused on PAI and intelligence analysis in the national security context. This review primarily collected and utilised peer-reviewed academic papers, traditional media and 'grey' literature from the past five years, including blogposts, podcasts, research reports, social media posts and white papers. The semi-structured interview format allowed the research team to maintain a broadly consistent line of questioning, while still creating space to probe specialised areas of experience and knowledge among participants.

The research team relied on a purposive sampling strategy to identify interviewees who could provide detailed responses to research questions. Government practitioners were selected based on their first-hand experience of the opportunities and challenges in deploying PAI and OSINT in the security context, while non-government representatives offered perspectives from different operational and legal environments. While most interviewees were identified via pre-existing institutional relationships, a number were identified during the production of the literature review, complemented by a snowball sampling strategy.

Due to the unclassified nature of this study, in-depth analysis of current operational capabilities and techniques across government was not possible. Additionally, a thorough assessment of the legislative frameworks pertaining to the use of PAI and OSINT (and whether potential changes would be required) was outside the scope of this project. For this reason, the authors offer observations based on practitioners' experiences of operating within these frameworks.

[1] Throughout the rest of this paper, an anonymised coding system is used to refer to interviewee data. The prefix 'A' refers to academia, 'CS' to civil society, 'HMG' to UK government, 'IND' to industry and 'LE' to law enforcement.

The paper begins with a brief overview of the context and summarises the challenges in defining and interpreting PAI and OSINT. Chapter I explores the increasing prominence of PAI and OSINT in the intelligence environment and examines the related drivers of change. Chapter II explores current and emerging capabilities within the field, including relevant skillsets. Chapter III analyses the tradecraft, cultural and partnership barriers to the national security community deploying PAI as a core investigative resource, before Chapter IV offers a series of observations regarding the existing legal, ethical and policy context. Finally, Chapter V concludes by summarising the paper's main findings and recommendations.

The Context

On 24 February 2022, analysts from the Middlebury Institute of International Studies watched online as a column of tanks edged from the Russian city of Belgorod towards the Ukrainian border. By following traffic-jam alerts across 40 km on Google Maps – alongside a tip-off from a commercial satellite company that showed an absence of tents, indicative of troop movement – researchers correctly identified offensive Russian activity before it was made widely public.[2] Shortly after, Google disabled live traffic data from being displayed in Ukraine for the safety of local communities.[3]

The use of PAI to create OSINT is fundamentally changing the way that conflict is analysed and understood.[4] This is now being recognised beyond the loose networks of dedicated open source researchers online – the 2022 Russian invasion of Ukraine signalled a step change in the way that Western governments think and communicate about intelligence. In 2011, the Syrian Civil War was likely the first prolonged conflict where smartphones were used to collect and disseminate information widely in real time.[5] In the years prior, the transformation of Web 1.0 (read-only) to Web 2.0 (user-generated) heralded a new era of OSINT, but related tools and tradecraft were still in their infancy. In the ensuing decade, progress has accelerated. Using PAI to make decisive connections and provide critical insights in public is now recognised as increasingly important to an effective national security posture. This is evidenced by UK Armed Forces Minister James Heappey's comments that having seen 'open source intelligence', key 'combat enablers' being moved by Russia to the Ukrainian border was one of a number of indications to 'suggest final preparations [were] being made for them to invade'.[6]

[2] Rachel Lerman, 'On Google Maps, Tracking the Invasion of Ukraine', *Washington Post*, 25 February 2022; Tim Stenovec, 'Google Has Gotten Incredibly Good at Predicting Traffic – Here's How', *Business Insider*, 18 December 2015; Morgan Meaker, 'High Above Ukraine, Satellites Get Embroiled in the War', *Wired*, 4 March 2022.

[3] Marc Cieslak and Tom Gerken, 'Ukraine Crisis: Google Maps Live Traffic Data Turned Off in Country', *BBC News*, 28 February 2022.

[4] Alexa O'Brien, 'Open Source Intelligence May Be Changing Old-School War', *Wired*, 24 May 2022.

[5] Author interview with IND6, 26 January 2022.

[6] Aubrey Allegretti, Julian Borger and Daniel Boffey, 'Russian Claims of Ukraine Drawdown is Disinformation, Says UK Minister', *The Guardian*, 17 February 2022.

Members of the public can now see an invasion coming at roughly the same time as government officials with access to privileged sources of information, which means that the reasonable threshold of secrecy – the justification required to keep certain pieces of intelligence secret – is being raised. This was expressed by the director of GCHQ, Sir Jeremy Fleming, who recently spoke of an increased appetite to declassify and share secret intelligence:

> It is already a remarkable feature of this conflict just how much intelligence has been so quickly declassified to get ahead of Putin's actions.
>
> From the warnings of the war. To the intelligence on false flag operations designed to provide a fake premise to the invasion. And more recently, to the Russian plans to falsely claim Ukrainian use of banned chemical weapons.
>
> On this and many other subjects, deeply secret intelligence is being released to make sure the truth is heard. At this pace and scale, it really is unprecedented.
>
> In my view, intelligence is only worth collecting if we use it, so I unreservedly welcome this development.[7]

In large part, this refers to daily briefings from UK Defence Intelligence (DI) on the conflict, sharing details on air strikes, civilian displacement and territorial gains.[8] This type of intelligence sharing is occurring in a crowded environment which adversaries are seeking to pollute with disinformation: increasingly, a range of stakeholders are realising that transparency is essential not just in raising the level of informed debate domestically, but discrediting false narratives being peddled both domestically and abroad.[9]

A wide range of questions need to be answered to ensure that innovation in the application of open source techniques during the Russia–Ukraine conflict can be sustained and institutionalised over a longer period. The evidence base motivating this is described in Chapter I, where five drivers of the importance of PAI in the modern-day intelligence environment are outlined. Without capitalising on this momentum and implementing ambitious changes at the policy level, there is a risk that this period will be one that promised a revolution in intelligence analysis but only delivered limited compromise.

Various barriers to progress exist, and there are several pertinent legal and ethical considerations that need to be navigated. This will require an enabling and unified approach across government.

[7] GCHQ, 'Director GCHQ's Speech on Global Security Amid War in Ukraine', 31 March 2022, <https://www.gchq.gov.uk/speech/director-gchq-global-security-amid-russia-invasion-of-ukraine>, accessed 11 May 2022.

[8] Ministry of Defence (@DefenceHQ), 'Latest Defence Intelligence Update on the Situation in Ukraine – 03 March 2022', Twitter, 3 March 2022, <https://twitter.com/DefenceHQ/status/1499278184407314437>, accessed 16 May 2022.

[9] Julian Borger, Shaun Walker and Dan Sabbagh, 'Russia Plans "Very Graphic" Fake Video as Pretext for Ukraine Invasion, US Claims', *The Guardian*, 3 February 2022.

One of the areas where the absence of this is most stark is in the differing definitions, interpretations and perceptions of PAI and OSINT.

Definitions, Interpretations and Perceptions of PAI and OSINT

One of the first recorded instances of the use of PAI for national security was the Foreign Broadcast Monitoring Service, set up in 1941 to monitor and analyse Axis propaganda.[10] Following the Second World War, the department was integrated into the CIA to provide intelligence on foreign state media.[11] As the role of PAI in the analytical toolkit continues to grow, so do the variety of definitions and interpretations of it. Within government, perceptions of PAI and OSINT are tied up with long-held views about the value of classified information, and this dynamic is explored in more detail in Chapter III.

The research identified a lack of consensus on what counts as PAI, and wide variation in the makeup and goals of the communities that use OSINT. While most stakeholders would agree that freely available ship-tracking data or a public social media post showing Russian tanks being transported[12] is classified as PAI, there would be a conflict of opinion as to whether the most cutting-edge (and expensive) commercial satellite imagery or data acquired from the relatively inaccessible 'dark web' counts as publicly available.[13]

These differences extend to UK government and law enforcement. The UK Ministry of Defence's (MoD) most recent published definition of OSINT dates back to 2011: 'intelligence derived from publicly available information, as well as other unclassified information that has limited public distribution or access … [and is] exploited by trained analysts'.[14] In 2016, the National Police Chiefs' Council (NPCC) gave a more codified definition as:

> the collection, evaluation, and analysis of materials from sources available to the public … Open source is defined as publicly available information … It includes books, newspapers, journals, TV and radio broadcasts, newswires, Internet WWW and newsgroups, mapping, imagery, photographs, commercial subscription databases and grey literature (conference proceedings and institute reports).[15]

[10] Its UK partner remains an active organisation to this day. See *BBC Monitoring*, <https://monitoring.bbc.co.uk>, accessed 11 May 2022.

[11] See Joseph E Roop, *Foreign Broadcast Information Service History, Part I: 1941–1947* (Washington, DC: CIA, 1969); Cameron Colquhoun, 'A Brief History of Open Source Intelligence', *Bellingcat*, 14 July 2016.

[12] Janes, 'Russia Continues Military Build-Up on Ukrainian Border', 3 February 2022.

[13] Recorded Future, 'What is Open Source Intelligence and How Is It Used?', 19 February 2022, <https://www.recordedfuture.com/open-source-intelligence-definition/>, accessed 16 May 2022.

[14] Ministry of Defence, 'Joint Doctrine Publication 2-00: Understanding and Intelligence Support to Joint Operations', 3rd Edition, August 2011.

[15] National Police Chiefs' Council, 'NPCC Guidance on Open Source Investigation/Research', April 2015, <https://www.suffolk.police.uk/sites/suffolk/files/003525-16_npcc_guidance_redacted.pdf>, accessed 16 May 2022.

In this paper, PAI is used as a catch-all term for data or information that can be accessed by the general public with limited technical and economic resources and does not require membership of a specific organisation. OSINT is created when PAI is used to achieve a targeted investigative outcome. This is a definition deemed suitable for the parameters of this paper, and the authors do not intend for this to be seen as a conclusive definition that all stakeholders must adopt. Another relevant caveat is that not all 'social media data' is open source. This paper is primarily concerned with open source 'SOCMINT'[16] – that which is freely publicly available with no user verification requirement. From a global perspective, it is also important to acknowledge the global disparities in internet access,[17] which mean the definition of what is 'publicly available' is context specific and will vary across jurisdictions.

[16] 'SOCMINT' is a term first outlined in a 2012 Demos report, defined not by the openness of information on which intelligence is based, but its existence on a social media platform. See David Omand, Jamie Bartlett and Carl Miller, *#Intelligence* (London: Demos, 2012).

[17] Joe Tidy and Becky Dale, 'What Happens When the Internet Vanishes?', *BBC News*, 25 February 2020.

I. PAI in the Present-Day Intelligence Environment: Drivers of Change

THE PROLIFERATION OF PAI is augmenting collection methods and analytical techniques, and creating new opportunities for both the government and the public. This development is allowing researchers outside the traditional security community to deploy innovative methodologies on PAI with significant results, abetted by the decreasing cost of PAI collection and analysis. At the same time, classified information is becoming more costly to collect and analyse. For stakeholders with access to such information, this raises the central question of what a mutually reinforcing relationship between PAI and classified information looks like. A key component of this relationship has played out over the course of the Russian invasion of Ukraine: the use of open source insights to parallel evidence secret insights has allowed national security stakeholders to be more forthcoming with their intelligence, thereby narrowing the distance between them and the public. At a fundamental level, this has reinforced the importance of the information space as a central domain of modern-day conflict.

The Accelerating Data Curve

The first driver of change identified in the research is the rapidly accelerating data curve. Traditionally, open source requests within government were directed to libraries[18] and involved housing and land registry records, newspaper archives and telephone directories. Now, information sources range significantly across technologies (such as remote sensing technologies[19] and messaging platforms) and outputs (such as satellite imagery data and social media posts). As consumer demand continues to increase for devices and products to be connected for convenience and efficiency, the accelerating data curve is unlikely to slow down.

The exponential increase in available data has distinct operational and analytical implications.[20] On the operational front, 'hiding' becomes more difficult as covert or clandestine operations are easier to uncover than before. 'Seeking', however, becomes easier as clues about threat actors' networks and activities are more identifiable. On the analytical front, the increasing volume, velocity and variety of data flows is both a challenge and an opportunity depending on the supporting infrastructure available to filter the 'signal from the noise' – or equally, the information from the data. Traditional skills, systems and processes will struggle in an

[18] Author interview with A4, 19 January 2022.

[19] Defined by the US Geological Survey (USGS) as 'the process of detecting and monitoring the physical characteristics of an area by measuring its reflected and emitted radiation at a distance (typically from satellite or aircraft)'. See USGS, 'What is Remote Sensing and What is it Used for?', <https://www.usgs.gov/faqs/what-remote-sensing-and-what-it-used>, accessed 16 May 2022.

[20] William H Arkin, 'Inside the Military's Secret Undercover Army', *Newsweek*, 17 May 2021.

environment of too many 'signals', potentially making it easier for 'intelligence failures' to occur. Developing and maintaining modern computing and data infrastructure requires significant resources and a flexible approach from commercial providers to frequently adapt to their users' needs. Moreover, processes need to keep pace with the emergence of new technologies (such as applied, mixed and virtual reality) and applications within evolving environments (such as smart cities[21]) which will further augment the data curve.[22]

Across even the most basic computing environments, investigators can already use an open source radar interference tool to locate military radar,[23] conduct due diligence on a prospective client[24] or challenge claims about troop movements across borders in near real time using satellite imagery.[25]

[21] Joseph Bogan and Aimee Feeney, 'Future Cities: Trends and Implications', Dstl, 2020, <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/875528/Dstl_Future_Cities_Trends___Implications_OFFICIAL.pdf>, accessed 11 May 2022.

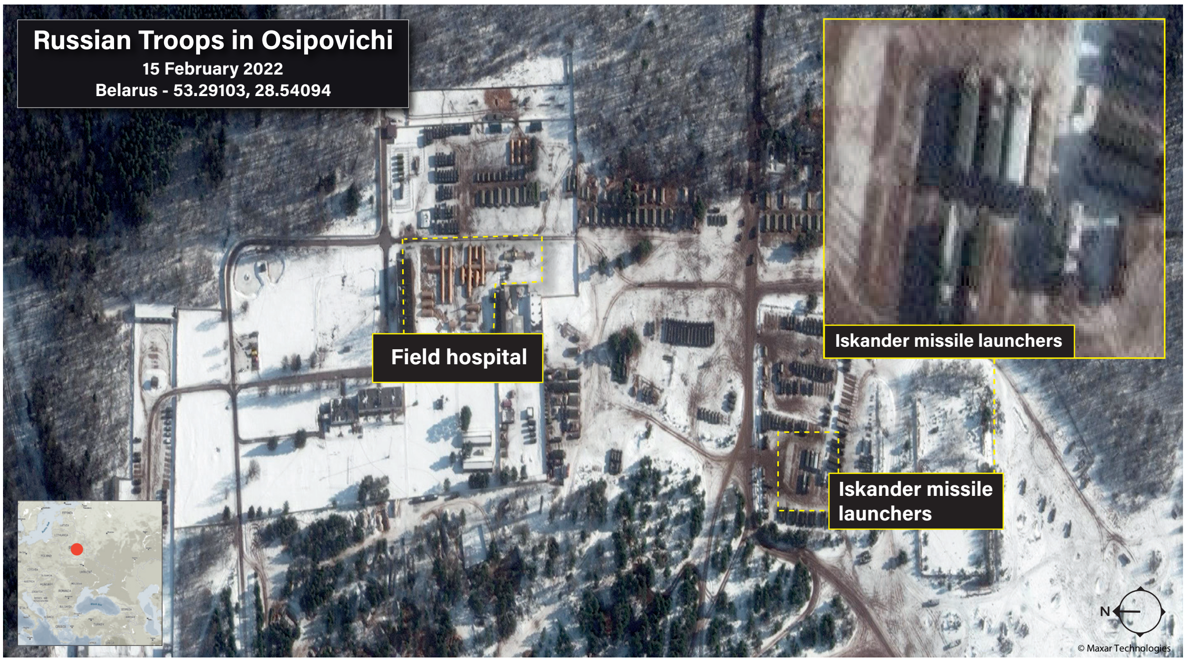
[22] Author interview with LE4, 27 January 2022.

[23] Ollie Ballinger, 'Radar Interference Tracker: A New Open Source Tool to Locate Active Military Radar Systems', *Bellingcat*, 11 February 2022.

[24] Companies House, <https://www.gov.uk/government/organisations/companies-house>, accessed 10 May 2022.

[25] Patrick Tucker, 'Satellite Images and Experts Challenge Russian Withdrawal Claims', *Defense One*, 15 February 2022, <https://www.defenseone.com/threats/2022/02/satellite-images-and-experts-challenge-russian-withdrawal-claims/362045>, accessed 16 May 2022.

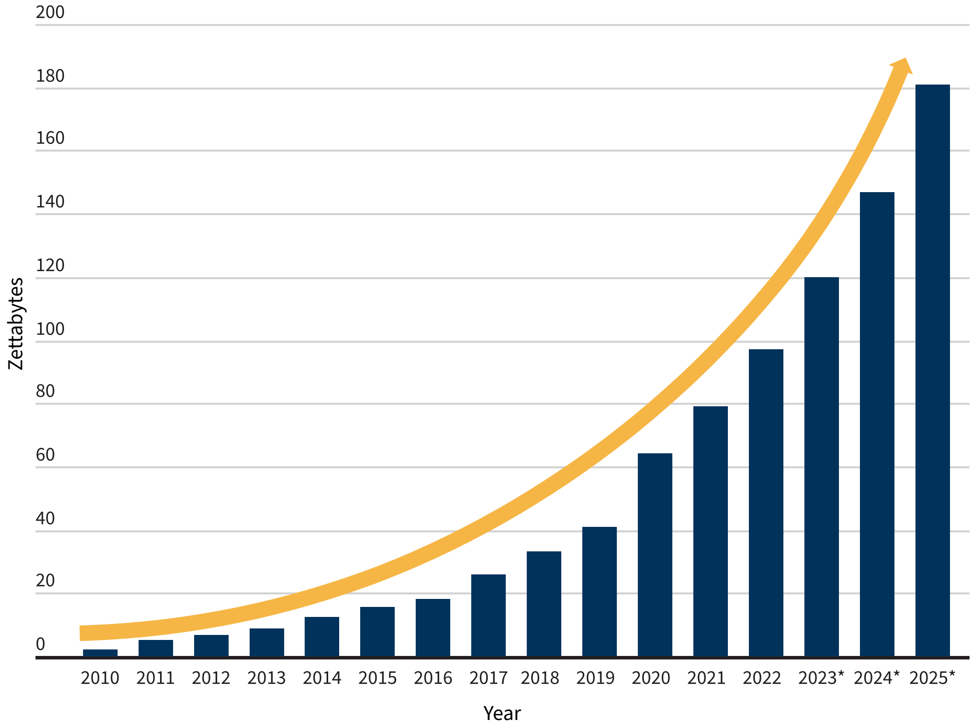
Figure 1: Example of Satellite Imagery Used to Assess Russian Troop Locations Prior to the Russian Invasion of Ukraine in February 2022


Russian Troops in Osipovichi
15 February 2022
Belarus - 53.29103, 28.54094
Field hospital
Iskander missile launchers
Iskander missile launchers
N
© Maxar Technologies

Sources: Maxar Technologies; RUSI Open Source Intelligence and Analysis.

As the amount of available data continues to increase, technological developments that help analysts locate and filter the most relevant data will reveal changes in the threat environment with lower latency. Investigators will thus benefit from better situational awareness and decision-makers will have a richer evidence base on which to draw analytic conclusions.

Figure 2: The Accelerating Data Curve



Sources: IDC; Seagate; Statista, 'Volume of Data/Information Created, Captured, Copied, and Consumed Worldwide from 2010 to 2025', <https://www.statista.com/statistics/871513/worldwide-data-created/>, accessed 1 June 2022.

Efforts are underway to support government analysts and decision-makers with the tools required to effectively manage these rapidly increasing data volumes. In UK government, the Cabinet Office is developing a platform – Information and Data Exchange (INDEX)[26] – that will help analysts process information across both publicly available sources and internal government reporting.[27] The overarching aim of INDEX is to serve as the definitive, shared space for users to find, connect and promote open source material. It should be a central function in the attempt to elevate the status of OSINT as a 'core' intelligence discipline and ensure it is fully integrated into analysis products for senior decision-makers.[28]

[26] Johnny Hugill, 'Introducing INDEX – A Cross-Government Digital Information and Data Exchange', *PUBLIC*, 18 October 2021, <https://www.public.io/blog-post/introducing-index-a-cross-government-digital-information-and-data-exchange>, accessed 16 May 2022.

[27] Author interview with HMG9, 25 January 2022.

[28] 'Professor Omand and Suzanne Raine called for the creation of a joint centre of expertise on open-source information to drive its use in the intelligence community and across Whitehall, and to ensure that "there is proper training available in the access to and safe use of open sources"'. See Joint Committee on the National Security Strategy, 'The UK's National Security Machinery: First Report of Session 2021–22', 19 September 2021, p. 34.

The type of innovation described above is reliant on either significant internal data science capability or external resourcing for contracting.[29] In either case, training and upskilling is essential to effectively utilise this improved analytic capability, as explored later in the paper.

Contributions of Non-Government Actors to OSINT Investigations

Since 2014, non-government organisations, such as Bellingcat, the *BBC* and the *New York Times*, have made significant strides in the use of open source tools and techniques to conduct investigations into the activities of individuals, entities and states.[30] These organisations and others have been able to follow Russia's military build-up near Ukraine,[31] document human rights abuses in Myanmar[32] and track North Korean sanctions evaders[33] by using PAI in innovative ways. One of the effects of this has been to expose disinformation and, in some cases, compel the acceptance of responsibility for actions previously denied.[34] The very act of exposing false narratives is itself not new, but there is a foundational shift in government stakeholders no longer being the ultimate arbiters of the public's access to the intelligence cycle. This has, in turn, expanded the number of routes through which states can be held to account, international law upheld and illegal acts exposed, thereby limiting future threats to society.[35]

Some open source investigations conducted by non-government stakeholders may misattribute a particular incident or individual with potentially serious consequences, as was the case in the 2013 'Find Boston Bombers' Reddit thread.[36] The challenges associated with verifying PAI and demonstrating analytical rigour are explored in more depth later in the paper, but it is worth noting here that there are dedicated processes to review and remedy such 'intelligence failures' within government, while equivalent processes outside government are likely to be more ad hoc and less regulated by law.

[29] Author interview with LE1, 21 January 2022; author interview with LE4, 27 January 2022.

[30] See Bellingcat, <https://www.bellingcat.com>, accessed 10 May 2022*; BBC Africa Eye*, <https://www.bbc.co.uk/programmes/w13xttpn>, accessed 10 May 2022; *New York Times*, 'Visual Investigations', <https://www.nytimes.com/spotlight/visual-investigations>, accessed 10 May 2022. Author interview with CS1.

[31] Christoph Koettl et al., 'Tracking Russia's Latest Military Movements Around Ukraine', *New York Times*, 14 February 2022, <https://www.nytimes.com/video/world/europe/100000008206442/russia-military-ukraine.html?playlistId=video/investigations>, accessed 16 May 2022.

[32] Myanmar Witness, <https://www.myanmarwitness.org>, accessed 10 May 2022.

[33] James Byrne et al., 'Black Gold: Exposing North Korea's Oil Procurement Networks', Project Sandstone Special Report, RUSI and C4ADS, March 2021.

[34] Farnaz Fassihi, 'Anatomy of a Lie: How Iran Covered Up the Downing of an Airliner', *New York Times*, 26 January 2020.

[35] Burma Campaign UK, 'Serbian Arms Exports to Burma Exposed by UN and Myanmar Witness', 25 February 2022, <https://burmacampaign.org.uk/serbian-arms-exports-to-burma-exposed-by-un-and-myanmar-witness/>, accessed 10 May 2022.

[36] *BBC News*, 'Reddit Apologises for Online Boston "Witch Hunt"', 23 April 2013.

Open source investigations can have significant impact. Box 1 highlights the 2018 *BBC Africa Eye* short, ‘Anatomy of a Killing’, which offers a notable example.[37]

Box 1: ‘Anatomy of a Killing’, *BBC Africa Eye*, 2018

In 2018, *BBC Africa Eye* released a short film which detailed the results of their investigation into the killing of two women and two young children in Cameroon. They were able to prove exactly where and when these executions took place and which military officials were responsible for them, despite initial dismissals from the Cameroonian government of the video evidence being ‘fake news’. Due to the precise and replicable evidence presented, as well as the international coverage of the investigation, the soldiers were eventually arrested.

This example shows the power of diligent, replicable and verifiable open source research in holding individuals, institutions and governments to account.

Source: *BBC News*, ‘Cameroon Atrocity: What Happened After Africa Eye Found Who Killed This Woman’, 30 May 2019.

The PAI–Classified Information Cost Equation

The complex picture of the present-day intelligence environment indicates the importance of all-source fusion,[38] which produces a reinforcing relationship between classified sources and PAI. It is no longer a matter of selecting one form of intelligence or the other – in isolation, they will not be enough to meet the increasing demands being placed on the UK national security community by the sheer scale of data.[39]

One aspect underlying this relationship is the changing nature of the cost equation between PAI and classified information: PAI is becoming cheaper to collect while the cost of secret collection and dissemination is increasing.[40] Some of the reasons for this include: electronically connected environments; the proliferation of data analytics and surveillance technology among mid-capability actors; rapidly developing capabilities in cyber security and increasing knowledge of investigative tradecraft among adversaries; and fewer vulnerabilities to exploit as a result.[41]

[37] *BBC News Africa*, ‘Cameroon: Anatomy of a Killing – BBC Africa Eye Documentary’, 24 September 2018, <https://www.youtube.com/watch?v=XbnLkc6r3yc>, accessed 16 May 2022.

[38] All-source fusion is the use of all available forms of intelligence within a product, team or organisation (for instance, the combined use of geospatial intelligence [GEOINT], human intelligence [HUMINT], OSINT, signals intelligence [SIGINT] and telemetry intelligence [TELINT]).

[39] Edward Lucas, ‘The Spycraft Revolution’, *Foreign Policy*, 27 April 2019.

[40] Author interview with A1, 7 January 2022.

[41] *Ibid.*

The term 'cost' is used in its broadest sense. First, there is a distinctly financial element to it. Physical surveillance operations are expensive to action and maintain – such costs might be challenging to justify within a resource-constrained environment if similar value can be gained from information obtained via PAI.[42]

Second, there is a cost in terms of practicality and convenience. For example, to access a piece of sensitive information, an analyst may be required to enter a different part of the building they are working in, log on to an inefficient and unwieldy system, and access the information which itself takes a very particular form and cannot be readily integrated with other datasets.

These factors highlight the importance of an agile approach to intelligence which continues to recognise the value of secret collection, but only insofar as it advances the cause of integrated all-source analysis.

The Reciprocal Investigative Relationship Between PAI and Classified Information

There are two important questions regarding the role of PAI in obtaining better security outcomes: what investigative ground do PAI sources help cover? And how should OSINT be fused into the intelligence effort to produce value alongside classified information? Research for this paper identified a reciprocal investigative relationship between PAI and classified information. In one instance, PAI can be more useful in providing the first layer of intelligence knowledge – moving from pre-suspicion to offender identification and initial understanding of activity – rather than serving the decisive piece of evidence that settles an investigation beyond reasonable doubt.[43] High-quality collection across more traditional intelligence disciplines (itself conducted in a more targeted fashion through effective use of OSINT) may then be best placed to enhance this foundational intelligence, filling in gaps, and helping to answer the 'so what' question of previously collected information. In another instance, it may be more traditional intelligence disciplines providing the foundational intelligence. Where possible, case studies should be shared within government to demonstrate this reciprocal dynamic and challenge a decades-old paradigm where classification is considered a de facto requirement for information to be of intelligence value.[44]

Parallel Evidencing

At the same time, more focus is needed on the role that PAI can play to substantiate and triangulate insights gathered from secret sources and – where relevant – protect sensitive sources and methods. This reinforces the narrative of a mutually beneficial relationship between PAI and classified information. Moreover, it speaks to the importance of using high-end state capabilities

[42] Author interview with IND3, 18 January 2022.

[43] Author interview with A2, 11 January 2022.

[44] Cortney Weinbaum and John N T Shanahan, 'Intelligence in a Data-Driven Age', *Joint Force Quarterly* (Vol. 90, No. 3, 2018), p. 9.

in the most proportionate way: if there is a piece of intelligence that can be gathered from a detailed internet search or from a public database of a government or international organisation, this becomes a more cost-effective and less intrusive course of action. Crucially, this frees up secret intelligence collection capabilities to generate genuine information advantage by revealing information that cannot be discovered from PAI, such as an actor's intent. This was illuminated in the run-up to Russia's invasion of Ukraine: PAI could demonstrate Russia's troop build-up in detail, but could not explain whether the reasons for it were to create coercive leverage in negotiations or to prepare for direct military action.

The emphasis on parallel evidencing seen in the context of Russia's invasion of Ukraine allowed government officials to make claims that they knew to be valid because of secret intelligence, while the public were able to 'see the evidence for themselves' and place certain actions and decisions within a wider context. Importantly, this added insight was courtesy not just of government officials and departments, but of reputable non-government stakeholders such as Bellingcat who attributed events in granular detail.[45] This facilitated a raising of the threshold of secrecy – the justification for keeping significant pieces of intelligence secret became more stringent. As a result, a greater onus could be placed on important stakeholders to take decisive action, while also delegitimising the stance of plausible deniability that adversaries usually adopt following claims about their hostile activity.[46]

[45] Bellingcat Investigations Team, 'Documenting and Debunking Dubious Footage from Ukraine's Frontlines', *Bellingcat*, 23 February 2022; Nick Waters, '"Exploiting Cadavers" and "Faked IEDs": Experts Debunk Staged Pre-War "Provocation" in the Donbas', *Bellingcat*, 28 February 2022.

[46] Author interview with IND6, 26 January 2022.

II. Current and Emerging Capabilities

THIS CHAPTER PROVIDES an overview of current and emerging capabilities in respect of leveraging PAI for OSINT and the various criteria that need to be met to ensure this is done effectively. It is split into two parts: the first considers data sources and techniques that are integral to practicing OSINT, while the second outlines the makeup of the 'ideal' open source investigative unit and the skillsets that need to be developed to support that.

Data Sources and Techniques

News, Social Media and Other User-Generated Content

Some of the most accessible data sources and techniques listed here are news, social media and other user-generated content. Today, breaking news is just as likely to be picked up first on social media as it is to be discussed in a government briefing room. This was recently emphasised by Shadow Foreign Secretary David Lammy:

> I'm coming home to my teenagers asking questions about Ukraine because they see it online every day. They feel as up to date as I am even though I get the intelligence briefings that you would expect … that raises very different questions about public attitudes to conflict that I suspect we are still digesting.[47]

Given the vastness of social media and news on the internet, having a process for approaching these information streams is imperative. A first step may involve deploying web scrapers on several specified sites to collect targeted information (such as commentaries, expert opinion data and 'grey' literature) via keywords about an event to improve situational awareness.[48]

Finding innovative ways to search for content is a key part of the open source toolbox. 'Google Dorking' is one example of an approach that uses specific modifiers on Google services to locate valuable content or data that is otherwise hard to find or to inspect vulnerabilities in websites.[49] One example of this targeted approach to social media searches was the geolocation of a video posted in early 2022 by Chechen troops in Ukraine to a town outside Kyiv,[50] providing intelligence

[47] RUSI, 'David Lammy on Russia's Invasion of Ukraine', Members Event, 17 May 2022.

[48] Johanna Wild, 'This New Tool Lets You Analyse TikTok Hashtags', *Bellingcat*, 11 May 2022.

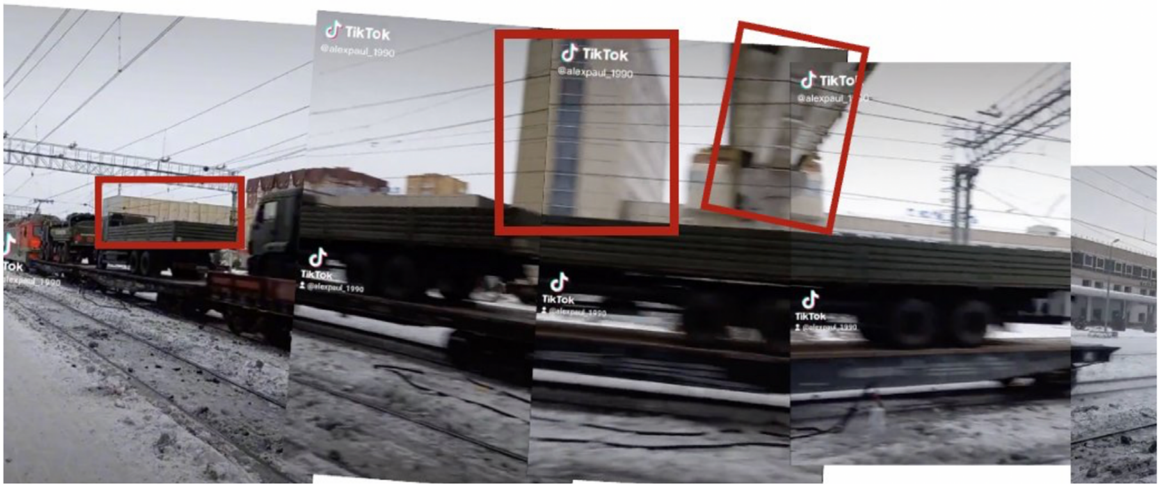
[49] Author interview with A3, 11 January 2022; Maltego Team, 'Useful Google Dorks for Open Source Intelligence Investigations', Maltego, 22 July 2021, <https://www.maltego.com/blog/using-google-dorks-to-support-your-open-source-intelligence-investigations>, accessed 16 May 2022.

[50] Benjamin Strick (@BenDoBrown), 'Geolocation of Kadyrov Chechen fighters in Babyntsi (Бабинці). Location: 50.640541, 30.023239 about 35km NW of Kyiv', Twitter, 7 March 2022, <https://twitter.

on specific military units and movements. The same individual in this video had their identity revealed in combination with other open source techniques such as facial recognition and social media analysis.[51] However, in many cases, assurance about the veracity of information found on social media can be challenging. This type of concern is discussed in more detail in the section on verification and analytical rigour below.

One important caveat is that not all 'social media data' is open source. To access social media pages that can only be viewed by logging into a user account, a law enforcement or intelligence agency would require the relevant legal authority. This restriction does not apply to private citizens or commercial organisations. This paper is primarily concerned with open source 'SOCMINT' – that which is freely publicly available with no user verification requirement.

Figure 3: Images Captured on Social Media of Russian Military Equipment Being Transported Towards the Ukrainian Border in February 2022



Sources: Ross Burley, 'Eyes on Russia Project: Latest Development', Centre for Information Resilience (CIR), 15 February 2022, <https://www.info-res.org/post/eyes-on-russia-project-latest-developments>, accessed 18 May 2022; TikTok.

com/BenDoBrown/status/1500800813916114948>, accessed 16 May 2022.

[51] Tactical OSINT Analyst (@OSINT_Tactical), '1/15 OSINT: Open Source Intelligence investigation & the use of facial recognition. Trigger to the investigation: Russian propaganda Military video posted on Telegram by what seems to be a Chechen Muslim fighter. #ukraine #islamicfighters #chechen #osint #Telegram #facialrec', Twitter, 1 March 2022, <https://twitter.com/OSINT_Tactical/status/1498694266754899978>, accessed 16 May 2022.

Remote Sensing Data

The importance of remote sensing data, such as satellite imagery, to OSINT in the conflict environment was highlighted at the beginning of this paper. This is a tool which commercial stakeholders are increasingly able to provide at high precision and relatively low cost.[52] Demand was recently demonstrated when Ukraine's vice prime minister, Mykhailo Fedorov, tweeted a request to commercial satellite operators to provide real-time synthetic aperture radar (SAR)[53] data to support Ukraine's armed forces with actionable geospatial intelligence regarding Russian troop movements.[54] Traditionally, this type of data was collected by a handful of states and classified at the highest levels. To offer a sense of historical scale, when Russia annexed Crimea in 2014, high-resolution commercial satellite imagery was still in its infancy and public-domain analysis 'as seen from space' could take weeks to emerge.[55] States are now openly requesting this same data from commercial organisations to assist in monitoring hostile troop movements; understanding the scale of an adversary's military arsenal; viewing damage caused by munitions; and changes in the climate or surrounding environment that might offer tactical and strategic advantage. Commercial imagery offers governments important opportunities to share data without referring to classified assets. The investment in expanding these capabilities was recently evidenced by the US National Reconnaissance Office (NRO) announcing their largest ever contract for commercial satellite imagery.[56] This can also be highly valuable for multi-national missions with non-Five Eyes partners who do not have access to classified imagery.

[52] Sandra Erwin, 'As Russia Prepared to Invade, U.S. Opened Commercial Imagery Pipeline to Ukraine', *Space News*, 6 April 2022, <https://spacenews.com/as-russia-prepared-to-invade-u-s-government-and-satellite-imagery-suppliers-teamed-up-to-help-ukraine/>, accessed 16 May 2022.

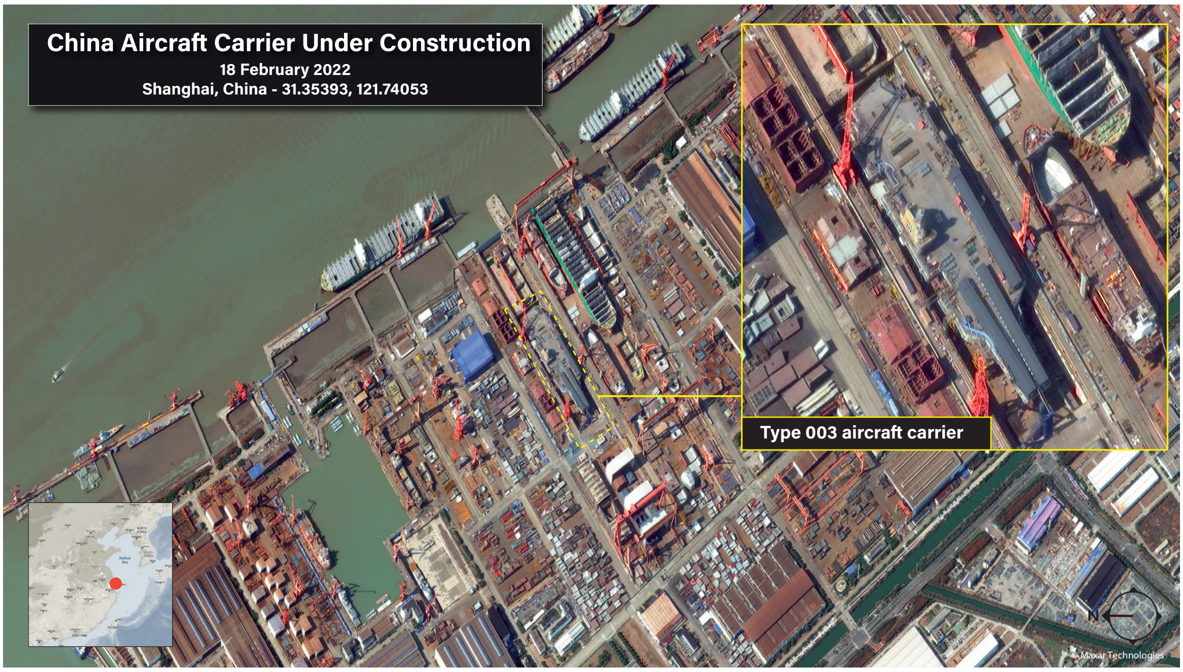
[53] SAR satellites emit their own energy and record the energy reflected back from the surface of the Earth. Their key advantage is that they can collect data during day, night and through cloud cover.

[54] Eric Berger, 'Ukraine Official Confirms Urgent Request for Western Satellite Data', *Ars Technica*, 1 March 2022, <https://arstechnica.com/science/2022/03/a-wartime-plea-to-western-satellite-companies-we-need-this-data-please/>, accessed 16 May 2022.

[55] American Association for the Advancement of Science, 'Satellite Imagery Assessment of the Crisis in Crimea, Ukraine – Part One: Sevastopol', 2014, <https://www.aaas.org/resources/geotech/sevastopol>, accessed 16 May 2022.

[56] National Reconnaissance Office, 'NRO Announces Largest Award of Commercial Imagery Contracts', 25 May 2022, <https://www.nro.gov/Portals/65/documents/news/press/2022/press_release_05-22.pdf>, accessed 25 May 2022.

Figure 4: Commercial Imagery of China's New *Liaoning*-Class Aircraft Carrier


China Aircraft Carrier Under Construction
18 February 2022
Shanghai, China - 31.35393, 121.74053
Type 003 aircraft carrier
N
© Maxar Technologies

Sources: Maxar Technologies; RUSI Open Source Intelligence and Analysis.

Other geospatial data such as radio frequency geolocation[57] and ship and flight tracking data[58] can also be crucial in advancing an investigation. Using satellite imagery and automatic identification system ship-tracking data, RUSI's Project Sandstone was able to demonstrate that North Korean vessels were violating sanctions in Chinese waters.[59] Additionally, the investigation suggested that China was likely aware of these violations, as imagery showed Chinese coast guard vessels sailing nearby (despite UN resolutions outlawing this activity).[60]

However, while it is now possible for a range of stakeholders to gain access to high-resolution remote sensing data, challenges in interpreting these outputs for analytical value should not be underestimated. Like other form of intelligence, analytical training is integral to exploiting raw

[57] HawkEye 360, <https://www.he360.com/>, accessed 10 May 2022.

[58] MarineTraffic, <https://www.marinetraffic.com/>, accessed 10 May 2022; ADB-S Exchange, <https://globe.adsbexchange.com>, accessed 10 May 2022.

[59] RUSI, 'Project Sandstone', <https://rusi.org/explore-our-research/projects/project-sandstone>, accessed 10 May 2022.

[60] Hamish Macdonald et al., 'North Korean Coal Smuggling Route to China Rebounds to Pre-Covid Activity Levels', *NK Pro*, 17 July 2020, <https://www.nknews.org/pro/north-korean-coal-smuggling-route-to-china-rebounds-to-pre-covid-activity-levels/>, accessed 16 May 2022; Patrick Tucker, 'This New AI Tool Can Help Spot an Imminent Invasion', *Defense One*, 14 December 2021, <https://www.defenseone.com/technology/2021/12/new-ai-tool-can-help-spot-imminent-invasion/359731/>, accessed 16 May 2022.

data and avoiding common pitfalls. For example, someone hired as an imagery analyst within Defence Intelligence's National Centre for Geospatial Intelligence will have to first complete the 16-week Defence Imagery Intelligence Analysis Course before formal enrolment.[61] As noted by Amy Zegart, 'imagery analysis requires considerable skill and training to know how shapes, shadows, sizes, scales, textures, perspectives, and contexts can obscure or delineate different objects seen from space'.[62] Although those without formal training can still usefully deploy remote sensing data, an awareness of verification and validation techniques (such as overhead imagery and differing weather conditions) is advantageous for any OSINT analyst examining this type of data source.

Other Commercial and Third-Party Data Sources

Outside of satellite imagery, news and social media sources, there is a wealth of third-party data that can be of significant value for OSINT purposes. For example, in Ukraine, location data from smartphones and devices – via Orbital Insight[63] – was used to calculate an estimated population of 5.4 million people in Kyiv the night before the Russian invasion. Approximately 10 hours later, the estimated population had decreased 20% to 4.4 million.[64] Although not without its limitations, from an operational and tactical perspective this is vital data that can help measure the scale and speed of civilian evacuation to support proactive humanitarian efforts.

Other relevant third-party data sources include corporate databases such as Companies House,[65] which provide details on company infrastructure and shareholders. In September 2019, in partnership with *The Times*, RUSI's Project Sandstone used this data to uncover UK front companies in ownership of vessels violating UN sanctions on North Korea.[66] The investigation demonstrated how individuals who operated companies sanctioned by the UN also had companies in the UK and were running coal-laden ships out of North Korea, possibly generating revenue for North Korea's ballistic and nuclear missile programme.[67] Using data from the Hong Kong and UK corporate registry in conjunction with ship-tracking data and satellite imagery, the investigation created a chain of evidence exposing this illicit activity. Following the investigation,

[61] Linkedin, 'Imagery Analysts', <https://www.linkedin.com/jobs/view/imagery-analysts-at-uk-ministry-of-defence-3036133557/?originalSubdomain=uk>, accessed 25 May 2022.

[62] Amy Zegart, *Spies, Lies, and Algorithms: The History and Future of American Intelligence* (Princeton, NJ: Princeton University Press, 2022), p. 241.

[63] Orbital Insight, <https://orbitalinsight.com/>, accessed 10 May 2022.

[64] *Nikkei Asia*, 'Satellite Images Reveal How Russia's Ukraine Invasion Unfolded', 25 February 2022, <https://asia.nikkei.com/Politics/Ukraine-war/Satellite-images-reveal-how-Russia-s-Ukraine-invasion-unfolded>, accessed 16 May 2022.

[65] Companies House, <https://www.gov.uk/government/organisations/companies-house>, accessed 10 May 2022.

[66] James Byrne et al., 'Project Sandstone Report 4: Down and Out in Pyongyang and London: North Korea's Coal Smuggling Networks Using UK Companies', RUSI, September 2019.

[67] Lucy Fisher, 'British-Owned Ships Smuggle Coal Out of North Korea', *The Times*, 26 September 2019.

several companies – some UK-based – and vessels were designated by the US Treasury for illicit North Korean shipping activity.[68]

It is worth noting that while many of these services can be freely available, large datasets such as satellite imagery catalogues or ship-tracking databases may be prohibitively expensive for individuals and small not-for-profit organisations conducting this type of work.

Analytic Capability

Freely available open source tools have significantly increased an analyst's ability to answer difficult questions. Tools such as QGIS and Google Earth Pro allow users to view, edit and plot geospatial data,[69] while SunCalc enables an analyst to gauge what time of day an image was taken via the length and presence of shadows.[70] These platforms provide an essential function in any OSINT investigator's toolbox by enabling geolocation (assessing exactly *where* an image or video was taken) and chronolocation (assessing *when* a piece of evidence was captured). A comprehensive and updated directory of available tools and capabilities can be located via Bellingcat's Online Investigation Toolkit[71] and OSINT Framework.[72]

[68] US Department of the Treasury, 'Treasury Sanctions Shipping Companies Transporting North Korean Coal', press release, 8 December 2020, <https://home.treasury.gov/news/press-releases/sm1204>, accessed 16 May 2022.

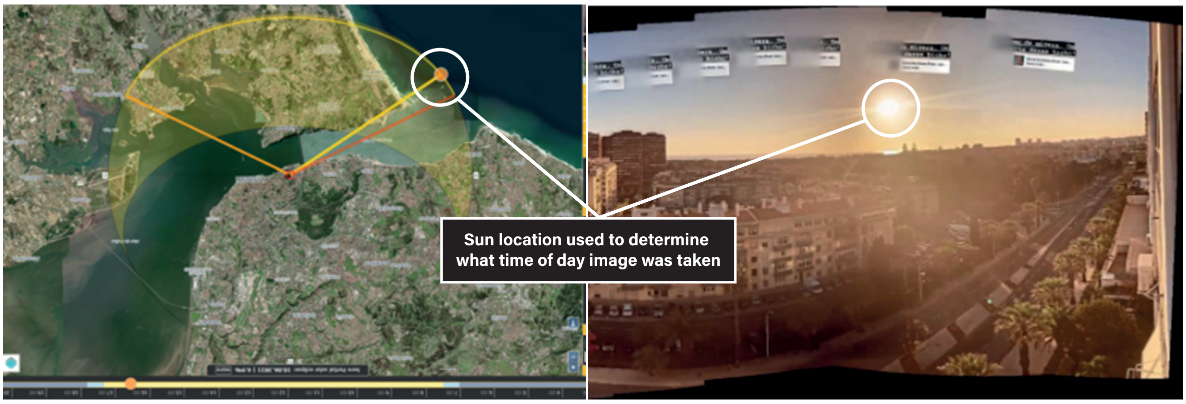
[69] QGIS, <https://qgis.org/en/site>, accessed 10 May 2022; Google Earth Pro, <https://www.google.com/intl/en_uk/earth/versions>, accessed 10 May 2022.

[70] SunCalc, <https://www.suncalc.org>, accessed 10 May 2022; Nick Waters, 'Unsure When a Video or Photo Was Taken? How to Tell by Measuring the Length of Shadows', *Bellingcat*, 18 May 2021.

[71] Bellingcat, 'Bellingcat's Online Investigations Toolkit', Version 6.8, 10 November 2021, <https://docs.google.com/spreadsheets/d/18rtqh8EG2q1xBo2cLNyhIDuK9jrPGwYr9DI2UncoqJQ/edit#gid=930747607>, accessed 16 May 2022.

[72] Justin Nordine, 'OSINT Framework', <https://osintframework.com>, accessed 16 May 2022.

Figure 5: Global Investigative Journalism Network Shows How to Chronolocate an Image from Social Media and Assess What Time a Photo Was Taken Using SunCalc


Sun location used to determine
what time of day image was taken

Sources: SunCalc; Youri van der Weide, 'Using the Sun and Shadows for Geolocating Photos and Videos', Global Investigative Journalism Network, 4 January 2021, <https://gijn.org/2021/01/04/using-the-sun-and-shadows-for-geolocating-photos-and-videos>, accessed 18 May 2022.

Developments in AI – and primarily in the sub-field of machine learning (ML) – are also a crucial component to consider when managing the opportunities and risks arising in the modern intelligence environment.[73] AI-assisted intelligence analysis can offer significant benefits in deriving insights from unstructured and disparate datasets.[74] This is of particular relevance to the open source context, considering the importance of the internet as a vector for intelligence insight, and the extent of human resource required for data discovery, collection and organisation.[75]

Cognitive automation of human sensory processing (particularly natural language processing[76] and audio-visual analysis[77]) is one area where investigators can hugely benefit from AI/ML, reducing the amount of time searching for keywords in text or recognising patterns in images that a machine could likely undertake more accurately and efficiently.[78] This could include

[73] Weinbaum and Shanahan, 'Intelligence in a Data-Driven Age', p. 9.

[74] *Ibid*., p. 11.

[75] Author interview with A3, 11 January 2022.

[76] Will Knight, 'As Russia Plans Its Next Move, An AI Listens to the Chatter', *Wired*, 4 April 2022.

[77] Intelligence Advanced Research Projects Activity, 'DIVA: Deep Intermodal Video Analytics', <https://www.iarpa.gov/research-programs/diva>, accessed 10 May 2022.

[78] Lindsay Freeman, 'Weapons of War, Tools of Justice: Using Artificial Intelligence to Investigate International Crimes', *Journal of International Criminal Justice* (Vol. 19, No. 1, 2021), pp. 35–53; João Rafael Gonçalves Evangelista et al., 'Systematic Literature Review to Investigate the Application of Open Source Intelligence (OSINT) with Artificial Intelligence', *Journal of Applied Security Research* (Vol. 16, No. 3, 2021), p.350; Esther Rolf et al., 'A Generalizable and Accessible Approach to Machine Learning with Global Satellite Imagery', *Nature Communications* (Vol. 12, No. 1, 2021), pp. 1–11; Eyal Weizman, 'Model Zoo', *Forensic Architecture*, 20 February 2020, <https://forensic-architecture.org/investigation/model-zoo>, accessed 16 May 2022.

identifying connections and similarities across datasets at an analyst's disposal, at which point particular subsets can be flagged or filtered for further human analysis.[79]

Box 2: Natural Language Processing

Natural language processing is a sub-field of AI that explores how computers process and interpret natural language data. It can be useful across the following tasks:

- **Automatic summarisation**: The rapid extraction of key insights from multiple text-based sources.
- **Entity and relationship extraction**: The creation of tailored information flows and knowledge graphing.
- **Machine translation**: The immediate availability of a multi-lingual information base.

The state of play of AI/ML analysis in the open source context is complex. A growing number of companies are formulating pitches around locating, collecting and organising unstructured data so that it can be best deployed within an open source investigation. This research identified tools focusing on systematic searching and the use of advertisement markers to assist geolocation; intuitive search tools which integrate with mainstream search engines and possess powerful video analytics capability; and software that allows biometric searching of images using facial recognition.[80] However, a large proportion of the field is still experimental with unrealised potential. A future scenario where there are tools that can ingest the latest news, achieve a degree of natural language understanding, react to a new piece of information and then isolate contradictory knowledge between them to produce a substantive answer (using natural language generation) is some way off.[81]

The primary value of ML is in reducing the time spent by analysts organising, prioritising and manipulating growing volumes of data. However, while ML-augmented analysis on PAI may help answer the 'what', it will not replace the human capacity to answer the 'so what', 'why' and 'what if' questions that are fundamental to intelligence analysis. Uniquely human traits, such as contextual understanding, counter-factual thinking and the application of theory will still be essential.

[79] Alexander Babuta, Marion Oswald and Ardi Janjeva, 'Artificial Intelligence and UK National Security: Policy Considerations', *RUSI Occasional Papers* (April 2020), p. 13.

[80] Author interview with A3, 11 January 2022; author interview with HMG6, 19 January 2022; author interview with LE4, 27 January 2022.

[81] Author interview with HMG4, 27 January 2022.

Currently, the wider application of cutting-edge AI/ML tools for OSINT across government remains limited. While a focus on better leveraging the potential of existing tools could yield significant gains, it is also important to ensure that future AI/ML tools reflect the specific needs of analysts in different types of roles. This is inherently linked to questions about training and upskilling, and procurement and partnerships, which are explored later.

People and Skillsets

As mentioned previously, OSINT is the product derived when PAI is used to achieve a targeted investigative outcome. For the purposes of this paper, the authors have modified the UK's Intelligence Handling Model used for lead development from 'Receive, Assess, Develop, Decide' to '*Collect*, Receive, Assess, Develop, Decide, *Disseminate*' to recognise the pertinence of active collection and publication within an OSINT investigation.[82]

To translate PAI into meaningful and usable OSINT products, a range of skills and expertise is required. Over the course of this research, the 'ideal' composition of an open source investigative unit was conceptualised across the following categories: information professionals; data scientists; analysts; and subject-matter experts. These skillsets are not mutually exclusive and are likely to be cross-cutting.

Information Professionals

An example of an information professional is a librarian – an individual who orders, collects, organises and provides information that matches (and potentially surpasses) a particular requirement. Knowledge information management remains one of the UK Civil Service Professions,[83] and professional qualifications can be gained from the Chartered Institute of Library and Information Professionals or the Information and Records Management Society.[84] These qualifications recognise the specialist nature of the precise location of information.

An informational professional might approach a problem in the following way:

- First, requirements within a question are interpreted and compartmentalised.
- Second, possible steps towards a comprehensive solution are outlined.
- Third, relevant information sources are explored.[85]

[82] David Anderson, *Attacks in London and Manchester, March–June 2017: Independent Assessment of MI5 and Police Internal Reviews* (London: The Stationery Office, 2017), p. 60.

[83] UK Government Knowledge and Information Management, 'About Us', <https://www.gov.uk/government/organisations/civil-service-government-knowledge-information-management-profession/about>, accessed 10 May 2022.

[84] Chartered Institute of Library and Information Professionals, <https://www.cilip.org.uk>, accessed 10 May 2022; Information and Records Management Society, <https://irms.org.uk/>, accessed 10 May 2022.

[85] Author interview with A4, 19 January 2022.

Being an information professional requires high-level understanding across a broad range of areas. For example, they will be able to quickly differentiate between what is or is not relevant for a particular outcome and be in possession of an acute sense of curiosity – those who see a problem as a puzzle to be solved rather than an obstacle.[86] When producing OSINT, an information professional would fulfil the function of 'data reconnaissance' – understanding what can be located and what can be leveraged to reach a desired outcome.

Data Scientists

Data scientists specialise in collecting, transforming and sharing datasets for the maximum extraction of value. In the research for this paper, it was suggested that focusing time and resources on these fundamental skills is more valuable than overspending on technical capabilities – such as monitoring tools – that can quickly become obsolete in a rapidly evolving technological environment.[87]

Establishing knowledge-sharing mechanisms to facilitate interaction between data scientists and other skilled professionals was also outlined as crucial within the present-day intelligence environment. For example, if focusing on collection and transformation, a data scientist plays an important role in validation, verification and visualisation. This might involve outlining the limitations of datasets to analysts and setting parameters for them to operate within.[88]

Analysts

An operational framework usually begins with questions, followed by the creation, dissemination and presentation of an intelligence product. It might conclude with the deployment of additional capabilities or resources. While an information professional will specialise in locating the most pertinent information, an analyst is focused on querying data sources, deriving insights from patterns, identifying courses of action and communicating findings to decision-makers.[89]

In analysis, the utility of data-centric skills – for example, being able to deploy a pre-written script that automates the collection of social media data – is becoming increasingly recognised across all sectors, including defence and journalism.[90] Other traits, such as creativity, enthusiasm and sound judgement, are also vital.

[86] *Ibid*.

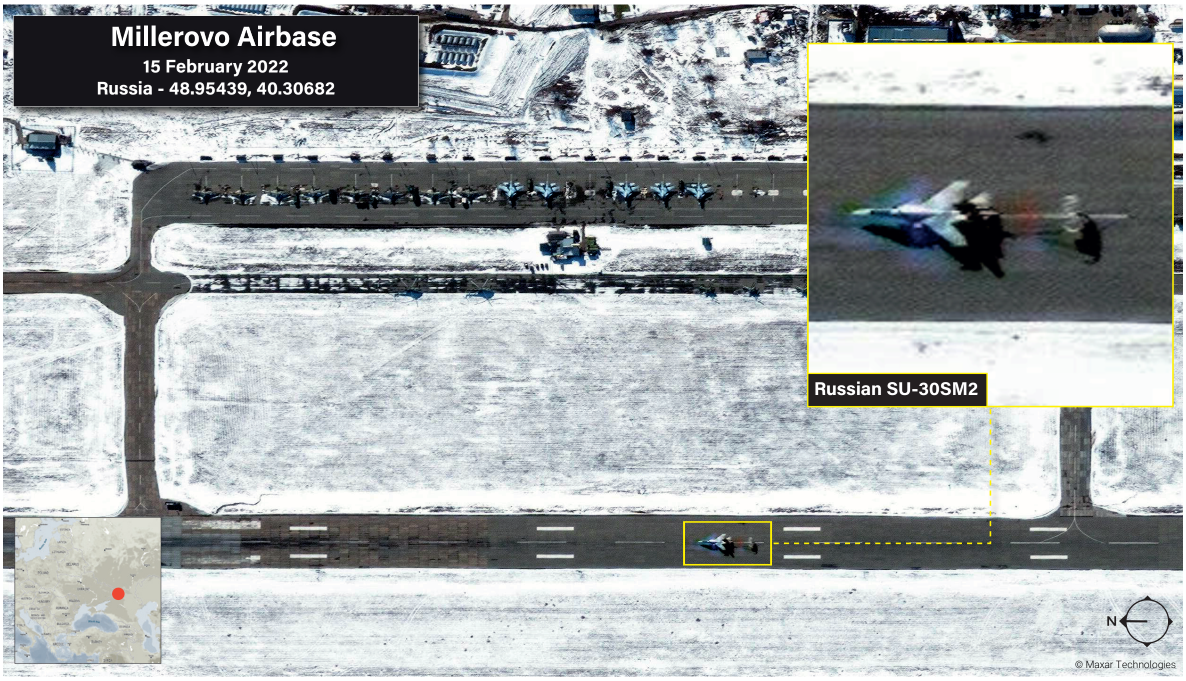
[87] Author interview with HMG1, 25 January 2022.

[88] Author interview with HMG3, 25 January 2022.

[89] GCHQ, 'How Does an Analyst Catch a Terrorist?', last reviewed 22 February 2019, <https://www.gchq.gov.uk/information/how-does-analyst-catch-terrorist>, accessed 16 May 2022.

[90] See, for example, James Kuht MBE (@KuhtJames), 'The jHub Coding Scheme means: 150 python coders, 90 web devs, 35 machine learning gurus, 44 agile practitioners, 13 NLP gurus, 8 app devs [and] 5 UX researchers are trained in Defence PER MONTH! In MOD? Want to get £ to learn to code? "jHub Coding Scheme" on Defnet', Twitter, 29 May 2020, <https://twitter.com/kuhtjames/

Figure 6: Satellite Imagery of a Russian Airbase with a Russian SU-30SM2 Landing on the Runway in February 2022



Sources: Maxar Technologies; RUSI Open Source Intelligence and Analysis.

Subject-Matter Experts

Subject-matter experts (such as behavioural scientists or linguists) can provide valuable insights that may not be gained by viewing data sources in isolation, such as in-country experience or social cues, including expressions or slang used within communities or groups.[91] However, subject-matter expertise – not unlike data science expertise – is resource-intensive to acquire, maintain and scale. Failing to meet this challenge may mean that vast swathes of foreign-language information are unassessed, thereby missing pivotal nuances in complex investigations.

A Whole Greater Than the Sum of its Parts

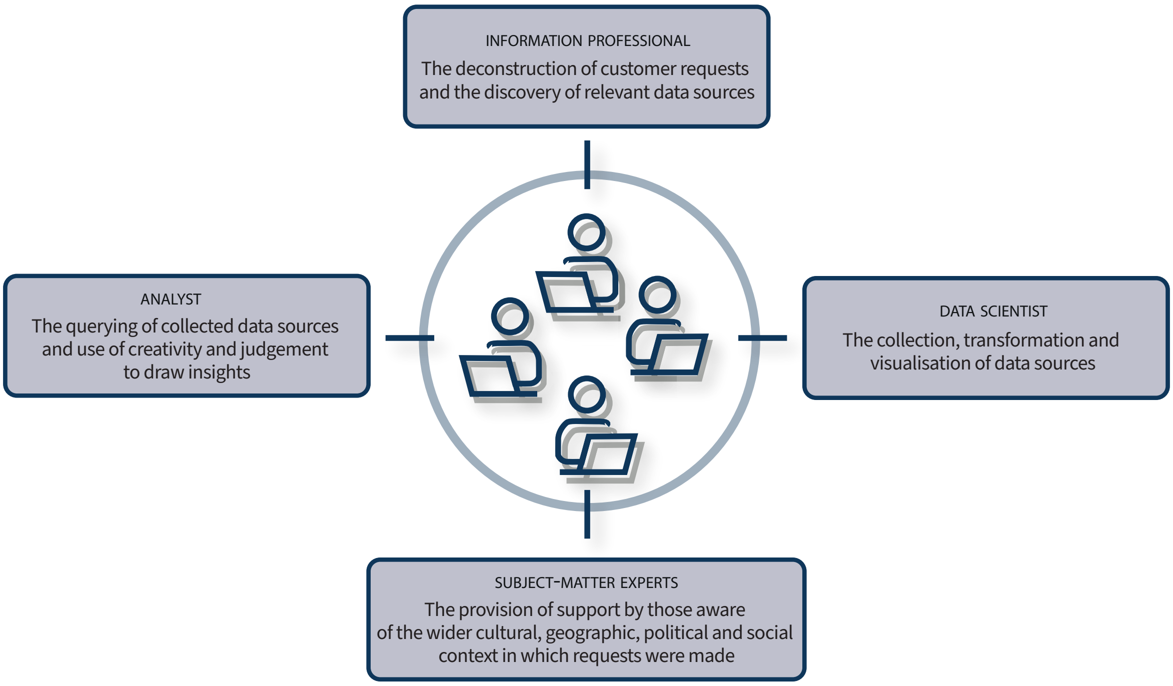
The above provides a conceptualisation of the 'ideal' open source investigative unit and the unique contributions of each of its components: the deconstruction of customer requests and the discovery of relevant data sources (the information professional); the collection, transformation and visualisation of data sources (the data scientist); the querying of collected data sources and

status/1266394161466482688?lang=en-GB>, accessed 16 May 2022; *Financial Times*, 'Data Visualisation', <https://www.ft.com/data-visualisation>, accessed 10 May 2022.

[91] Author interview with HMG1, 25 January 2022.

use of creativity and judgement to draw insights (the analyst); and the provision of support by those aware of the wider cultural, geographic, political and social context in which requests were made (the subject-matter expert).

Figure 7: The ‘Ideal’ Composition of an Open Source Investigative Unit


INFORMATION PROFESSIONAL
The deconstruction of customer requests and the discovery of relevant data sources
ANALYST
The querying of collected data sources and use of creativity and judgement to draw insights
DATA SCIENTIST
The collection, transformation and visualisation of data sources
SUBJECT-MATTER EXPERTS
The provision of support by those aware of the wider cultural, geographic, political and social context in which requests were made

Source: Author generated.

Overlap will likely exist between all these roles – a subject-matter expert helping an information professional to refine questions or a data scientist assisting an analyst in communicating their findings via data visualisations. Additionally, value placed on particular skillsets is likely to vary in accordance with customer requests and intended outcomes. It is important to recognise the fluid nature of investigations and that diversity across experience and skills is essential to ensure that the whole is greater than the sum of its parts.

III. Barriers to Progress

THE PREVIOUS CHAPTERS have aimed to lay out the context behind the rising importance of PAI in today's intelligence environment and provide some detail on the sources, techniques and skillsets required to effectively leverage PAI and OSINT. This chapter explores the barriers to progress in the national security community's engagement with PAI as a core investigative tool, broadly categorised into three segments: tradecraft barriers; cultural barriers; and resourcing and partnership barriers.

Tradecraft Barriers

Given how much of the UK national security community has been designed to service 'secret' processes and sources, there remains a lack of formal professionalisation of the skills required for analysing, curating, disseminating and, most importantly, integrating PAI throughout the intelligence lifecycle. Here, the first place to start is verification and analytical rigour with regard to PAI.

Verification and Analytical Rigour

While the range of sources and techniques used to gather PAI is impressive – and continues to grow – its utility is greatly reduced if rigorous processes are not in place to evaluate the veracity of any information output. Like other more traditional types of intelligence, there are significant risks in formulating products without proper verification processes in place.

A common example of this is where footage taken of one incident (such as a violent incident at a riot) later appears online claiming to prove a completely different incident taking place (such as a war crime). Often, falsely attributed media of this kind circulates rapidly through private messaging groups and social media platforms before it is discredited and established as being unconnected to the relevant incident at that time.[92]

Situations like these are indicative of an information environment infected by misinformation (accidental dissemination of incorrect information) and disinformation (intentional dissemination by threat actors with the purpose of deception or sabotage). Over the course of the authors' research, this was identified as one of the biggest risks to open source investigations. Worryingly, this problem was predicted to grow, with the prospect of more sophisticated deepfakes and dataset manipulation posing future threats to effective use of PAI.[93] An important trend to monitor will be whether technology that can detect and analyse content manipulation in audio, image, text and video keeps pace with the technology that exists to manipulate content in the

[92] Author interview with CS2, 17 January 2022; Marianna Spring, 'Coronavirus: The Viral Rumours That Were Completely Wrong', *BBC News*, 6 August 2020.

[93] Author interview with HMG2, 25 January 2022.

first place.[94] However, even if content is proven to be fake – as was the case with one deepfake of Ukrainian President Volodymyr Zelensky[95] – the initial impact of the disinformation is commonly so significant that it outstrips its disproval. With this caveat in mind, the outcome of this technological race is likely to have profound implications for public trust and trust in visual evidence – emphasising the importance of more formally professionalising OSINT capability as an integral part of wider intelligence skillsets. This research found that visual evidence forms a large basis of the impact of open source investigations due to its ability to cut through noise and link sequences of events to provide an inexorable conclusion in a manner which can be straightforwardly interpreted.[96]

There is a risk that adversaries may seek to understand the type of data that the UK national security community is interested in and then seek to influence (or 'poison') such data prior to collection. Approaches that might help to mitigate this risk include: a strong understanding of the information environments of both adversaries and allies; and pairing skilled subject-matter experts with innovative technologies that are able to detect and measure changes in particular themes or targets within the information space. Verification processes should evolve in correlation with new trends – a method for verifying information from one year ago may not work in two years' time, suggesting the need for ongoing review and updating of professional OSINT standards across the security community.[97]

A Gap in OSINT Tradecraft

It is helpful to distinguish between three main types of assurance activity:

- Validation of sources.
- Verification of individual pieces of information (see Box 3).
- Establishing robust analytic tradecraft for generating and testing analytic judgements (see Box 4).

Despite occasional scepticism levelled at PAI (largely driven by the misconception that non-secret information is less accurate or reliable than secret intelligence), the underlying tradecraft principles for OSINT are much the same as for classified intelligence. Accuracy, currency, operational utility, relevance, replicability via alternative sources and reputability are some basic

[94] Author interview with CS3, 19 January 2022; see, for example, Jack Berkowitz, Mona Gogia and Zig Hampel-Arias, 'Fake Finder: A Platform for Deepfake Detection', IQT, 9 September 2021, <https://www.iqt.org/fakefinder-a-platform-for-deepfake-detection>, accessed 18 May 2022.

[95] Tom Simonite, 'A Zelensky Deepfake Was Quickly Defeated. The Next One Might Not Be', *Wired*, 17 March 2022; Jane Wakefield, 'Deepfake Presidents Used in Russia–Ukraine War', *BBC News*, 18 March 2022.

[96] Author interview with CS3, 19 January 2022.

[97] Author interview with IND6, 26 January 2022.

criteria which apply in both contexts.[98] When producing an intelligence report, information has to pass through various levels of analysis and include the relevant caveats, regardless of whether the information is classified or open source.[99] Analysts working in intelligence are accustomed to operating with ambiguous, complex and partial data – for example, information received from a human intelligence (HUMINT) source will be closely cross-examined and questioned, rather than taken at face value.[100] This research established a clear need to ensure these core analytical practices and skills are embedded into the open source information environment and make use of data science expertise to ask nuanced questions of the PAI.

However, despite some of those shared principles, this research found that the structures in place to safely collect, analyse, curate and disseminate secret information are more labour intensive than what is routinely applied to PAI. The tried and trusted methods for assessing HUMINT – such as dealing with behaviour, motivations, uncertainty in actions and provision of guidance to ascertain confidence levels – have been developed extensively over decades and seeped into the institutional memory of the UK security community. The same cannot yet be said for OSINT, where an analyst carries the burden of responsibility for reliability and precision checks on information located on the internet.[101] The absence of a wider, standardised infrastructure in support of PAI analysis encourages a greater reliance on neatly packaged classified information and risks an undesirable situation where an OSINT product is treated differently despite being reliable on its own terms. Filling this gap in tradecraft should be a policy priority in the near future.

The authors encountered several examples of good practice from non-government stakeholders which could inform this approach, two of which are presented below.

Box 3: Geolocation

If content can be placed in time and space, misrepresenting it as false becomes much more challenging. Taking time to identify contextual clues within the content and corroborating it with other databases helps in identifying elements of fact and whether secondary analysis is required (for instance, comparing railway carriage numbers with a country's railway database).

Sources: Author interview with CS2, 17 January 2022; author interview with HMG9, 25 January 2022.

[98] Author interview with A3, 11 January 2022; Harry Kemsley, 'In OSINT We Trust?', *The Hill*, 1 September 2021.

[99] Author interview with HMG4, 27 January 2022.

[100] Author interview with HMG3, 25 January 2022.

[101] Author interview with HMG7, 25 January 2022.

Box 4: Evidence Documents, Red Teaming and Peer Review

Behind most published open source investigations lie an evidence document running between 100 and 150 pages, containing annotations, screenshots and transcripts which footnote the entry of every data point. Every assumption made in these evidence documents will be interrogated by a senior analyst or manager to ensure all assumptions are beyond reproach. Next, representatives external from the investigative team are asked to act as a 'red team' (namely, finding fault with what is being presented).

The importance of professionalised peer review and editorial policy is also increasingly recognised, particularly for investigations which may be used in legal proceedings.

Sources: Author interview with CS3, 19 January 2022; author interview with CS2, 17 January 2022.

Verification of PAI for the Evidential Context

The difficulties of verification in a saturated information environment and the approaches that can be used to address this carry extra importance in the context of legal and evidential proceedings. In this setting, having a defensible presentation of the investigative steps that were taken – supported by effective software tools (such as screen-recording software[102]) and detailed audit trails – is essential.[103] One example is Hunchly, an open source tool which allows the user to record their investigations and automatically databases information from the pages which were visited, creating an audit trail for an investigation.[104]

Where PAI is used for judicial outcomes, it is often received from reputable organisations (such as credit referencing agencies) which have an established tradition of cooperating with law enforcement and government. Transparency is also provided regarding the provenance of that information.[105] For these reasons, this is usually accepted as useful primary evidence. Where transparency may not be possible – if it is not in the public interest – evidence would be presented as being additional rather than primary.[106]

While the risks when OSINT is brought to a courtroom may not be higher than other types of intelligence, there are fewer precedents and a smaller body of case law on which to make this judgement.[107] Known examples of open source analysis such as geolocation verification being presented in court are limited. However, one notable case includes Bellingcat's linking

[102] See, for example, HM Government, 'Long Arm', https://www.digitalmarketplace.service.gov.uk/g-cloud/services/210748352904117>, accessed 25 May 2022.

[103] Author interview with A2, 11 January 2022.

[104] Hunchly, <https://www.hunch.ly/>, accessed 25 May 2022.

[105] Author interview with LE1, 21 January 2022.

[106] *Ibid*.

[107] Tom Simonite, 'The Race to Archive Social Posts That May Prove Russian War Crimes', *Wired*, 11 April 2022.

of an assassination in Berlin to Russian intelligence services, where the investigative methods adopted were deemed both admissible and proportionate.[108] Initiatives such as Syrian Archive and Myanmar Witness were also established to help document war crimes and preserve court-admissible evidence before it is removed.[109] More recently, the Centre for Information Resilience (also responsible for Myanmar Witness) started a crowdsourced effort to document significant incidents in Ukraine.[110]

Nevertheless, like other forms of intelligence, the majority of OSINT material is unlikely to reach the high provenance and accuracy thresholds required for evidential purposes. Where PAI is intended to be ultimately presented as evidence, additional verification and handling requirements must be imposed from the point of collection onwards to maintain integrity of the complete evidential chain.

Cultural Barriers

Biases favouring classified information and internal datasets, approaches to training and upskilling, and the fragmentation of open source activity across government are three core areas which illuminate the cultural barriers to progress in open source investigation.

Biases Favouring Classified Information and Internal Datasets

Institutional Factors: Inertia and Access

Two central institutional factors have impeded growing use of PAI in intelligence and investigation. The first is inertia and affects all organisations. Changing 'how things have always been done' – especially in relation to infrastructure for collecting, verifying and analysing internal datasets – will be a challenge. The central question is: what proportion of investigative value is going to come from open source, and what balance of technology and skills effort is required to exploit that?[111] The answer will have direct implications for future funding.

Assigning values to these questions is difficult, but one government interviewee suggested that funding for open source investigation may be as little as 1% relative to classified sources, while the percentage of intelligence stemming from open source investigation stands anywhere between 35% and 90%.[112]

[108] Author interview with CS2, 17 January 2022; Bellingcat Investigations Team, 'Berlin Assassination: New Evidence on Suspected FSB Hitman Passed to German Investigators', *Bellingcat*, 19 March 2021.

[109] Syrian Archive, <https://syrianarchive.org>, accessed 11 May 2022; Myanmar Witness, <https://www.myanmarwitness.org>, accessed 11 May 2022; James Clayton, 'Are Tech Companies Removing Evidence of War Crimes?', *BBC News*, 31 March 2022.

[110] CIR, <https://www.info-res.org>, accessed 11 May 2022; MapHub, 'Russia–Ukraine Monitor Map', CIR, <https://maphub.net/Cen4infoRes/russian-ukraine-monitor>, accessed 11 May 2022.

[111] Author interview with A2, 11 January 2022.

[112] Author interview with HMG8, 25 January 2022.

Comparatively, the chief of Defence Intelligence has stated that ‘significant, highly classified information … is perhaps about a fifth of the information that I need to be able to access’.[113] Better clarity on this matter must be achieved by departments that use this information in all-source assessments to understand the value that different intelligence types provide.

The second factor – operational in nature – is the limited access to non-classified sources that many intelligence community analysts may have in their working environment. Due to a combination of regulatory and technical constraints, it may be more challenging for an intelligence analyst to conduct the sort of open source information gathering that private citizens can conduct freely and unrestricted. One academic interviewee described security-cleared intelligence analysts as sitting in a fortress, invulnerable to attack from outside forces, but by the same token shielded from osmosis of information.[114] The corresponding analogy is one of a weather forecaster surrounded by the latest technology and big data models, yet unable to see outside the window and register the torrential rain.[115]

This disparity in on-tap access to openly available sources inevitably feeds into bias and preferential treatment for classified sources which – as previously explained – benefits from dedicated credibility and verification processes. There are other identifiable reasons for why these datasets often remain the first port of call in the national security community[116] – some investigative queries are so specific in nature that answers can only be gained via closed internal datasets.

The Power of the ‘Secret’ Label

A large part of the bias towards internal datasets is due to the influence that a security classification of ‘secret’ may possess. A consistent finding of this research was that a ‘secret’ stamp on a piece of intelligence is likely to ensure it commands more attention in the analytical process than something found on a search engine (even before that open material has been properly synthesised).[117]

This can be explained in two ways: trust and verification challenges surrounding OSINT; and an assumption that ‘secret’ intelligence will provide more interesting information in comparison to other types of intelligence. For example, officials attending an emergency meeting regarding a national security incident will enter that meeting with a baseline level of information gained via the news and other open sources – when someone in that meeting explains what the ‘secret’

[113] Angus Batey, ‘DSEI 2021: UK Chief Says Defence Intelligence at a Tipping Point’, Mönch Publishing Group, <https://monch.com/dsei-2021-uk-chief-says-defence-intelligence-at-a-tipping-point>, accessed 16 May 2022.

[114] Author interview with A4, 19 January 2022.

[115] *Ibid*.

[116] Author interview with LE6, 2 February 2022.

[117] Author interview with IND3, 18 January 2022.

intelligence says, the centre of gravity of the debate shifts in its favour.[118] What is lost in this process is the understanding that the 'secret' classification is often about the sensitivity of the source rather than the value of the information. Culturally, those lines have gradually blurred to the detriment of open source material.[119] It is important to recall that the history of intelligence is rife with examples of highly inaccurate and unreliable 'secret' sources. Certain kinds of 'secret' sources are likely to be more accurate than others and each case needs to be judged on its merits rather than cognitive biases and preconceptions. However, as outlined earlier, concerns about accuracy are prevalent in the open source space, too.[120]

The risk of overestimating 'secret' sources may be more germane with policy officials in government who carry the responsibility of executing final decisions than intelligence personnel who are more steeped in the nuances of the discipline.[121] By that token, there is undoubtedly an educational remedy to the classification bias problem, including conversations at leadership levels about the value of open source work and the technology behind it. Alongside some of the technical and skills-based recommendations advocated in the final chapter, these measures would help to ensure that OSINT is fully integrated into the overall intelligence effort.

Approaches to Training and Upskilling

Current approaches to training and upskilling the security community to better leverage PAI are inadequate for the scale of the challenge and opportunity. For open source techniques to fulfil their potential, a professionalised and mainstreamed approach to training is required. This will go a long way to ensuring that these techniques are not seen as an afterthought in practitioner training but a central component of it.[122]

While PAI may come from other sources, the practice of OSINT largely revolves around using the internet in smarter ways. This involves being more creative in the way that data is queried and more granular in the use of search engines.[123] Improving this baseline takes time and investment and the cultivation of a new type of analyst that is cognisant of the complex and overlapping paradigms of compliance, data governance, emerging technologies and investigations.[124]

A vital step will be moving away from the existing model where such approaches are seen as the niche responsibility of a limited number of people within an agency or department. Focusing a disproportionate amount of OSINT training resource on analysts within the UK intelligence community may be counterproductive considering the existing costs of maintaining their 'secret' work. From an efficiency perspective, it may be optimal to prioritise the upskilling of users who

[118] Author interview with HMG8, 25 January 2022.

[119] *Ibid*.

[120] Author interview with A4, 19 January 2022.

[121] *Ibid*.

[122] Author interview with HMG8, 25 January 2022.

[123] Author interview with HMG9, 25 January 2022.

[124] Author interview with LE4, 27 January 2022.

do not bear the infrastructural and operational costs of secrecy, while at the same time ensuring that the UK intelligence community is best equipped to partner with and integrate information from OSINT specialists.

In the policing context, PAI can be used to speed up investigations and assist in identifying individuals or locations of interest. Bringing front-line officers into the fold so they understand what they can and cannot do in relation to PAI is central to ensuring better outcomes and maintaining the legitimacy of the police in an era of information overload.[125] It is becoming essential to have a cadre of analysts who are proficient enough to understand code, and data scientists who can build bespoke solutions for these analysts.[126] This would mean that when companies develop AI/ML tools to support practitioners in the security community, analysts are confident enough to identify what is useful based on their collection requirements, and acquire only the tools that enable them to apply their own skills to the data effectively.[127]

Agile recruitment and upskilling of employees need to be complemented by effective training packages – for example, a curriculum that highlights the utility of OSINT and instils confidence and creativity within those who undertake it in equal measure. In this regard, this research found that there would be value in a generalist cross-government OSINT curriculum combined with specialist modules and secondment opportunities. The recently announced College of National Security,[128] alongside the Intelligence Assessment Academy and the MoD Open Source Centre of Excellence, would be ideally placed to lead on producing a generalist OSINT course for practitioners across government.

Fragmentation of Activity

A bird's-eye view of the approach to open source capability within government yields a slightly confused picture where there are encouraging pockets of activity without necessarily adding up to a coherent whole. Moving away from a scenario where open source activity is fragmented, sub-scale and underinvested requires sustained joint activity and collaboration, which in turn means moving beyond disputes about 'ownership of OSINT' within government.[129]

Sustaining Joint Activity and Collaboration

A scaled capability for leveraging open source techniques should recognise that one of the competitive advantages of government is its ability to collaborate. Even so, there are various

[125] *Ibid*.

[126] Emily Harding, 'Move Over JARVIS, Meet OSCAR: Open-Source, Cloud-Based, AI-Enabled Reporting for the Intelligence Community', CSIS, 19 January 2022.

[127] Author interview with IND6, 26 January 2022.

[128] Cabinet Office, 'New National Security College Founded to Boost UK and Australian National Security', press release, 28 February 2022, <https://www.gov.uk/government/news/new-national-security-college-founded-to-boost-uk-and-australian-national-security>, accessed 16 May 2022.

[129] Author interview with A1, 7 January 2022.

nuances that can make this challenging: two government departments each collecting satellite images may express a desire to collaborate, but on doing so find that the types of images they are collecting are crucially different.[130] In other cases, different agencies may have varying levels of appetite to replicate approaches to data and technology, as well as varying levels of permissions and access at different classification levels which prevent full-scale collaboration.[131] Nonetheless, resisting the temptation to retreat into the institutional inertias and classification biases discussed previously is crucial.[132]

In this vein, the Joint Intelligence Organisation within the Cabinet Office – in its role as Professional Head of Intelligence Analysis – should be primarily responsible for the professional exploitation of PAI for OSINT. This would align with its existing responsibilities to establish common tradecraft programmes which foster communities of expertise and common platforms (such as INDEX) which ensure better visibility of analytical outputs across government. Additionally, there are cross-government incentives that could be leveraged to drive efficient collaboration: funding programmes which encourage departments to bid jointly for shared PAI analysis capabilities; and enshrining alignment in accreditation policies which enable OSINT tools to be rolled out across multiple departments at once.

Going a Step Further: A Centralised Agency?

Some commentators (mostly abroad) have advanced the idea of a centralised OSINT agency as the solution to the challenge of fragmentation of activity.[133] Based on research for this paper, this is currently not the optimal course of action for the UK national security community. Rushing to a centralised agency model before taking stock of the challenges laid out here would risk several unintended consequences (such as further siloing of OSINT capabilities) and a premature abandonment of the ideal scenario where PAI and OSINT are viewed firmly in the context of the other types of intelligence and integrated into analytical approaches across the UK security community.

[130] Author interview with HMG3, 25 January 2022.

[131] *Ibid*; author interview with LE4, 27 January 2022.

[132] Author interview with IND3, 18 January 2022.

[133] Brian Katz, ‘Maintaining the Intelligence Edge: Reimagining and Reinventing Intelligence Through Innovation’, CSIS, January 2021; Kevin Johnston, ‘It’s Time to Give OSINT Its Own Agency’, *Fair Observer*, 25 February 2022, <https://www.fairobserver.com/region/north_america/kevin-johnston-osint-us-intelligence-community-internatinoal-security-news-35271>, accessed 16 May 2022; Bob Ashley and Neil Wiley, ‘How the Intelligence Community Can Get Better at Open Source Intel’, *Defense One*, 16 July 2021; Byron Tau and Dustin Volz, ‘Defense Intelligence Agency Expected to Lead Military’s Use of “Open Source” Data’, *Wall Street Journal*, 10 December 2021.

Resourcing and Partnership Barriers

As with any potentially significant development in national security architecture, the scale of ambition will be reflected by resourcing considerations and the extent to which effective partnerships can be formed with external stakeholders.

Investment and Procurement Challenges

Underinvestment in analysis tools and data resources is costly in an era where maintaining technological competitiveness is so reliant on it. Specifically, this means investment in an OSINT infrastructure that is commensurate with its ever-increasing value. Understanding the value of existing curated data and comparing that to the value of PAI to achieving an organisational mission – asking whether £1 spent on PAI is more useful than £1 spent on traditional operations – should lead to an investment equation which is more reflective of strategic and operational requirements rather than tactical history.[134] Giving sufficient weight to different types of intelligence tradecraft in a community which has been designed to service 'secret' processes and sources is a challenge, and even where the value of PAI to national security activities is appreciated, taking the leap to actively prioritising the resources for it carries its own difficulties. PAI that is nominally 'free' (some of it will not be) may still require expensive technical expertise to collect, develop pipelines and maintain the broader supporting infrastructure.[135]

Another important element of the resourcing problem is the distribution of capability across different stakeholders within government. Investment that is driven by the UK intelligence community, for example, may not necessarily benefit other government departments.[136] This includes policing, where there can often be an unclear perception of available resources and access to technology.[137] In an environment where a wide range of stakeholders within the UK security community need a more mature PAI and OSINT capability, concentrating the vast majority of funding within small pockets of it will not be a feasible solution.

In some cases, the best technical solutions will be possible and known to individuals already working within government, and the main barriers to progress are the permissions and empowerment afforded to them to progress those ideas. In other cases, the data sources and techniques that are essential for leveraging PAI for national security purposes will need to be brought in from external parties. Procurement processes necessarily vary depending on context and requirement, but this research found that in the OSINT field, a dynamic approach of short-term contracts enabling the 'swapping out' of suppliers is necessary to stay in tune with the latest innovations and capitalise on fleeting opportunities.[138] This is particularly important considering that purchased datasets are only as good as those providing the filtration – many

[134] Author interview with A2, 11 January 2022; author interview with HMG4, 27 January 2022.

[135] Author interview with HMG4, 27 January 2022; author interview with A1, 7 January 2022.

[136] Author interview with HMG6, 19 January 2022.

[137] Author interview with LE2, 25 January 2022.

[138] Author interview with A2, 11 January 2022; author interview with HMG1, 25 January 2022.

different products have claims to uniqueness, but in reality access similar datasets and offer similar packages.[139]

To protect against such risks, it is essential to have specialist commercial officers with the adequate level of expertise in the data and technology field, so that they can work closely with suppliers to ensure the data being provided is secure and accurate.[140] The security aspect of this includes concerns about where the servers of the supplier may be based, how and where their data is sourced, the nature of their auditing process, their legal compliance, and relevant data governance regimes.[141] Once these standards are deemed to have been met, internal assessments will be conducted to accommodate the use of a new tool.[142] Commercial officers with a more generalist background may struggle to be effective in the environment described above, so focusing investment on training at this level is vital.[143]

Despite this, there are encouraging developments across the fields of defence, national security and technological innovation. Government-led initiatives such as jHub,[144] the National Security and Innovation Exchange,[145] the Accelerated Capability Environment and the National Security Strategic Investment Fund[146] are demonstrating success in strengthening partnerships across sectors, enhancing capability development and establishing a dual-use technology ecosystem within the UK.

Partnerships with External Stakeholders

The matter of collaboration between different departments and agencies within the government umbrella was addressed earlier in this paper, but thornier issues exist in the context of the UK government's relationships with non-government stakeholders. This encompasses both the commercial sector and civil society organisations.

With respect to the former, concerns about security and information leakage are prominent. For example, using unclassified cloud architecture could theoretically present adversaries with the chance to discern national security priorities by monitoring the type of queries being made on datasets.[147] These queries are usually classified, although there is growing anticipation about

[139] Author interview with LE4, 27 January 2022.

[140] Author interview with HMG3, 25 January 2022.

[141] Author interview with LE4, 27 January 2022.

[142] *Ibid*.

[143] Author interview with HMG3, 25 January 2022.

[144] jHub, <https://www.gov.uk/government/organisations/jhub-defence-innovation>, accessed 11 May 2022.

[145] National Security Technology and Innovation Exchange, <https://www.gov.uk/government/organisations/national-security-technology-and-innovation-exchange>, accessed 11 May 2022.

[146] British Business Bank, 'National Security Strategic Investment Fund', <https://www.british-business-bank.co.uk/national-security-strategic-investment-fund>, accessed 11 May 2022.

[147] Harding, 'Move Over JARVIS, Meet OSCAR'.

the role that privacy-enhancing technologies could play in giving non-government partners more clarity and granularity on national security problem sets.

In any case, the security apparatus is likely to favour traditional defence contractors rather than small and medium-sized enterprises, which are less likely to have the resources to meet the demands and timelines imposed by the national security community.[148] To avoid perpetuating this lock-in effect into the future, evolution of security accreditation systems in ways that enable more agile and dynamic partnering will be necessary.[149]

Data governance and compliance requirements are another consideration, where the potential of new tools provided by the commercial sector is often not realised due to a low risk appetite among government stakeholders.[150] This is explored further in the next chapter.

Regarding potential collaboration between government and civil society organisations, two significant barriers emerged during this research. On one hand, the direction of information flow in this type of partnership can be one-way; on occasions where open source investigators or academics offer their insight to a government stakeholder, there is no clear path for reciprocation or additional context.[151] Some of the reasons for this are perfectly intelligible, but it is nonetheless clear that a relationship where one party takes without giving back is unlikely to be sustainable. Feedback loops which continually improve analysis are a key feature of the online open source community, and this effect is greatly limited in the context of government collaboration.

Another pressing issue is the reputational risk inherent in this type of collaboration to both parties. From the point of view of a civil society organisation, being seen by the public to be working closely with the defence or intelligence community may damage their credibility or neutrality. The actors that these organisations often work to expose could also seek to undermine the credibility of an investigation by way of its association with a government actor. While some investigations carried out by civil society organisations may provide information that is useful to the national security community, it is not borne out of an objective to do so.[152] If there are threats to life, there are obligations incumbent on a responsible citizen to engage with the relevant authorities. However, the purposes of the activities conducted by government and civil society parties are divergent, and collaboration may not always be desirable.[153]

[148] Author interview with IND3, 18 January 2022.

[149] Some encouraging steps have been made. For example, in October 2020, the UK government announced the launch of the Vetting Transformation Programme which hopes to improve agility and efficiency. See Government Security Profession, 'We Are Transforming Vetting', news story, 29 October 2020, <https://www.gov.uk/government/news/we-are-transforming-vetting>, accessed 12 May 2022.

[150] Author interview with LE4, 27 January 2022.

[151] Author interview with A3, 11 January 2022; author interview with CS1, 10 January 2022.

[152] Author interview with CS2, 17 January 2022.

[153] Author interview with A4, 19 January 2022.

From a government stakeholder's point of view, the priority is to supply accurate information to decision-makers; the quality of this information must be balanced with operational risk. Where partnerships are expanded, operations are exposed to future risk.[154] An assessment of the 'who', 'what', 'when', 'where' and 'how' of the partnership, as well as checks to ensure that whatever is being shared is current and reliable, would be necessary.[155] As will be explored next, the legal boundaries that government stakeholders must abide by are different to those for non-government stakeholders, and this can critically affect any prospect of cross-sector collaboration.

[154] Author interview with HMG4, 27 January 2022.

[155] Author interview with A5, 24 January 2022; author interview with A6, 24 January 2022.

IV. The Legal, Ethical and Policy Context

THIS CHAPTER OFFERS four main observations on the interplay between legislation and policy in the context of PAI and considers the ethical and pragmatic dilemmas of revealing investigative approaches in the most granular detail.

Legislative Background and Applications

The use of PAI and OSINT in the UK security context is governed by a combination of primary legislation, statutory instruments and internal departmental policies. However, there remains a lack of clear guidance on how these regulatory and policy requirements should be operationalised in practice. The research identified four main findings in relation to the interaction between legal requirements and OSINT policy.

Disconnect Between 'Law' and 'Policy'

There is a discrepancy between the restrictions placed on the national security community by legislation, the interpretations of that legislation in the form of internal policy guiding day-to-day activities and the interpretations of that policy which are themselves baked into organisational culture.[156] While legislation is generally well understood, the translation of this into internal policy and the reasons underlying that are less so, creating areas of uncertainty and encouraging a precautionary principle where practitioners lean against using certain methods or datasets for fear of compliance breaches.[157]

For this reason, better clarity is required on two fronts: first, on the ways that treatment of PAI differs from the treatment of sensitive information in legal and regulatory terms; and second, the circumstances under which collection of PAI would require a RIPA or IPA authority. The five-year review of the IPA 2016 presents an important opportunity to address these issues and encourage an environment where individuals ask what can be done safely and legally rather than what cannot be done.

Differences in Legal Thresholds Across Sectors and Departments

Unlike the private sector, government agencies are subject to additional restrictions governing their collection and use of publicly available data. Law enforcement and intelligence agencies are required to apply for warrants (via RIPA 2000 and IPA 2016) to access certain types of PAI that

[156] Author interview with HMG1, 25 January 2022; author interview with HMG4, 27 January 2022.

[157] Author interview with HMG5, 14 January 2022.

private and third-sector stakeholders may access and use freely (subject to the DPA 2018). While this can largely be explained by the different nature of the relationship between citizen and state compared with the private sector and civil society, it remains the case that the need for approval for relatively minor actions can deter potentially valuable queries and interactions with PAI.[158]

Differences across public sector stakeholders also exist. What is deemed acceptable activity for a central government department differs from a law enforcement agency, which differs from an intelligence agency.[159] This is inherently linked to the level of granularity that these stakeholders must achieve and the type of decision-making it is informing. For example, a government PAI monitoring service will primarily be interested in high-level strategic themes; a police officer must have enough detail in their information to be able to arrest and charge specific individuals; and a military commander must have an extremely high level of confidence in their information before engaging in direct conflict.[160]

Changes to Privacy Regulation

Some social media platforms are taking action in the name of user privacy to restrict the amount of information that third parties can view or access.[161] There have been instances of legal action being instigated against those responsible for popular online OSINT tools utilising access to social media platforms, contributing to the closure of those tools.[162] What is deemed open source today may, in the near future, be deemed a closed source due to regulation that is either self-imposed by companies or directed by governments.[163] This could significantly affect the ‘art of the possible’ in open source investigation, particularly for non-government parties who depend on unfettered access to open source data.

Relevance of Necessity and Proportionality Principles

The Human Rights Act 1998 requires that any activity that interferes with an individual’s right to privacy must be carried out only as necessary for an agency’s statutory functions and subject to the required ongoing human rights proportionality assessment. This would include collection and analysis of data that reveals sensitive information about an individual, whether obtained through intrusive techniques or otherwise.

These principles are particularly relevant given that open source methodologies are largely less intrusive than the covert alternative. If information can be found via a publicly accessible search engine

[158] Author interview with A1, 7 January 2022.

[159] Author interview with HMG5, 14 January 2022.

[160] *Ibid.*

[161] Author interview with A3, 11 January 2022.

[162] Anjuli R K Shere, ‘Now You [Don’t] See Me: How Have New Legislation and Changing Public Awareness of the UK Surveillance State Impacted OSINT Investigations?’, *Journal of Cyber Policy* (Vol. 5, No. 3, 2020), pp. 429–48.

[163] Author interview with A4, 19 January 2022.

instead of more intrusive covert techniques, then there is a legal duty to use the less intrusive option (under the human rights proportionality principle).[164] Re-emphasising this is essential in justifying greater attention and resourcing for open source techniques in the national security community.

The Impacts of Greater Transparency: Risk vs Reward?

The growth of open source methods and investigations has encouraged a debate on transparency – namely, whether the detailed publication of techniques used within investigations has provided advantages to threat actors, thereby making it harder to sustain successes into the future. For example, it is well-established that the world's militaries have been briefed to be far more careful with the digital footprints they leave behind in response to some of the granular reporting published by civil society organisations.[165] In practice, this can have the effect of reducing the amount of 'low-hanging fruit' in open source investigation. There are also instances where law enforcement capabilities and limitations are exposed during a publicised investigation. An example of this would be differentiating between specific online platforms that the police can and cannot intercept – the very fact of interception is not secret but detailing the 'who' and the 'where' can undermine tradecraft.[166] This brings to light the careful balance to be struck between protecting operational capability while ensuring foreseeability of the law. However, transparency and openness are non-negotiable tenets of open source tradecraft, and the net advantage of publicising methodologies and information is likely to outweigh those concerns.[167]

This raises an interesting contrast: government stakeholders are concerned about revealing operational and tradecraft capabilities, while the opposite is true for civil society stakeholders who risk having their credibility undermined without maximum transparency.[168] This speaks to a potentially foundational difference between traditional approaches to intelligence and the modern open source alternatives: while the former revolves around gaining information advantage among adversaries and groups, the latter attempts to solve problems with all participants at the table leveraging similar sources of information.[169] Replicability and iterative development are the principles on which the open source community draws its trustworthiness, both to ensure that analysis is rigorous and the information and techniques used to form that analysis can be recreated by others.[170] In this way, what may be viewed by one party as an ethical issue – risking future investigations and by extension public safety by being very granular – may be viewed by another as a pragmatic one – ensuring continual improvement in open source innovation and by extension public safety in the long run.

[164] *Ibid*.

[165] See, for example, *BBC News*, 'Russia Bans Smartphones for Soldiers Over Social Media Fears', 20 February 2019. Author interview with CS3, 19 January 2022.

[166] Author interview with A2, 11 January 2022.

[167] Author interview with A4, 19 January 2022.

[168] Author interview with CS1, 10 January 2022.

[169] *Ibid.*

[170] Author interview with A3, 11 January 2022.

V. Summary of Findings and Recommendations

CHAPTER I HIGHLIGHTED the increasing prominence of PAI and OSINT in the present-day intelligence environment and outlined the factors driving that change. Chapter II explained the current and emerging capabilities in the field, while Chapter III explored how tradecraft, cultural, resourcing and partnership factors are hindering progress. Chapter IV highlighted some key observations in the existing legal, ethical and policy context. This chapter provides a summary of the paper's main findings and policy recommendations.

First, **there is a fragmentation of open source investigation activity within government which leads to varying definitions, interpretations and perceptions of PAI and OSINT**. This makes the challenge of professionalising OSINT tradecraft and achieving cohesion and standardisation across government more difficult.

The most recently published government definitions of OSINT are several years old. Given the pace of change in this field, these should be revisited and an updated outlook should be provided across the UK government.

- **Recommendation A**: The Information and Data Exchange (INDEX) platform – by virtue of being a shared space for users to find, connect and promote open source material – will provide a mechanism for the open source community to connect across teams and share best practice. Departments should support INDEX's roll out and put in place mechanisms to build and sustain this cross-government capability.
- **Recommendation B**: The Cabinet Office should take the lead in actively promoting the professional exploitation of PAI in all-source assessment – through its Professional Head of Intelligence Assessment – with the aim of elevating the status of OSINT as a 'core' intelligence discipline and ensuring it is fully integrated into analysis products for senior decision-makers.

Second, while some of the concerns surrounding the veracity of OSINT are overplayed, **there is an absence of a wider, standardised infrastructure in support of PAI analysis in the national security community**. The structures in place to analyse secret information – as well as curate, package and make it accessible to the right people – have a longer tradition and are more labour intensive than what is routinely applied to PAI. This feeds into preferential treatment and a reliance on classified intelligence and internal datasets, and risks a situation where an OSINT product is treated differently despite being reliable on its own terms. It is essential that existing tools and skillsets within government agencies are supplemented by a training offer that establishes OSINT as a core intelligence skill rather than a specialist one. Further research is also needed to establish the appropriateness of using evaluation techniques meant for classified information on OSINT.

- **Recommendation A**: The College of National Security, the Intelligence Assessment Academy and the MoD Open Source Centre of Excellence should lead on developing a generalist open source investigation curriculum for practitioners to open up current training capability across departments. This should be supported by ongoing review and updating of professional OSINT standards across the security community.
- **Recommendation B**: For more specific training needs, where practitioners rely on internal training, the Cabinet Office should explore how relevant modules can be shared securely across organisations and establish secondment opportunities for practitioners to acquire specialist experience from outside their usual domain. These responsibilities could sit with the proposed tradecraft school within the College of National Security.

Third, **there is friction between legislation that is imposed on the national security community, the interpretation of that legislation in the form of internal policy and the subsequent interpretations of that policy manifested in organisational culture**. This creates uncertainty and encourages a potentially counterproductive precautionary principle regarding the use of certain open source datasets.

- **Recommendation**: The Home Office and the National Police Chiefs' Council should issue internal guidance for government and policing respectively, clarifying how the treatment of PAI differs from the treatment of sensitive information in legal and regulatory terms, and the circumstances under which collection of PAI would require a RIPA or IPA authority.

Fourth, **PAI is getting cheaper and becoming more available, while the costs (in the broadest sense) of secret collection are increasing**. This is a cost equation which should be at the forefront of policymakers' minds. Investment in OSINT capabilities will enable deeper and more transparent engagement with the public on national security topics and allow the national security community to engage with foreign actors in the information domain more effectively. However, PAI collection and secret collection can also be complements in economic terms. As the value of insight available from PAI increases – and becomes available to a broader range of state and non-state actors – the value of insight into genuine secrets also increases as investment in secret capabilities are focused more efficiently in pursuit of asymmetric information advantages. Focusing a disproportionate amount of OSINT training resource on analysts within the UK intelligence community may be counterproductive considering the existing costs of maintaining their secret capabilities.

- **Recommendation**: The allocation of the UK government's investment in PAI collection and analysis should be efficiently weighted towards departments and agencies who do not already bear the costs of maintaining secret capabilities and infrastructure. However, the UK intelligence community should be adequately equipped to partner with and integrate information from OSINT specialists across government.

Fifth, **PAI and OSINT provide new opportunities enabling the national security community to be more open and transparent with the public and with other partners, both domestically and internationally**. The ability to cover secret insight with publicly available insight is raising the level of detail and therefore the informed nature of public debate, while also removing plausible deniability from adversaries. Currently, these insights largely reach the public via private media briefings, one-off statements from policy officials, or individual government posts on social media during a crisis – there is no platform which gathers, orders and disseminates these insights to the public in a coherent fashion.

- **Recommendation A**: As part of their wider humanitarian response efforts, the FCDO and Defence Intelligence should jointly establish a platform dedicated to publishing timely briefs of open source insights into a digestible format for the public. This would be driven by clear policy objectives relating to public and strategic communications.
- **Recommendation B**: The UK government should commission a public survey to ensure that this platform builds on lessons learned from public engagement during the Russian invasion of Ukraine.

Sixth, collaboration between government stakeholders and civil society stakeholders in this field is problematic due to reputational and legal considerations. However, **one of the competitive advantages of government is its ability to collaborate and, as such, targeted collaboration across government itself needs to be better supported**.

- **Recommendation**: To facilitate a more systematic approach to government collaboration, the National Security Secretariat and the Joint Intelligence Organisation within the Cabinet Office should:
    - Ensure clear tasking and prioritisation for OSINT analysis, alongside the processes in place to prioritise secret collection.
    - Advocate for programme funding which encourages departments to bid jointly for shared OSINT analysis capabilities, and alignment in accreditation policies which enable OSINT tools to be rolled out across multiple departments at once.

Finally, in some cases, individuals in government will be aware of and able to develop the best technical solutions, but be hindered in progressing those by certain permissions and a lack of empowerment. **Where solutions need to be sourced externally, a dynamic, fluid approach to procurement of PAI and OSINT tools is essential**, but must be complemented by due diligence to ensure what is being supplied to government stakeholders is appropriate**.** Many different products have claims to uniqueness but use similar datasets and offer similar capabilities – specialist commercial officers are needed to establish the viability of potential suppliers, where their data or servers are hosted, and the nature of their auditing and compliance processes. Furthermore, the security apparatus is likely to favour traditional defence contractors rather than small and medium-sized enterprises, risking an undesirable lock-in effect. In both cases, it is difficult for non-government parties to achieve clarity and granularity on national security problem sets due to security concerns.

- **Recommendation**: All government departments with a role in procuring tools and datasets leveraging PAI for intelligence purposes should invest in building a cadre of commercial officers with specialised expertise in the data and technology field. This should be part of a longer-term effort to evolve security accreditation systems and utilise the potential of privacy-enhancing technologies to enable more agile partnering with non-government parties.

Conclusion

THE RUSSIAN INVASION of Ukraine – and subsequent reporting and analysis across publicly available sources – emphasises the findings and recommendations of this research. There is a significant opportunity before the UK government: to encourage analysts to think creatively about the use of PAI in investigations and embrace the higher threshold of secrecy.

This paper has also outlined the range of data sources and techniques – including AI/ML, satellite imagery, news and social media content and other commercial and third-party datasets – that can help leverage PAI for OSINT outcomes. The rate of technological innovation and levels of accessibility are likely to continue to increase rapidly across all these areas. As such, internal development of roles and skillsets will be essential to keep pace and maximise these opportunities for the public good.

Three barriers to progress in the leveraging of PAI in the national security sphere were outlined and consist of: tradecraft barriers; cultural barriers; and resourcing and partnership barriers. Differences were also highlighted in the legal, policy and ethical contexts.

The recommendations outlined in this paper should open new avenues for further exploration. They are a positive 'call to arms' to encourage best practice across a wide range of stakeholders and outline next steps. If fully adopted, they will make a tangible difference to the way that PAI and OSINT can be best leveraged across the UK security community.

About the Authors

Ardi Janjeva is a Research Fellow at RUSI. His research on technology and national security spans four main themes: technology-enabled threats; technology-driven intelligence innovation; technology-based geostrategic alliances and competition; and technology-dependent economic resilience.

Alexander Harris is a Research Associate in the Defence and Security (D&S) programme in The Alan Turing Institute, working across a research portfolio exploring the application of data science and AI to a range of national security and defence challenges.

Joe Byrne is a Research Fellow in RUSI's Open Source Intelligence and Analysis team.