Please Read Carefully Detailed Formatting Guidelines for Preparing Your HICSS Final Paper with Author Names

Franz Kayser
ESG
franz.kayser@esg.de

Thomas Mayer
ESG
thomas3.mayer@esg.de

Michael Bücker
FH Münster – University of Applied Sciences
michael.buecker@fh-muenster.de

Abstract

Open Source Intelligence (OSINT) is currently experiencing an intensive discourse, heightened since the Russian invasion of Ukraine. However, despite numerous attempts at standardized definitions, the intelligence discipline remains ambiguous. This paper introduces a practice-validated OSINT trend radar, categorizing technologies by maturity, intelligence cycle phase, and use case. Serving as a profound knowledge base and tool for identifying research gaps, the radar emerges from a structured design process. Sixty studies underwent categorization and validation through expert interviews, revealing the absence of a comprehensive, automated third-generation OSINT system in Germany. Technological gaps, especially in the planning, direction, dissemination, and integration phases, are evident. Although intelligent support technologies were identified, practical implementation lags behind theory. The human factor therefore remains central to the OSINT process. Future research should thus prioritize developing applications for underserved phases, probing reasons for limited widespread implementattion of proven applications, with emphasis on legal, ethical, political, and social parameters.

1. Introduction

OSINT is a currently more debated research field than ever before. Obtaining intelligence from publicly available data [1] has become undeniably important since the Russian invasion of Ukraine in 2022. In this context, real-time analysis, especially of social media, has revealed highly relevant insights [2,3]. However, OSINT itself is not a new technique [4,5], but one of the oldest intelligence disciplines [6]. Despite numerous attempts to define OSINT [vgl. z.B.: 7–9], the concept of intelligent analysis remains controversial to this day [10–12]. This is not the least because every definition of OSINT is subject to advances in computer and data science, which are continuously developing improvements in (intelligent) collection and analysis capabilities [11,12]. Moreover, this is accompanied by numerous novel communication channels, which have led to a veritable ”information explosion” [9,1,7]. Today’s problem therefore no longer lies in acquiring information, but in processing its sheer volume [6]. In addition, technologies originally restricted to defense and intelligence services are now accessible to the general public, primarily via the Internet [7,12]. The understanding of intelligence thus changed completely [13]. At the same time, the increasing speed of development makes it almost impossible to predict the future shape of OSINT and its consequences [14].

To date, there has been a lack of decisive, fundamental scientific publications to pervade the opacity of the subject area [15] and address its rapid developments [vgl. 11,12]. In particular, there is a lack of current studies that reveal the actual technologies behind OSINT in detail and determine their characteristics. The question of whether ”third generation” OSINT systems in the sense of robust, self-managing solutions [4,8] already exist has therefore not yet been clarified [11,8,9]. Moreover, the majority of studies focus exclusively on analyzing the OSINT trend area ”cyber security” [cf. e.g.: 7,9,4]. The literature thus missed to cover the topic in its entirety. Important application scenarios (”use cases”) of OSINT have therefore remained unconsidered in research to date [cf. also 16,13,11]. In addition, supplementary qualitative field research is absent, for example in the form of expert interviews, which contrast theory with the corresponding practical implementation. Although OSINT has a major impact on topics such as security and defense, there is a lack of insight into these sectors [4,15]. This paper is hence dedicated to answer the research question: *How can the current trends in OSINT in the form of the technologies used and their characteristics, in particular the maturity level and the use case, be presented in a trend radar and validated by experts within the security sector?*

The aim in answering the research question is to identify the technologies used in OSINT applications and to present them systematically in a trend radar, according to their characteristics. Through expert interviews, the identified trends will then be validated and compared with the common practical ”reality”. In this way, a well-founded knowledge base will be compiled, and existing research gaps of practical relevance will be identified. This will enable a coordinated exploration of the research field. The structure of this study thereby follows the iterative ”Design Science Research Model” (DSRM) [17], an open research paradigm for the creation of an innovative artifact [18]. Within this framework, the relevant literature on OSINT will first be analyzed and classified using a systematic literature review [19]. Based on this, the OSINT technologies and their characteristics identified will be visualized in the form of a trend radar. The radar will then be validated using systematizing interviews [20] conducted with experts in the security sector. The interviews will then finally be evaluated using a qualitative content analysis [21].

2. Theoretical Background

This chapter sets out the theoretical background required to understand the paper. First, the terms Open Source Intelligence (OSINT), Open Source Data (OSD) and Open Source Information (OSIF) as well as intelligence are defined. Subsequently, the previous papers with a similar research question are presented.

2.1. Open Source Intelligence (OSINT)

Despite numerous attempts to define OSINT, it remains controversial in the literature to this day [12]. This is mainly due to the fact that OSINT is largely dependent on the developments of advancing computer and data science. Its domain is continuously expanding as a result of the resulting improved recording and analysis possibilities [16,11,12]. In addition, advances in information and communication technology and the associated new means of communication have made OSINT an increasingly complex discipline [16,14,22,12]. One of the earliest and still frequently referenced definitions (see also 1) originates from the handbook published by NATO in 2001. OSINT according to this definition is information that has been deliberately discovered, discriminated, distilled, and disseminated to a select audience, [...], in order to address a specific question. OSINT, [...] also applies the proven process of intelligence to the broad diversity of open sources of information and creates intelligence [23]. However, today the discipline is no longer seen as a purely governmental matter. Private research institutions and organizations outside the security sector [24,25] are also massively driving the development of such systems, e.g. for competitive analyses or marketing activities [16,13,11]. The focus is thereby shifting to developing OSINT into a robust, self-managing solution and fully automating the process from data collection to analysis [4,8,9].

2.2. Open Source Data (OSD)

The starting point for all OSINT activities lies in data. Data forms the basis of the analysis and the conclusions derived from it [26]. In this context, OSD refers to non-processed [1], general raw data that is openly available [6] and legally and ethically accessible [5,23]. In practice, sources whose access requires additional effort [24] or must be acquired commercially [12,23,27] are not excluded. At the same time, however, the addition ”legally and ethically acceptable” implies that not all publicly accessible data should automatically be treated as OSD [28,5].

2.3. Open Source Information (OSIF)

OSD are of little use on their own and only become relevant to intelligence when they are aggregated [12]. Before intelligence can be obtained from them, the data must therefore be subjected to a preparation process that includes a certain amount of filtering, validation and summarization [1,23]. The result of this organization of the data [5] is referred to as OSINF. It provides the basis for the resulting knowledge creation [1,5].

2.4. Intelligenz and Intelligencecycle

The core task of OSINT is to generate intelligence [7,13] from the condensed information in the sense of a well-founded basis for decision-making [29,30]. The generation process of such an intelligence product is also referred to synonymously as the intelligence cycle [15,31]. It represents the central element of every intelligence discipline, regardless of the underlying sources or their accessibility [32,13]. The representation of the process as a cycle [33] goes back to the CIA’s ”Factbook” published in 1987 [31]. To this day [34], the CIA defines the process as consisting of five successive phases: Planning and Direction, Collection, Processing, Analysis and Production, and Dissemination [31]. The link between the phases is that the result of the preceding phase serves as input for the subsequent phase [27,35]. Similarly, the product of one cycle serves as a starting point for refinement in the next [13,26]. Furthermore, the individual phases within the

cycle are not linear, but are continuously iterated due to the fulfillment of previous requirements and new demands [27]. Accordingly, the JCS supplemented the Intelligence Cycle in 2013 with an evaluation and feedback process that is subject to all phases [27] (see Figure 1). However, to improve the representation of external influences or the assignment of responsibilities [36-38], numerous other modifications can be found in the literature today [32,24]. The Intelligence Cycle should therefore be seen less as a guideline and more as an informal coordination element, with a partly very intuitive [29] interpretation [7]. Thus, as a proven instrument, the original CIA cycle continues to provide a good starting point for managers and analysts [36,23].

The planning and direction phase combines the identification, definition and prioritization of the requirements for the cycle and the intelligence product. It is also responsible for developing the activities required to achieve these [28] as well as monitoring their coordinated implementation [27,28]. The phase is therefore responsible for controlling the entire cycle [31]. The survey phase refers to the collection of raw data [31]. The core of this phase consists of the iterative repetition of research [23] in order to make the query more precise with each run [8]. The processing and utilization phase is concerned with condensing these data volumes into valuable and action-relevant information for further processing [33,8,27]. Analysis and production refers to the synthesis of the information obtained into a user-oriented, timely and accurate intelligence product [28,33,7]. The final phase consists of handing over the finished product to the ”customer” in a usable form [31,12,28]. In doing so, it is important to adhere to the deadlines and reduce the product to the relevant content [28], while at the same time ensuring its completeness (see also 36). Evaluation and feedback are not to be regarded as individual phases within the cycle, but take place continuously throughout the entire process. The aim is to achieve progressive optimization [33,23,27].

2.5. Previous studies

In the literature, eight previous, publicly accessible literature reviews can be identified on the topic of OSINT. In 2017, Dos Passos [1] highlighted the benefits of incorporating OSINT into everyday life in his literature review. In particular, he emphasized how big data and data science can make the decision-making process more useful and effective. Pastor-Galindo et al. then published further studies in 2019 [4] and 2020 [8]. In these, they described the current state of OSINT with a focus on services and techniques to improve cybersecurity. They also examined the role of OSINT in the public sphere of governments, using Spain as an example. Moreover, they are credited with the first and only approach to a rudimentary inventory of trends in OSINT. Their findings show that OSINT is used in the three spheres of social opinion and sentiment analysis, cybercrime and organized crime, as well as cybersecurity and cyberdefence. Two further literature reviews were published in 2020. In their research, García Lozano et al [39] identified methods and techniques for computer-assisted veracity assessment of public information. Herrera-Cubides et al [15] investigated how the production of research and educational materials in the field of OSINT has developed. They came to the conclusion that the number of publications is lower compared to other trending topics. Finally, in 2021, Yogish and Krishna [9] explored the state of implementation and use of AI (Artificial Intelligence) technologies in the context of cybersecurity. The result of this study showed that machine learning (ML), pattern recognition and natural language processing (NLP) can and will simplify OSINT in view of increasing data volumes and are already being used in isolated cases. In the following year, Hwang et al. [7] conducted a further literature review on security trends and investigated security threats and cybercriminality in the context of OSINT misuse. Based on this, they proposed preventive security measures. In a literature review published in 2023, Ghioni et al. [11] then examined the political, ethical, legal and social implications of OSINT in combination with AI. They came to the conclusion that there is still no framework for discussing these. They also found that third-generation OSINT is still in its early stages and that human components cannot yet be replaced.

3. Graphics/Images

All images must be embedded in your document or included with your submission as individual source files. The type of graphics you include will affect the quality and size of your paper on the electronic document disc. In general, the use of vector graphics such as those produced by most presentation and drawing packages can be used without concern and is encouraged.

- Resolution: 600 dpi
- Color Images: Bicubic Downsampling at 300dpi
- Compression for Color Images: JPEG/Medium Quality

- Grayscale Images: Bicubic Downsampling at 300dpi
- Compression for Grayscale Images: JPEG/Medium Quality
- Monochrome Images: Bicubic Downsampling at 600dpi
- Compression for Monochrome Images: CCITT Group 4

If your paper contains many large images they will be down-sampled to reduce their size during the conversion process. However, the automated process used will not always produce the best image, and you are encouraged to perform this yourself on an image by image basis. The use of bitmapped images such as those produced when a photograph is scanned requires significant storage space and must be used with care.

4. Main text

Type your main text in 10-point Times, single-spaced. Do not use double-spacing. All paragraphs should be indented 1/4 inch (approximately 0.5 cm). Be sure your text is fully justified—that is, flush left and flush right. Please do not place any additional blank lines between paragraphs.

Figure and table captions should be 9-point boldface Helvetica (or a similar sans-serif font). Callouts should be 9-point non-boldface Helvetica. Initially capitalize only the first word of each figure caption and table title. Figures and tables must be numbered separately. For example: "Figure 1. Database contexts", "Table 1. Input data". Figure captions are to be centered below the figures. Table titles are to be centered above the tables.

Figure 1. Sample figure with caption.

5. First-order headings

For example, "1. Introduction", should be Times 12-point boldface, initially capitalized, flush left, with one 12-point blank line before, and one blank line after. Use a period (".") after the heading number, not a colon.

5.1. Second-order headings

As in this heading, they should be Times 11-point boldface, initially capitalized, flush left, with one blank line before, and one after.

5.1.1. Third-order headings. Third-order headings, as in this paragraph, are discouraged. However, if you must use them, use 10-point Times, boldface, initially capitalized, flush left, followed by a period and your text on the same line.

6. Footnotes

Use footnotes sparingly and place them at the bottom of the column on the page on which they are referenced. Use Times New Roman 8-point type, single-spaced. To help your readers, try to avoid using footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence). asdöfj

7. References

List and number all bibliographical references in 9-point Times, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example [1, 2] and [2]. Where appropriate, include the name(s) of editors of referenced books.

References

[1] C. D. Jones, A. B. Smith, , and E. F. Roberts, "Book title," *Journal*, pp. 20–25, Date.

[2] A. B. Smith, C. D. Jones, and E. F. Roberts, "Book title," *Journal*, pp. 7–15, Date.