Guidelines for Public Interest OSINT Investigations

ObSINT

Publication date
March 2023

Version
1.0

Table of content

Introduction and Legitimacy **3**
Chapter One: Principles **4**
Chapter Two: Public Interest **6**
Chapter Three: Methodology **7**
3.1 Design 7
3.2 Data collection & preservation 8
3.3 Data analysis 9
Chapter Four: Outputs **11**
4.1. Support to the community 12
4.2. Follow-up 13
Chapter Five: General Work Practices **14**
5.1 A diverse, safe and caring work environment 14
5.2 Accountability 15

Introduction and Legitimacy

In recent years, online investigations using open-source data have flourished. Inspired by the work of Bellingcat and other global leaders, "OSINT" has become increasingly seen as a fundamental part of the research and media landscape and has generated interest both online and offline .

Dr Colette Cuijpers has wisely said, "the fact that data are openly available does not mean that they can be processed without regard to legal and ethical standards. Put in other words, the mere fact that data are publicly available does not imply an absence of restrictions to researching them"[1].

With a constantly developing Open-Source Intelligence ("OSINT") community using a diverse range of approaches and methodologies, both practitioners and users of OSINT have identified the need for a practical framework of ethical principles and good practices to guide public-facing OSINT work.

With the support of a European Union ("EU") Project aiming to establish a European Fact-Checking Standards Network, this group of public-interest oriented OSINT practitioners established a working group to develop this set of Guidelines. The Guidelines build on existing frameworks, such as the Berkeley Protocol on Digital Open Source Investigations for investigating violations of international law. They are designed to be inclusive of the diversity of fields in which OSINT researchers work and provide practical examples of standards and good practices that can be adopted by individuals and organisations, whatever their size and makeup. They are designed primarily for those working on public-facing investigations conducted to fulfil public-interest objectives (as discussed in Chapter 2), although certain elements may also be of assistance to those conducting private, client-facing investigations.

These Guidelines are designed by a small group of practitioners as a starting point. It is hoped that they will be strengthened and expanded with the support of the wider OSINT community; feedback and challenge is warmly welcomed.

[1] Dr Colette Cuijpers, "Legal aspects of open source intelligence – Results of the VIRTUOSO project", https://www.sciencedirect.com/science/article/abs/pii/S0267364913001647

Chapter One: Principles

The Working Group identified five overarching principles that relate to all stages of the public interest, public-facing open-source investigation process. These principles flow through every chapter of these guidelines.

Accuracy

Investigative **processes should be objective and reliable**, with an aim to build trust in the research, the findings and the people working on them. **The research should be clear, self-explanatory** and written in a way that makes it possible for others to evaluate and replicate.

Community

Researchers should keep in mind that **we are part of a wider community** that includes past and future research, a diverse set of stakeholders, and a wide audience. **Forging positive community relationships and giving back to the community are crucial to sustaining a more diverse and reliable industry**. Due regard for the safety and well-being of team members, data subjects and audiences of OSINT outputs is also essential for the protection of this community.

Diversity

Diverse inputs and inclusive participation in OSINT research and investigative processes; a reflective approach to identifying research gaps and limitations and; willingness to seek continuous improvement from a diverse range of stakeholders are necessary to create robust products and reduce risks of doing harm. Such practices are likely to consider multiple perspectives, strengthen the quality of OSINT outputs and minimise risks of bias and unintended consequences.

Accountability

Being **transparent, open and accountable** for the processes followed and actions taken within an open-source investigation, recognising our responsibilities towards both data subjects and audiences and the potential impacts of our work.

Balance and Responsibility

Striving for the appropriate **balance between technical capability and best serving the public interest, focusing on what should be done over what technically can be done**. Organisations should seek to **minimise the risks of causing harm to individuals and the society at large and safeguard fundamental rights**. For instance, to balance the right to privacy of data subjects with the right of the public to have access to information.

Chapter Two: Public Interest

Public interest is not universally defined, and definitions may vary under law in different countries. There are often competing definitions of the public interest and trade-offs between different objectives which an organisation will need to navigate throughout the course of its work.

Public interest in the context of OSINT and journalism is generally understood to involve revealing information that is conducive to the common good and welfare of the general public, for example to expose corruption, crime and wrong-doing; to hold malicious actors accountable; and to ensure that the public has access to reliable information so as to make informed decisions. It is distinct from 'what the public is interested in', focusing on public interest topics, which, if left unexposed and unaddressed, would jeopardise the democratic fabric of society and negatively affect people's wellbeing.

Organisations conducting public-facing OSINT work should be clear how their research is serving the public interest. Every operation should have a clear statement on how and why it contributes to the public interest and the impact it intends to achieve. Organisations should regularly review how their operations are contributing to the public interest throughout the research process, using this as a handrail to guide decision-making.

Examples of good practices:

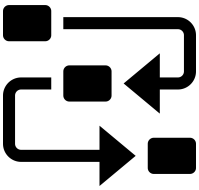
1. **Strengths, Weaknesses, Opportunities and Threat (SWOT) analysis for public interest** - An assessment of the public interest motivation for research, analysing the strengths and weaknesses of your argument(s) for how and why your research serves the public interest. This exercise should also assess opportunities to strengthen the benefits of the research, risks and trade-offs between different objectives. This should be periodically reviewed to ensure your work remains true to the public interest principles underpinning it.

2. **Risk Assessment** - To identify the potential unwanted consequences of the research, and preliminary mitigation strategies which can be developed during the research methods stage. It may be helpful to refer to the breakout scale[2], and the risk of amplifying information which could cause public harm.

3. **Public Interest Statement** - Set out the specific public interest objectives of the research and which can act as a guideline throughout the course of the investigation.

[2] Ben Nimmo, "The Breakout Scale: Measuring the impact of influence operations", https://www.brookings.edu/research/the-breakout-scale-measuring-the-impact-of-influence-operations/

Chapter Three: Methodology



The diversity of the open-source intelligence community generates a diversity of methodologies - there is not a single methodological process that will apply across all forms of research. This chapter sets out a general set of principles to guide the development of bespoke research methodologies to support OSINT investigations.

Organisations should strive to achieve the highest standards of objectivity, transparency, replicability and accuracy within the constraints of the resources available to them. Research methodologies should be clear about methodological limitations, the scope of data/sources collected and used and make this information easily accessible to the reader. Research methodologies will need to balance the priorities of serving the public interest and respecting individuals' fundamental rights. Organisations should regularly re-evaluate their process throughout their research and maintain an accurate record of all steps taken to reach their conclusions.

3.1 Design

Research design is an ongoing process that begins at the outset of the research and is continually refined throughout. This should clearly define the objectives of the research and the process, tools and methods used to identify, collect, analyse, preserve and verify data, including how data samples will be constructed. It should consider limitations, including data or knowledge gaps, risks of bias and identify mitigation strategies to tackle the same. This includes clearly acknowledging what data and data sources are in and out of scope, and the potential implications of this for research findings. It should also consider risks to data subjects and team members and apply appropriate measures to mitigate these.

Examples of good practices:

1. **Research plan** - A live document that sets out the objectives and details steps taken to collect, analyse and preserve information;

2. **Risk assessment** - Conducted at the outset of research to understand and mitigate risks to team members and data subjects, taking into account security, privacy, psychological welfare and legal aspects;

3. **Research skills** - Consideration of what technical expertise, diversity, linguistic and cultural understanding is needed and assembly of a team that meets this to the best of your abilities;

4. **Data limitation assessment** - Set out gaps, limitations or potential biases in the underlying dataset, the steps taken to mitigate these and their implications for findings in the report.

3.2 Data collection & preservation

The right to privacy is a fundamental human right[3]. Researchers will need to adhere to data protection laws, for example the European General Data Protection Regulation (GDPR), as applicable in their country. Data collection needs to strike a balance between the public interest purpose of the research and the individuals fundamental rights.

Organisations may need to take steps to protect the security of their researchers, such as the use of sockpuppet accounts. However, data collection should not be based on deception.

Data collection should seek to minimise the amount of data collected - particularly personal data - and limit it to what is needed to meet defined research objectives. It should include a process of data cleansing to detect, correct or remove corrupt, incomplete or inaccurate records from a dataset. It should also consider scope for data pseudonymisation and anonymisation at the earliest feasible opportunity to avoid unnecessary additional exposure of personal data.

Data should be securely stored and safely archived to ensure the replicability of the research and minimise the risk of a data breach.

Some investigations combine open-source methods with other research methods, such as interviews to provide additional context. In these cases, it is important to be clear about which sources were used to draw conclusions, and the extent to which that has been verified by investigators.

Examples of good practices:

1. **Data protection policy** - A statement that sets out how your organisation protects personal data. It normally consists of a set of roles and responsibilities, rules and procedures on how personal data is handled within your organisation to ensure compliance with data protection laws. It also includes a register of data processing used inside the organisation.

[3] Universal Declaration of Human Rights Article 12, The Universal Declaration of Human Rights (UDHR). New York: United Nations General Assembly, 1948.

2. **Data Protection Impact Assessment (DPIA)** - A DPIA is a process designed to help systematically analyse, identify and minimise the data protection risks of a particular project or research plan. Under the GDPR, there are particular contexts in which you must conduct a DPIA. Further guidance on this can be found at the European Commission's website[4].

3. **Information Technology (IT) Security Policy** - Set out the approach, rules and procedures to ensure the secure processing and handling of all information and data held by an organisation and protect it from both internal and external cyber and other security threats. IT Security requires preservation of confidentiality (protecting unauthorised disclosure of data); integrity (unauthorised or accidental modification of data) and; accessibility (data is available to only those who need it, when they need it).

4. **Data pseudonymisation:** A process to protect privacy rights and prevent the re-identification of data subjects, organisations should implement data pseudonymisation[5] or data anonymisation processes.

5. **Long-term archiving tools:** Organisations should plan for secure, future-proof archiving solutions that would enable replicability of their research analysis while protecting the integrity of data. This could include a combination of local and cloud copies of the data. Access to archives should be regulated to individuals with a legitimate interest in consulting the data.

3.3 Data analysis

Data analysis should be objective, accurate and allow for replicability. It should consider the entirety of the dataset and not cherry-pick evidence in order to support research hypotheses. Data analysis includes acknowledging bias and limitation of tools and measures used to process data and establish findings. It also includes providing an honest confidence level of the findings obtained during data processing. If the confidence level of the data analysis is not good enough, additional measures should be taken to provide a solid assessment. Findings should be backed by the use of multiple sources as far as possible in order to ensure information is accurate and verified.

[4] European Commission, "When is a Data Protection Impact Assessment (DPIA) required?", https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/when-data-protection-impact-assessment-dpia-required_en

[5] European Data Protection Supervisor, "Pseudonymous data: processing personal data while mitigating risks", https://edps.europa.eu/press-publications/press-news/blog/pseudonymous-data-processing-personal-data-while-mitigating_en

Examples of good practices:

1. **Research Logs** - Used by analysts to register possible dilemmas and bias during the research process and which can be periodically reviewed to identify implications for research findings.

2. **Confidence Levels** - Design a set of confidence indicators which grade the strength and quality of the evidence underpinning each finding and which are used consistently throughout the research analysis.

3. **Peer Review** - Implement an external or internal peer-review process to genuinely analyse the data and findings; identify weaknesses, gaps or alternative explanations and; make suggestions on how to improve. Ensuring peer review from diverse perspectives e.g. cultural as well as technical, can help further improve robustness of findings.

4. **Technical Stack** - Create a technical research document detailing the software, scripts and other technical elements used to collect, analyse and filter the data. This document should include technical limitations and bias encountered during the use of the tools.

Chapter Four: Outputs

Outputs should be accessible and comprehensible for the target audience and presented in the clearest way possible.

The main output should be editorially independent and focus on investigation results, accurately presenting findings in precise, objective and non-emotive language. Relevant contextual factors should be considered, and disclaimers and methodological definitions made clear. Technical, methodological and other editorial decisions should be accessible via supporting documents to ensure that the reader has all the necessary information to replicate the investigation and understand how and why certain conclusions were reached. Language and terminologies used should reflect the level of confidence achieved in data analysis, especially in listing conclusions and when assumptions and attributions are made. Wherever possible and relevant for the public interest, supporting evidence should be provided. Where this is not possible, for example for security or privacy reasons, there should be clear justification for how conclusions were reached. Outputs should credit sources of data - including relevant information about the reliability of the source (contextualisation) - as long as it is safe for the source to do so.

For media publications, relevant journalistic deontology standards such as the Charter of Munich may also apply.

Examples of good practices:

1. **Public interest assessment of publication/privacy reviews** - In order to strike the balance between sharing references to back up the conclusions made and individual rights, organisations should reflect on restricting inputs, data and information to what is absolutely necessary to the public interest. This could be conducted through a privacy review of the final written output and a public interest statement accompanying the output detailing the assessment made.

2. **Provide supporting evidence -** Wherever possible, relying on primary sources for every claim, statement or assumption made in the analysis. Set out relevant evidence that supports the claim or analysis, as well as relevant information which appears to undermine it and data gaps. An additional step could involve the application of confidence levels to the evidence.

3. **Methodological Annexures -** Or supporting documents which detail the methodological process used.

4. **Disclaimers** - Set out contextual information relevant to the research findings, such as possible conflicts of interest and developments subsequent to data collection which might affect future findings or any significant updates to the findings. It can be helpful to clearly state the timeframes within which an investigation is conducted and data cut-off points, as well as the date the report is published.

5. **Consider your target audience -** Adapt the type and design of output to best suit the target audience. Ensure multimedia, visual or graphical representations of data are clearly labelled and described, including an explanation of any manipulation of the data or images undertaken in the process of analysing or presenting the data.

6. **Editorial guidelines -** Create guidelines which provide a framework on the style, language and content of research outputs in line with an organisation's values and standards.

4.1. Support to the community

OSINT as a field derives its strength from both the sharing of information by affected populations, such as civilians posting footage on social media, and the sharing of knowledge among practitioners. As such, it is strongly encouraged to share knowledge and insight gained through investigations back to both affected communities and the wider OSINT community and its stakeholders, where legally permissible and safe to do so.. As far as possible, this should include sharing relevant data, tools, methodologies and findings, either publicly or with key stakeholders. It also includes reporting threats and vulnerabilities to relevant actors, platforms, regulators and affected users among others. Organisations doing OSINT work should try and support the development, accessibility and sustainability of open source tools.

Examples of good practices:

1. **Facilitate access to knowledge -** Build and maintaining open libraries and repositories of data.

2. **Contribute** - To common glossary and create common manuals of methodologies.

3. **Make available expertise -** Share experience with peers, for example through learning events and publications.

4. **Participate or provide training -** Feedback to smaller organisations, especially those with limited resources.

4.2. Follow-up

The application of these principles should continue past the dissemination of findings. A reasonable effort should be made to follow up on the impact of the publication. The relevance and efficacy of the methodology followed and processes used to serve the public interest should be evaluated, in order to continually improve work practices.

Examples of good practices:

1. **Impact Assessment** - To understand the extent to which the research achieved its public interest objectives, and to identify any unintended consequences it may have had.

2. **Retrospective Review** - Conduct a review with the relevant stakeholders to understand the lessons learned what worked well, key challenges and recommendations for continuous improvement of the investigative process, including how to develop more robust methodologies and tools.

3. **Version Control Procedures** - To ensure a clear and transparent process for integrating new information or refining the findings of an investigation.

4. **Archiving Policy** - Define and implement a solid policy for how to securely and accurately store and preserve data for future use, including creating and enforcing data retention policies;

5. **Implement design features for future readers -** To inform readers about the original context of publication (date, etc.) and potential content obsolescence. This could apply to images, text, titles and snippets displayed on social media.

Chapter Five: General Work Practices

The guidance set out in this document is intended to create resilience and trust in and between organisations, especially about their objectives, working methodologies and risk mitigation processes.

Organisations should strive towards continuously improving work culture, for which they are held accountable through internal and external inputs. Work practices should reflect a commitment to the "open source" philosophy, including a diverse, safe and caring work environment and an active and positive participation in the OSINT community.

5.1 A diverse, safe and caring work environment

Organisations should design and nurture an inclusive, safe and caring work culture and environment. Best efforts should be made to protect staff well-being and the safety of people involved in research and dissemination. This should include promoting a culture of openness and support around mental health, recognising the risk of vicarious trauma involved in OSINT research.

Objectivity and internal accountability should be fostered in diversity, among staff and organisational structure, as well as in research topics, target audiences, methods and techniques used.

Examples of good practices:

1. **Diversify tools and techniques** - To try and avoid reliance on a single approach and invest time and resources in new accurate methodologies and tools.

2. **Recruitment and Working Practices** - which promote diversity of researchers and staff, for example through flexible working practices and foster **work environment** focussed on the principles of diversity, equality and inclusion.

3. **Cybersecurity and Online Safety Training** - Providing staff with adequate support and learning material to improve their protection against cybersecurity or other online risks.

4. **Prevention and Mitigation of risks for staff -** Make regular risk assessments regarding safety and protection of staff, including risks to mental health. Build and regularly update safeguarding policies; including policies on how to handle illegal and/or traumatic content and content involving minors.

5. **Employee Assistance Programme -** Establish a programme that includes practical support for staff with mental health and other issues.

6. **Cultivate feedback, knowledge and experience sharing -** Build regular work processes (internal and external), such as lesson learning reviews or professional development events.

7. **Whistleblowing Policy -** Create a policy that provides a safe and protected means for staff members or contractors to raise concerns about organisational policy, practice or potential wrong-doing.

5.2 Accountability

Just as organisations doing public-facing OSINT work seek to hold those in power to account, they are themselves also accountable for the actions they take in carrying out their operations. This means being responsible for their research methodologies, the data they hold and the public outputs they disseminate.

OSINT organisations should publicly share a clear mission statement.

Examples of good practices:

1. **Accessibility and Responsiveness -** Making organisational contact details transparent and accessible. Respond in due time to good faith communications. Consider and integrate external contributions when reviewing data analysis.

2. **Acknowledgement** - Give credit to sources who have contributed or helped author the piece where safe to do so.

3. **Corrections Policy** - Update your publication in a timely and honest manner when required.

4. **Self assessment** - Build processes that guarantee regular self evaluation, for example wrap-up sessions at the end of each investigation which consider strengths, weaknesses and lessons learnt from the research and publication process .

5. **Transparency** - Take reasonable efforts to provide relevant organisational transparency, for example by publishing a mission statement and/or a statement of methods.

ObSINT

Guidelines for Public Interest OSINT Investigations

This project has been conducted in the framework of the European Fact-Checking Standards Network Project. EFCSN is supported by the European Union under the 2020 work programme on the financing of Pilot Projects and Preparatory Actions in the field of "Communications Networks, Content and Technology"