applied sciences

MDPI

Article

Open-Source Intelligence Educational Resources: A Visual Perspective Analysis

Jhon Francined Herrera-Cubides [1,*], Paulo Alonso Gaona-García [1] and Salvador Sánchez-Alonso [2]

[1] Faculty of Engineering, Universidad Distrital Francisco José de Caldas, Bogotá 11021-110231588, Colombia; pagaonag@udistrital.edu.co

[2] Department of Computer Science, Universidad de Alcalá de Henares, 28805 Madrid, Spain; salvador.sanchez@uah.es

* Correspondence: jfherrerac@udistrital.edu.co

Received: 15 July 2020; Accepted: 14 August 2020; Published: 29 October 2020

Abstract: Given the growing application of open-source intelligence (OSINT), which has facilitated fast decision-making, this study aims to explore how research and educational material production in OSINT has evolved. For this analysis, two OSINT material sources are examined: the research dissemination databases and educational resources repositories. Considering that web information may or may not be publicly available, web Scraping and querying web interface strategies are used to metadata extraction. Finally, we suggest a findings hierarchical classification for the metadata retrieval results. Our main results: (1) Google Scholar and NewsBank are the centralizing axes of OSINT publications; (2) OSINT presents a broad development in the areas of defense and security; thus, presenting itself a promising future; (3) it is necessary both to generate educational resources that complement OSINT training processes and documenting existing resources with a metadata structure defined for this purpose; (4) pay increased attention to the last stages of the OSINT process, to use this knowledge in more assertive ways. This study allows guiding the researchers to the current state of research and education in OSINT and promotes a useful metadata description to make resources accessible and reusable in the educational environment.

Keywords: OSINT; Open-source intelligence; open data; educational resources; open sources; metadata; resources

1. Introduction

Currently, there is a diversification of services offered on the web, which has led to an evolution of a growing mass of digital data [1]. These data can be accessed by Application Programming Interfaces (APIs), or different services, applications, etc. As an example, people who make use of services available on the web to register sites of interest, travel, political or religious affiliations, photos, among other sources of disclosure of a clearly public nature, can be found. However, not everyone is aware that a large proportion of this information is publicly exposed and can be used by individuals or organizations with different purposes [2]. This means that all information published on social networks, discussion forums, and group chats, among other sources, is free and accessible to anyone, considering the restrictions that may apply [3]. Nevertheless, even when large amounts of data are found, these are, themselves, considered as unevaluated material obtained from any source. Yet, when such data are elaborated and treated, acquiring meaning and utility, they are transformed into information. Furthermore, if experience, understanding, and codification are added to this, such information becomes knowledge. Once this is made available to a person interested in the purpose of helping the

decision-making process, intelligence takes place [4–6]. The activities of gathering and correlating such information through the use of tools is called Open-source intelligence (OSINT) [6].

Although this work of collecting and correlating information is not a recent activity, some more contemporary definitions of OSINT can be given.

"Unclassified information that has been deliberately discovered, discriminated, distilled and disseminated to a select audience to address a specific question" [7,8].

"Intelligence that is produced from publicly available information and is obtained, used and disseminated in a timely manner to an appropriate audience for the purpose of responding to a specific intelligence request" US DoD (Department of Defense) [9].

"Collect, process and correlate public information from open data sources such as the media, social networks, forums and blogs, public government data, publications or commercial data" [10].

From the aforementioned definitions, it is possible to identify key aspects in OSINT: (a) open sources, (b) process of obtaining and treatment of information, (c) tools and techniques, and (d) intelligence and decision-making process. With these aspects in mind, when performing an elementary surface web search on Google using the keywords "OSINT" and "Open-source Intelligence", a list of between 1.48 and 684 million related sites is obtained respectively, with information about one or several of the key aspects contrasted in the definitions, which allows and generates a clear interest on the topic of OSINT.

Given the growing need to implement intelligence processes in different fields, some works can be identified as examples: courses on how to apply OSINT, sites teaching how to use tools on open sources or research on the application of OSINT in specific areas such as cyber threats inspection [11,12], as well as disaster management [13], privacy dilemmas [14], among others.

While there is a growing interest in open-source intelligence, which is shown in research such as those carried out by [10,15], an interesting research question—that this research looks at—is how research and educational material production in OSINT has evolved. Most developed works have focused on specific OSINT applications. However, they do not pay attention to justify open-source intelligence from the perspective of OSINT research production and dissemination, which is an important indicator that shows OSINT trends and applications. Furthermore, OSINT educational material production needs to be considered because it shows how OSINT training processes are being supported. Consequently, existing researches have not shown quantitative evidence of work about these indicators in the last 10 years.

The main contributions of this paper are as follows:

- a set of indicators is proposed to support and justify the research and educational materials production in OSINT;
- the useful and complete metadata description and documentation are promoted to make resources more accessible and reusable in the educational environment;
- the design and use of educational material that supports OSINT training processes are promoted.

Briefly, our contribution allows stimulating research and advances within the OSINT ecosystem, both in application domains, as well as in the generation of resources, which can lead to supporting the growth that this strategy has been experiencing.

To conduct this study, a brief exploration of state-of-the-art OSINT is presented. Subsequently, a methodology to address the quantitative collection of information is proposed. This will, consequently, lead to carry out a tabulation and analysis of the results obtained in order to evaluate the interest that the OSINT topic arouses in the academic and research world. Once the survey was completed, some conclusions and future work are drawn. The collection of data required for this research—originated from open sources—was carried out between 1 February and 25 April 2020.

2. State-of-the-Art OSINT

Considering that the main aim of our research is to evaluate how the research dissemination and production of educational material in OSINT have evolved, we see OSINT research dissemination as a vital input to evaluate the behavior of resource openness and availability. Additionally, reviewing the production of OSINT educational materials allows us to verify if educational resources have been produced in this topic and how educational resources are being described (metadata) in these open sources.

For the reasons outlined above, three essential aspects of OSINT are summarized in this section. Firstly, the concept of open-source is reviewed, as the base of the open-source intelligence process, to subsequently address both some tools used in the OSINT process and the availability of information on the web. Finally, some approaches using open-source intelligence will be briefly reviewed.

2.1. OSINT Sources

According to [16], OSINT sources are public sources—independent of whether their content is commercialized or free. They can be documents with any content, in any medium, under any means of transmission or mode of access. OSINT sources are distinguished from other forms of intelligence because they must be legally accessible to the public without violating any copyright or privacy laws [17]. These open-sources include [18–20]:

- government data and public reports, budgets, hearings, telephone directories, press conferences, websites, and speeches;
- professional and academic publications, information acquired from journals, conferences, symposia, academic papers, dissertations, and theses;
- commercial data and images, financial and industrial evaluations, as well as databases;
- grey literature, technical reports, preprints, patents, working documents, commercial documents, unpublished works, and bulletins;
- photos and videos, including metadata;
- geospatial information (e.g., maps and commercial imaging products).

On the other hand, open-source information is not limited to what can be found using the main search engines [21]. In fact, performing a search on any engine produces as a result massive sources of information, which are far from being the only sources given. On the web, there are multiple open-source for different types of searches: videos, images, texts, etc.

2.2. Tools for OSINT

As for tools or applications that allow open-source intelligence, there are applications in different fields. Several specialized software tools include big data software, video analysis, text analysis, visualization tool, cybersecurity, web analysis, and social network analysis. On the web, there are tools for multiple purposes, with valuable information resources and uses in decision-making and measures in different fields, such as [2,3,22–24]: Maltego (security and forensic investigation), Shodan (search engine for hackers), The Harvester (email and domain information), among others.

2.3. Availability of Information

Over the years, the improvement and specialization of the technologies and services available on the web have generated exponential growth in the amount of data available on the network. The increase in active users on social networks [25] is an example of this growth. Apart from promoting the strengthening and expansion of platforms such as TikTok or WhatsApp, social networks have generated large amounts of public data available on the web that can be queried through techniques, such as OSINT [26,27]. On the other hand, the maturity of open data in several countries has improved

the availability of public information [28,29]. Therefore, repositories of this type of information become reliable and official sources for OSINT [10].

Lastly, publicly available information can be queried using standard search engines and can be accessed using standard web browsers [30,31]. Conversely, there is information that needs credentials, password, encryption, or APIs to be accessed [31]. These access requirements define the way that data can be queried.

2.4. Related Works

Regarding OSINT and its different applications, examples of research are identified in areas, such as application of OSINT tools in the private sector and public sector [1], cyber intelligence [32], analysis of opinions on Twitter [33], as well as classification in recruitment processes [34]. There are many other related works such as [20,35], Oryon C Portable, and Facebook Graph Search (project closed in 2019), among others, whose detailed study is beyond the scope of this paper.

On the other hand, [5,10], apart from describing the current state of OSINT by making a comprehensive review of the paradigm focusing on services and techniques that improve the field of cybersecurity, also raises challenges on OSINT, such as (i) automation of capture processes, (ii) improvement of knowledge analysis and extraction processes, (iii) filtering of irrelevant data, among others. Regarding the main trends on OSINT, identified by authors such as [5,10,36,37], are focusing on the defense and security analytics segment (video analytics, reducing network traffic, providing a real-time indication of external threats, detection and prevention of inside threats, and monitoring of suspicious activity in the organization, and forth).

As for the design of OSINT educational material, the literature review did not identify researches related to planning, design (didactic or pedagogical), or implementation of OSINT educational resources. However, materials such as slides, videos, blogs, as forth, that are being used to carry out the transmission of knowledge on OSINT were identified.

The review of these components shows that, on OSINT, some work has been carried out in different knowledge domains with the aim of exploiting open-source data to make decisions or design measures for different behaviors. However, such work does not mention two key factors, which, in fact, give rise to the current study: (a) there has not been research that shows the academic research and dissemination behavior provided by OSINT in recent years, which, in addition to analyzing trends, could provide elements that allow decisions to be made regarding its application and use. (b) Although there is evidence of the development of materials supporting the different OSINT proposals, in many cases, these are not presented as materials that allow their accessibility and reuse for academic and/or training purposes, among other benefits.

3. Materials and Methods

A systematic mapping method was implemented to perform our research based on the construction of classifications and obtaining information on the existing knowledge on a specific topic [38]. This approach allowed us to analyze the source of information in order to identify findings both in the OSINT research dissemination and in the description of OSINT educational materials. Based on this approach, and following the method described by [38], the research design is explained below (Figure 1).

3.1. Definition of Research Questions

To develop this study, according to our approach, two macro-questions were raised to solve with the systematic mapping:

- How has the academic-research and dissemination behavior provided by OSINT evolved?
- How are the educational materials produced by OSINT being described?

Two strategies were developed to work on these questions.

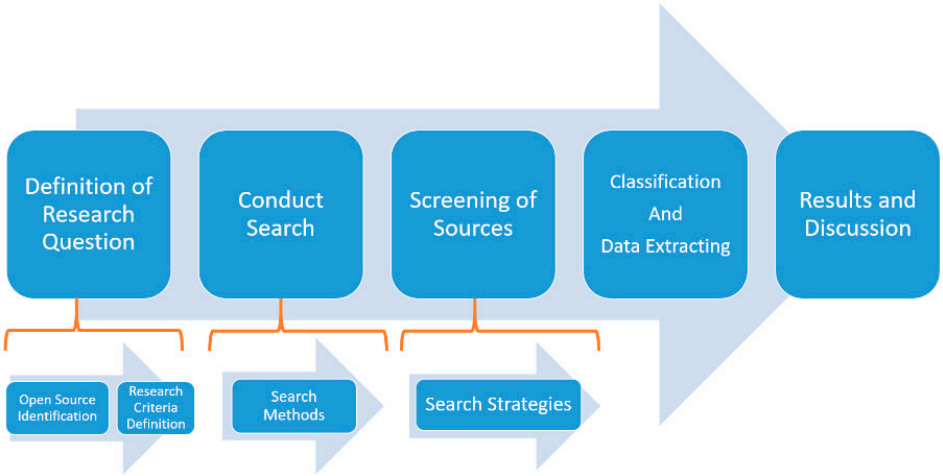

Figure 1. Research design.

3.1.1. Open-Sources Identification

Regarding the research questions, sources that allow identifying research dissemination and educational materials production in OSINT were selected:

- as to research dissemination, sources such as YouTube, Google Scholar, Latindex, SciELO, ScienceDirect, among others, were selected;
- as to educational resources, sources such as Carnegie Mellon University, Eduteka, MIT Open Courseware, Course Hero, among others, were selected;
- as to Massive Open Online Courses (MOOCs) dissemination, sources such as Udemy, Alison, The Open University, among others, were selected.

These sources were selected at random, taking into account their relevance in the academic and research world. The complete list of sources queried can be seen in Appendix A.

3.1.2. Research Criteria Definition

In order to perform the search processes, and considering the number of applications used in OSINT, the filtering of information in the selected sources was proposed in two stages. (a) For the search of resources, the following parameters were set: either in the title, summary or key words should contain the keys "OSINT" or "Open-Source Intelligence". (b) Subareas or sectors of OSINT application were identified in order to detail the findings on OSINT, which are provided by the OSINT body of knowledge:

- security type: human intelligence, content intelligence, dark web analysis, link/network analysis, data analytics, text analytics, artificial intelligence, big data, others;
- technology: big data software, video analytics, text analytics, visualization tool, cyber security, web analysis, social media analysis, others;
- application: military and defense, homeland security, private sector, public sector, national security, education, others;
- intended audience: investors and consultants, security agencies, government organizations, research/consultancy firms, technology solution providers, IT solution providers.

3.2. Conduct the Search: Definition of Search Methods

In order to identify sources that provide open information about OSINT materials, the strategies described below were defined.

In the first step, and in order to identify sources, a general surface web to search videos, documents, sites, and any other OSINT materials (also called OSINT resources), was carried out by running a

query on the Google search engine. For this search, a web scraping that allowed to filter the results by: (a) OSINT subareas, (b) sources that provide OSINT resources, and (c) type of OSINT material, was used. This was made in order to establish the digital format in which OSINT resources are published.

In the second step, and in order to execute a basic deep web search on OSINT resources, a source scan was carried out including sources of educational resources. This exploration was based on the following information retrieval strategies:

a. information retrieval using APIs or web scraping techniques. These sources were chosen because of the availability of their API, as well as for being key sources in terms of provision and dissemination of academic and scientific works;
b. information retrieval using web interfaces from sources. These sources were selected because of their focus on the dissemination of academic and scientific works. However, they do not have an API available for consumption;
c. retrieval of information from Massive Open Online Courses (MOOCs). In addition to academic and scientific distribution, it was relevant to explore the production of educational resources such as MOOCs in the OSINT area, since these types of courses enhance knowledge dissemination by individual organizations with the spirit of sharing and collaboration [39]. This exploration was performed manually since the sources did not have an accessible API or query services to exploit their content;
d. retrieval of information from other repositories. Subsequently, the exploration of sources specifically oriented to educational resources was executed in order to identify the existence of OSINT resources catalogued as Open Educational Resources (OER).

3.3. Screening of Sources: Design of the Search Strategy

For the design of the web scraping, and since Google has blocked robots, a user agent belonging to the Firefox browser that runs on Ubuntu Linux was used. This user agent renders web pages using the Gecko engine [40,41]. However, even in this case, after a certain number of pages, Google classifies these requests as those sent by a robot, so other strategies, such as virtual private network (VPN), had to be used. For this Agent, the following were used as libraries: (a) BeautifulSoup for reading text in HTML format as an object, (b) requests to the Google search engine, (c) Operating System (OS) for managing directory paths, (d) JavaScript Object Notation (JSON) for writing and reading files in JSON format, and (e) sys for handling controlled errors. The type of search used was "Term1" AND "Term2" (Figure 2), forcing the search engine to use the exact word or term, whereas the AND defines that both terms must be present in the results. In our case, term 1 was always the word OSINT, whereas the second term was iterated over the set of predefined areas corresponding to the identified OSINT sub-domains (Figure 3).

```
query = "research firm"

query_edited = query.replace(' ','+')

url = "https://www.google.com/search?client=ubuntu&channel=fs&q=%22osint%22+AND+%22"+query_edited+"%22&ie=utf-8&oe=utf-8"
```

Figure 2. Search criteria.

```
["human intelligence","content intelligence","dark web",
"network analysis","link analysis","data analytics","text analytics",
"artificial intelligence","big data","big data software","video analytics",
"visualization tool","cyber security","web analysis","social media",
"military","defense","homeland security","private sector","public sector",
"national security","education","investor","consultant","security agency",
"security agencies","government organization","it solution provider",
"consultancy firm","research firm","technology solution provider"]
```

Figure 3. Search terms.

For the query of the source APIs, the APIs are accessed through their endpoints. In some cases, you have to use an API key to make authenticated requests to the platform. Each API may query data utilizing a GET request method. You have to carefully read the instructions to use each API. Below are a few examples of API endpoints:

- arXiv API: https://arxiv.org/help/api
- Elsevier Developers: https://dev.elsevier.com/api_docs.html
- Udemy: https://www.udemy.com/developers/affiliate/
- YouTube: https://developers.google.com/youtube/v3/
- Twitter: https://developer.twitter.com/en

3.4. Classification and Data Extracting

In this stage, three categories have been made (Figure 4). The first category corresponds to the sources that do not have OSINT resources published. The second category corresponds to the sources that publish OSINT resources without metadata. Finally, the last category corresponds to the sources that publish OSINT resources with metadata.

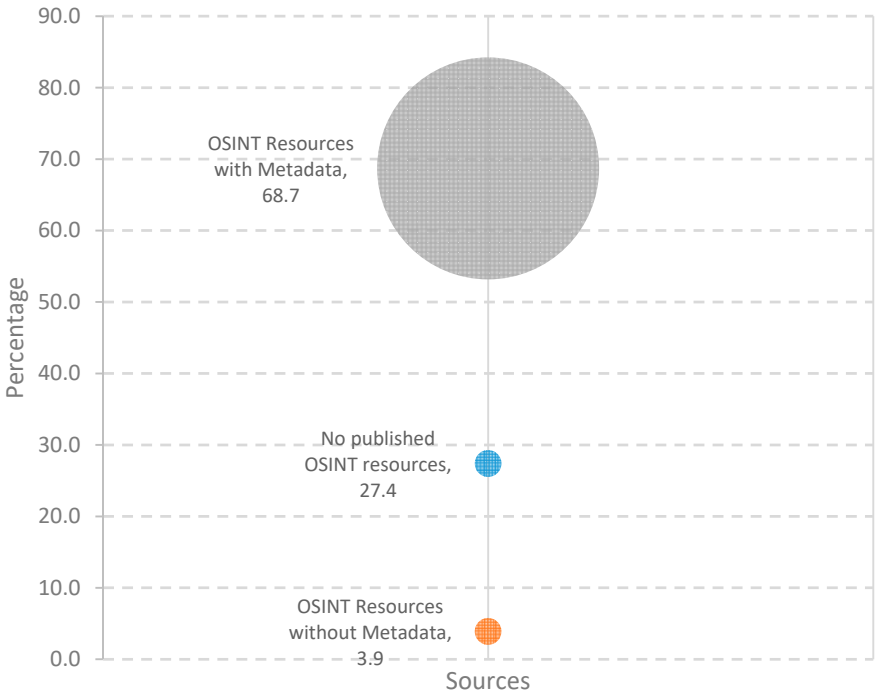

Figure 4. Classification of results.

3.4.1. Sources without OSINT Resources

A total of 27.4% of the sources consulted do not present any kind of OSINT resources (Appendix B). This finding is derived from the specificity of some of the sources consulted (music, varieties and entertainment, etc.). However, the lack of presence of OSINT topics in electronic library projects, such as Latindex or SciELO may be due to factors such as the low development of OSINT in the countries which comprise these projects. Another factor to take into account could be the low interest or scope suggested by this type of means to disseminate the research results in OSINT.

3.4.2. OSINT Resources Identified without Metadata

Sources such as: (a) Booklick, which identifies 230 OSINT resources (journals and book chapters) redirected to the universities supporting the documents; as well as (b) Dialnet, which identifies nine OSINT resources (seven journal articles, two book articles), do not provide a metadata structure to

classify the results found. From this, the need to have an adequate metadata structure that allows not only the description of resources, but also the provision of timely results to the search processes, can be inferred.

Nowadays, on the web, there are both paywall or subscription-based sources, and sources that offer publicly available information. However, it is essential to share metadata information to allow describing, enriching, finding, sharing, and reusing resources. We should publish the metadata necessary for resource identification regardless of the type of services used.

4. Results and Discussion

The results obtained from the exploitation of sources were analyzed using two different approaches: (a) the increase of available materials on OSINT specifically of a didactic or educational nature; and (b) the interest that OSINT arouses in the academic and research world.

4.1. Web Scraping Application Results

The following results were obtained from the Surface through the use of the web scraping:

With regard to the OSINT subareas, the subarea with the most work corresponds to security analysis on topics such as: web Analysis (traffic analysis), video analysis (data, behavior, attitudes, etc.), link analysis (criminal activity, security analysis, etc.) (Figure 5). To do this, products based on big data software are used, which allow the analysis of large amounts of disparate data, such as those provided by the aforementioned subareas. In general terms and seen from different perspectives, it is evident that security, as well as public and government environments, correspond to the areas of greatest interest in terms of the work and application of OSINT.

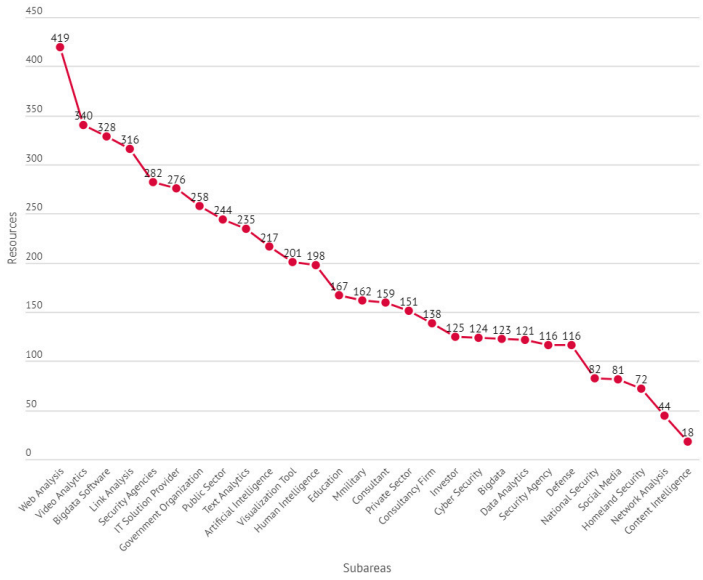

Figure 5. Open-source intelligence (OSINT) subareas identified by the web scraping.

It is important to consider that intelligence allows anticipating both opportunities and risks, the latter being critical factors for the survival of an organization or country. In general terms, security focuses on preventing risks and threats. Therefore, the combination of intelligence and security become key elements given, for example, in the different scenarios present in social, political, and economic environments worldwide.

With regard to the sources that provide OSINT resources, the sources with the highest participation rates correspond to the digitized book service offered by Google, as well as the ResearchGate Academic Collaborative Network. The latter carries out its searches in databases such as PubMed, CiteSeer, arXiv, and the National Aeronautics and Space Administration (NASA) Library, among others (Figure 6).

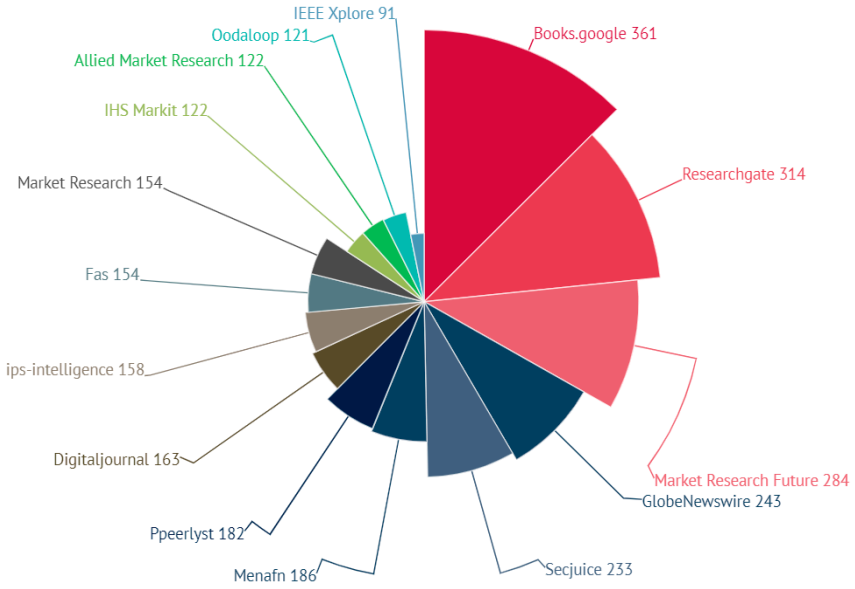

Figure 6. OSINT resource sources identified by the web scraping.

Regarding the type of material given in OSINT, the images and videos pertain to the most used OSINT formats, with 4447 and 1850 resources, respectively. Apart from that, below can also be found the Portable Document Format (PDF), blogs, and Wikis, with 113, 79, and 65 resources, respectively. These results show that in the vast majority of the sources queried, audio-visual media are used as one of the main strategies for the transmission of knowledge regarding the different OSINT topics. This type of behavior stems from the fact that in the process of managing open sources, the configuration and use of the tools are of great importance, for which the audio-visual material becomes the preferred tool, since involving skills, such as attention, application of learning, and understanding.

4.2. Metadata Reported by Sources

In relation to the queries made through the APIs or using web scraping techniques, a wide dissimilarity is observed in both the metadata published by the sources and the metadata that is allowed to be accessed through the services. The latter case is more restricted, as can be seen in the metadata results provided by the platforms for the description of their resources (Figure 7).


```
{'_filled': False,
 'bib': {'abstract': 'This paper introduces the concept of Open Source '
                     'Intelligence (OSINT) as an important component for '
                     'understanding human problem solving in the 21st century. '
                     'OSINT is in many ways the result of changing '
                     'human–information relationships resulting from the '
                     'emergence …',
         'author': 'M Glassman and MJ Kang',
         'eprint': 'http://www.academia.edu/download/59839022/j.chb.2011.11.01420190623-17064-1taekoe.pdf',
         'journal': 'Computers in Human Behavior',
         'publisher': 'Elsevier',
         'title': 'Intelligence in the internet age: The emergence and '
                  'evolution of Open Source Intelligence (OSINT)',
         'url': 'https://www.sciencedirect.com/science/article/pii/S0747563211002585',
         'year': '2012'},
 'source': 'scholar'}
```

```
[{'items': [{'id': 'vBIkHCa4sW0', 'snippet': {'publishedAt': '2020-03-19T13:18:44.000Z', 'defaultAudioLanguage': '
en'}, 'statistics': {'viewCount': '57'}}]}, {'items': [{'id': '4Iq4O3QhUZw', 'snippet': {'publishedAt': '2020-03-1
3T22:05:38.000Z'}, 'statistics': {'viewCount': '23'}}]}, {'items': [{'id': '7Aod-I81fDU', 'snippet': {'publishedAt
': '2020-03-12T23:19:17.000Z'}, 'statistics': {'viewCount': '63'}}]}, {'items': [{'id': 'tRzGiR4DS7w', 'snippet':
{'publishedAt': '2020-03-12T20:16:43.000Z', 'defaultAudioLanguage': 'en'}, 'statistics': {'viewCount': '45'}}]},
{'items': [{'id': 'qLwGwh1W9cs', 'snippet': {'publishedAt': '2020-03-03T07:00:09.000Z', 'defaultAudioLanguage': 'd
e'}, 'statistics': {'viewCount': '32'}}]}, {'items': [{'id': 'uWq6qZaqLig', 'snippet': {'publishedAt': '2020-03-03
T07:00:10.000Z', 'defaultAudioLanguage': 'en'}, 'statistics': {'viewCount': '35'}}]}, {'items': [{'id': '-fntwDgJQ
bU', 'snippet': {'publishedAt': '2020-03-03T03:53:09.000Z', 'defaultAudioLanguage': 'ta'}, 'statistics': {'viewCou
nt': '23'}}]}, {'items': [{'id': 'IBn6_J4aHoE', 'snippet': {'publishedAt': '2020-03-02T14:32:35.000Z', 'defaultAud
ioLanguage': 'en'}, 'statistics': {'viewCount': '54'}}]}, {'items': [{'id': 'SDwumTLLplw', 'snippet': {'publishedA
t': '2020-02-28T11:45:30.000Z', 'defaultAudioLanguage': 'en'}, 'statistics': {'viewCount': '36'}}]}, {'items': [{'
id': 'OCMSzwe04iM', 'snippet': {'publishedAt': '2020-02-28T00:30:00.000Z', 'defaultAudioLanguage': 'en'}, 'statist
ics': {'viewCount': '26'}, 'recordingDetails': {'location': {'latitude': 27.763383, 'longitude': -82.5436722}}}]}]
```


Figure 7. Metadata example: Google Scholar and YouTube.

The limited amount of data provided by the platforms' services restrict different data comparisons, such as data models and controlled vocabularies used to describe their resources, domain tagging, for instance. Although each data source needs specific metadata to describe their resources, there are common metadata that all resources have to manage.

Regarding the description of resources, most sources provide a title, abstract, year, and author metadata, whereas few sources provide metadata of keyword, language, and document type. Those metadata are examples of the mandatory metadata that all data sources should be managed and universally spread.

One of the sources with the greatest problem for its consultation was YouTube, since it allows access to resources through a consultation quota, restricting the search processes. On the other hand, key assignment request responses for searches are not answered in a timely manner.

4.3. Findings on the Identified OSINT Resources

Given the fact that the APIs of the sources consulted do not provide a uniform metadata scheme, the findings identified according to the metadata provided by the source(s) consulted are, therefore, presented below.

4.3.1. OSINT Resources Published

As for the production of OSINT resources in the last 10 years, Google Scholar shows a growing interest in publishing of these types of resources, whereas ScienceDirect tends to decline as a source of such resources (Figure 8). As for Twitter, a rate of 15,750 tweets related to OSINT has been observed so far in 2020. Regarding YouTube, it went from 279 videos in 2019 to 1092 videos in 2020, so far.

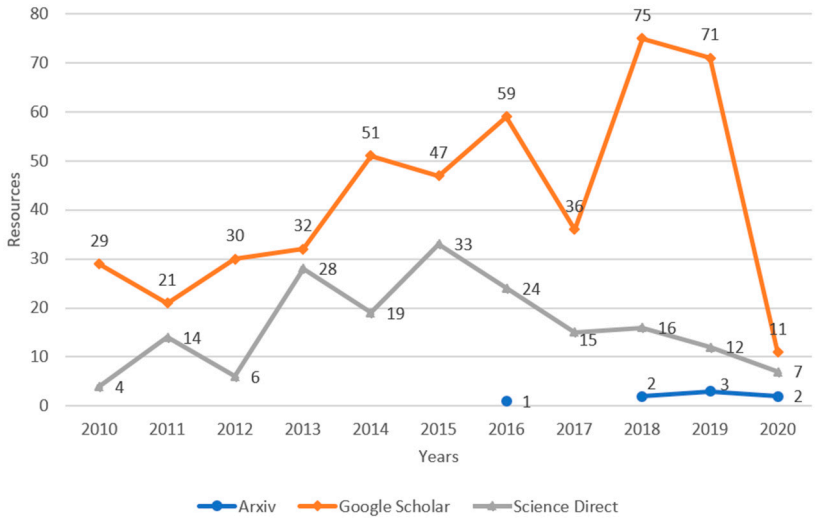

Figure 8. OSINT resources in the last 10 years.

These results make it possible to show what was presented by [42], research work that describes that Google Scholar has the advantage of constantly finding the highest percentage of citations in all areas (93%–96%), well ahead of Scopus (35%–77%) and Web of Science (27%–73%). Additionally, Google Scholar found almost all Web of Science (95%) and Scopus (92%) citations, and most of their citations come from non-journal sources (48%–65%), including theses, books, conference papers, and unpublished materials. For these reason, Google Scholar maintains its position over the use of other tools, considering the fact that it is a free tool. Regarding the OSINT resources published in the Database, Figure 9 presents the evolution in the period 2011–2019.

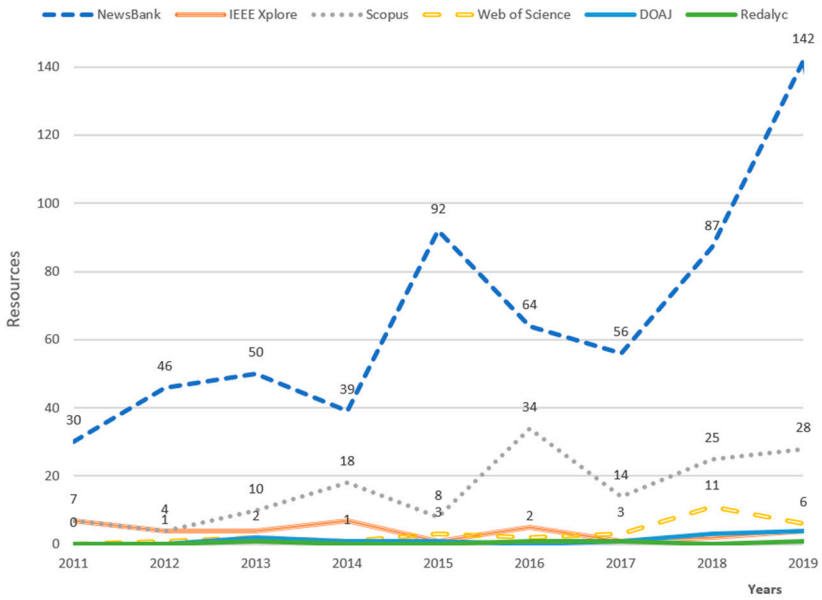

Figure 9. OSINT resources published on Databases.

Among the Databases consulted, NewsBank is the fastest growing database in OSINT publication, followed by Scopus with 80.3% less of OSINT resources published in the last year. NewsBank consolidates information from newspapers, cable news, web editions, blogs, videos, broadcast transcripts, business magazines, periodicals, government documents, and other publications. This factor contributes to agglomerate different types of publications in areas, such as OSINT. Finally, each item in the NewsBank database has key metadata, such as headline, source, date, Lexile/readability, source type, author, standard title, and title as published.

All of the above show that: (a) Google Scholar has a greater preference for publishing OSINT resources in relation to specialized and paid databases such as Scopus or Web of Science; and (b) although NewsBank is a subscription service, it is consolidated as the central axis of the different OSINT publications made on the web.

4.3.2. OSINT Subareas Worked

With regard to the subtopics or areas on which the OSINT resources are developed (Figure 10), the accumulated sources (Google Scholar, Udemy, YouTube and ScienceDirect) allow cyber security to be identified as the area of greatest interest in the work about OSINT. In relation to the areas mentioned in the tweets published so far in 2020, cybersecurity (1350), information and security-infosec (1350), security (900), and cybercrime (900) are the subareas with the highest participation rate in publications about OSINT on Twitter.

In general, it is evident that the concern for security—digital, physical, national, or organizational—has become one of the most worked domains in OSINT. Within this area, the use of link, video, text, and data analyses, etc., can be identified. In certain cases, this allows the detection and the establishment of some type of predictions about behavior, as well as people who may become potential threats to security.

4.3.3. OSINT Publication Languages

With regard to the academic works on OSINT published in Google Scholar and YouTube, English prevails with 619 resources published as the language with the highest dissemination in the publication of resources on OSINT, followed by Spanish (28), British English (19), Korean (15), and Italian (14), respectively. As for YouTube, the highest rate of publication languages of OSINT resources pertain to English (532) and Italian (112), respectively. The aforementioned leads to the conclusion that English is

consolidated as the universal language for the transfer of knowledge regarding OSINT, followed by Italian and Spanish.

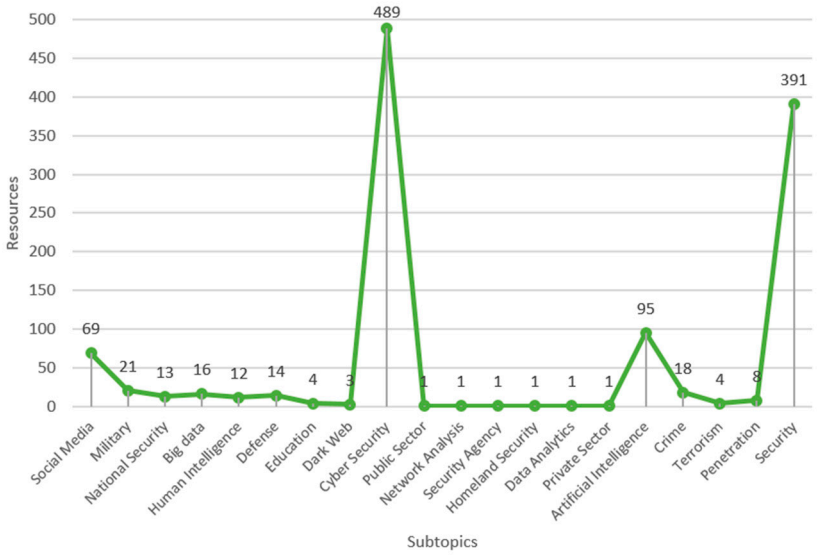

Figure 10. OSINT subtopics.

4.3.4. Country of Origin of OSINT Publications

The databases that report the country of origin of the published resources (NewsBank, Scopus and Elton B. Stephens Company - EBSCO), identify the United States as the country with the largest number of publications on OSINT, for which its publication focus is the cable news, provided by NewsBank (Figure 11).

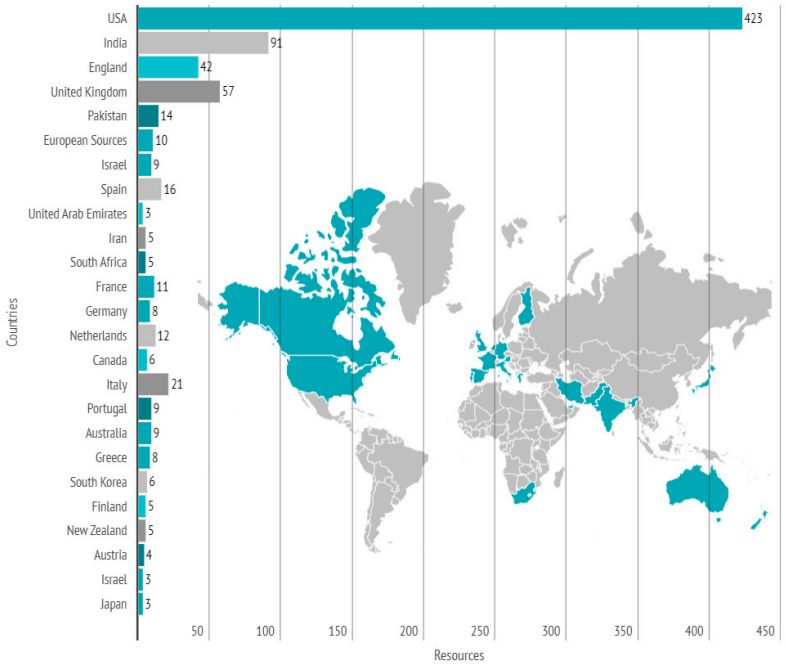

Figure 11. Origin of OSINT resources.

On the other hand, the lack of participation in OSINT publications from Latin American, Central, and East Asian countries (except Japan) is observed in the same queried sources.

4.3.5. Types of Publications of OSINT Resources

Within the databases consulted (NewsBank, Oxford University Press, Sage Journals, Sage Knowledge, ProQuest, Springer, Scopus, Web of Science, EBSCO, and Directory of Open Access Journals—DOAJ), cable news—provided by NewsBank and the research articles—provided by 8 of the 10 databases, as well as book chapters—pertain to the most published OSINT resource types (Figure 12).

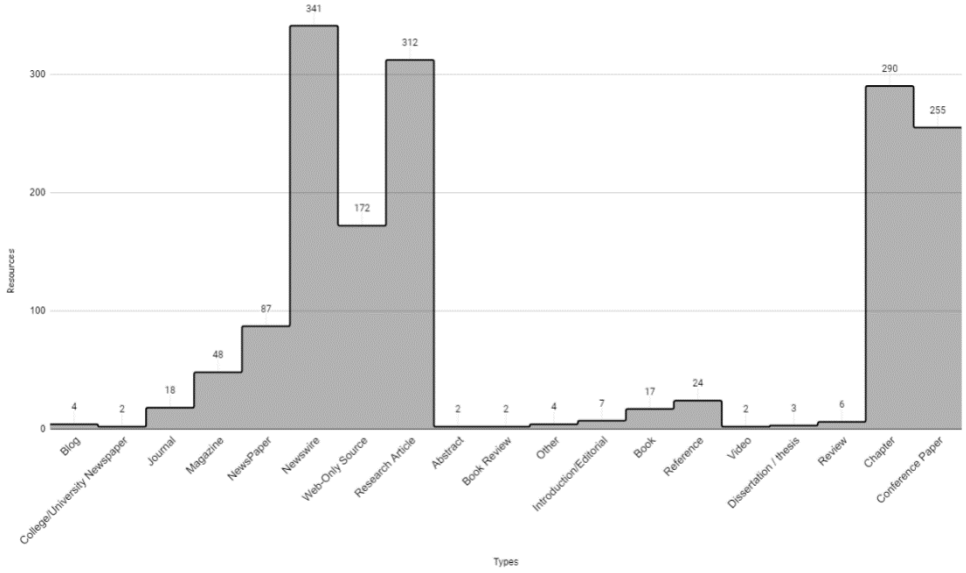

Figure 12. Types of OSINT publications.

4.3.6. Scope of Journals and Conferences in Which OSINT Is Published

Regarding the scope of journals and conferences, the metadata provided by three data sources (Taylor & Francis Group, The Institute of Electrical and Electronics Engineers–IEEE Xplore, and Oxford University Press) were reviewed. In general terms, the scopes with the greatest number of resources pertain to intelligence (100), security (77), social networks (12), science (5), computing (4), economy (3), and medicine (1). Most of the scopes are oriented to issues and challenges that must be addressed by both government and private institutions, especially when making contemporary decisions and policies related to intelligence and security.

4.3.7. General Areas in Which OSINT Is Published

For the review of general categories or subareas in which OSINT resources are published, the metadata provided by seven different databases (Taylor & Francis Group, Sage Journals, Sage Knowledge, Springer, Scopus, Web of Science and Redalyc) was reviewed. According to this review, the Taylor & Francis Group, Springer, Scopus, and Web of Science databases have greater coverage in subareas of OSINT resource publishing. On the other hand, Scopus has the largest number of resources (360) published under different OSINT subareas.

Figure 13 identifies the areas of Computer Science and Politics as well as International Relations as those with the main bibliographic production, showing that OSINT is based on two macro scenarios: (a) in the systemic study to describe and transform information using the application of computer systems; and (b) in its use and application in a global context, considering the existing complex dynamics.

However, given the current conditions of increasing security and defense problems, as well as the usefulness that OSINT has proven to tackle these, as can be seen in [43], which is a European Union Agency for Law Enforcement Training, it is not ruled out that many processes performed in other areas are not fully documented publicly.

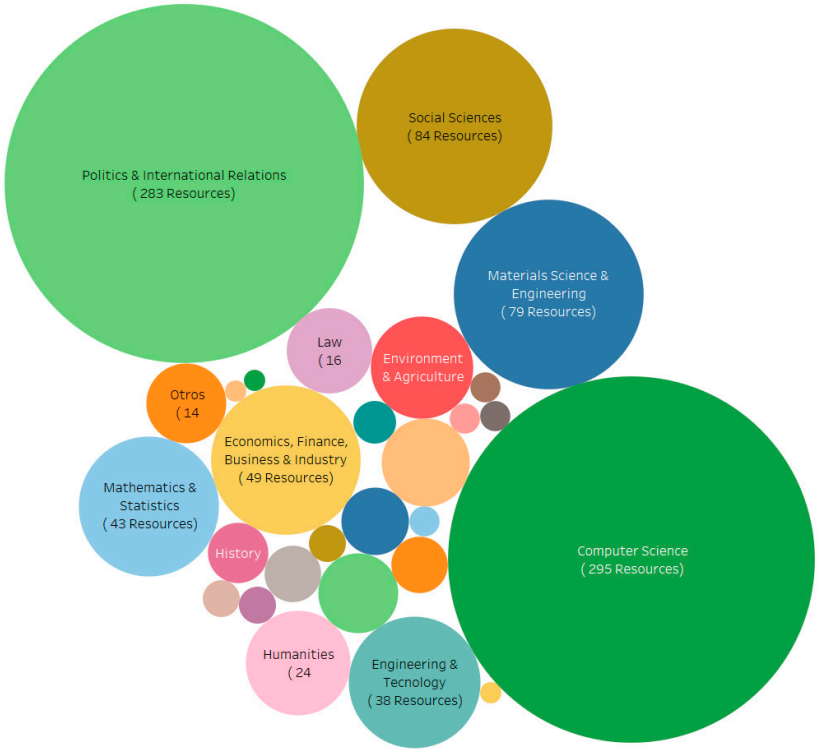

Figure 13. OSINT post categories.

4.3.8. Specific Subareas in Which OSINT Has Been Published

In terms of specific OSINT publication subareas, data protection and management, as well as security and defense (Figure 14) are identified as the main areas of work in OSINT, given that they comprise two major work fronts for performing political, military, scientific, and sociological intelligence, among others. As for the tweets registered in specific sub-topics, there are 450 tweets registered under the dark web and military subareas. The above confirms the impact that OSINT has had in the security area, extending this to security in data management, as well as having a defense approach used against threats, either internal or external.

4.4. Educational Resources and MOOCs OSINT

With regard to the "educational" resources available on OSINT, search engine queries were made with combinations of key words such as (OSINT) (Open source intelligence) and (OER) (Open Educational Resources) (training), considering the use of operators to refine the search, such as those defined in [10]. For each combination of keys, groups of responses that exceed 100,000 results in each query are obtained. In broad terms, these can be identified as: blogs, certifications, courses, tools, projects, videos, podcasts, books, etc., such as:

- video reading: Open-Source Intelligence, author: Clive Best, Joint Research Centre;
- course: Open-Source Intelligence from the American Public University System;
- guide: A Guide to Open-Source Intelligence (OSINT)—Michael Edison Hayden, Columbia Journalism Review;
- workshop: Advanced Web Intelligence (OSINT), Institute for Competitive Intelligence;
- podcast: Technology and Media, Timothy de Black;
- forum: OSINT tools and how you learn how to use them, Peerlyst;
- site: Exploiting Public Information for OSINT, Irfan Shakeel.

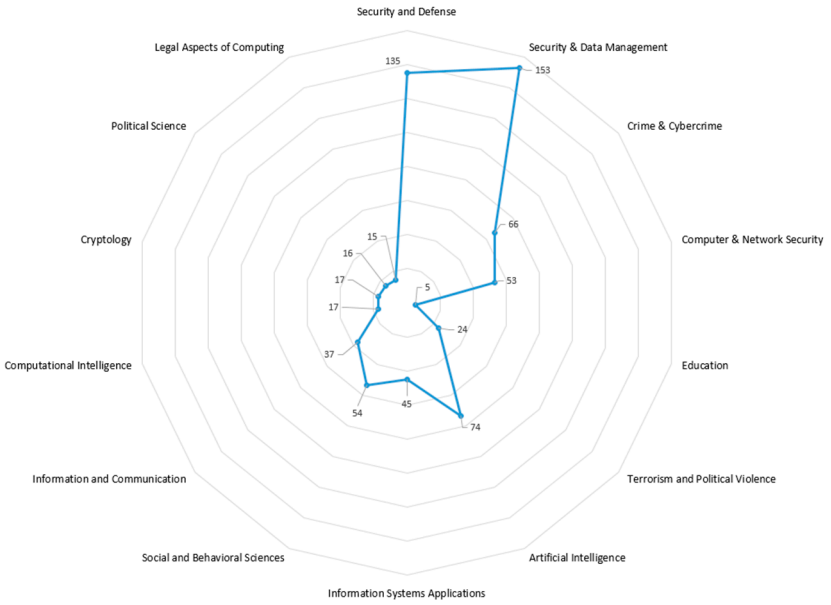

Figure 14. Specific subareas for OSINT publications.

However, even though no OSINT resources are clearly identified as "Educational Resources" within the exploration carried out, those which were identified are susceptible to be used in training processes as support material.

With regard to search engines and MOOC repositories, the query raised about OSINT resources yielded the results shown in Table 1.

Table 1. Exploring Massive Open Online Courses (MOOCs) in repositories.

Repository	MOOCs on OSINT
The Open University	0
Iversity	0
Alison	0
Open Learning	0
Coursera	1 Course: Area: Defense
edX	0
Udemy	18 Courses: Levels: All (25), Beginner (11), Intermediate (6), Expert (1). Language: English (17), Spanish (1) Approaches: Espionage, Social Engineering, Defense, Security.
SANS	1 Course: Area: Cyber Defense

As seen in these results, most MOOCs are redirected to Udemy. MOOCs published in the Udemy platform, are focused on applications, tools, and techniques of OSINT, and Cybersecurity and investigation. On the other hand, Udemy uses a simple set of metadata information to describe their resources (title, author, date, language, description, requirements). Finally, all of these resources are paid courses. In such query, MOOC search engines were also reviewed, giving the following results (Table 2).

Although most MOOCs are described using metadata such as platform, provider, effort, length, language, credentials, and Uniform Resource Identifier (URI); some of these platforms do not manage metadata information that allows describing essential data, such as their educational purpose and skills, or a complete provenance and contributor schemas. That metadata information is so important in the educational domain in order to classify adequately the resources.

Table 2. MOOCs Search Engine Exploration.

MOOC Search Engine	MOOCs on OSINT
Class Central	6 MOOCs that are redirected to Udemy and Coursera
Accredible	0
CourseBuffet	0
Open Education Consortium	0
MOOC List	1 MOOC that is redirected to Udemy
Course Talk	3 MOOCs that are redirected to Udemy and PluralSight
MOOCLab	0
Stanford Online	0
Tutellus	1 MOOC about Hacking
Miriada X	0

Generalizing the results obtained in this section, it is therefore evident that:

a. OSINT material, such as videos, images, etc. has been created and these resources are used in different sites with different purposes as well;
b. although these materials do not have all the characteristics of educational resources, they can be used in this context, thus complementing aspects of a pedagogical and didactic nature;
c. there is a very low rate of resources that are framed within the educational and MOOC context, in relation to the number of courses, certifications and trainings found on the web about OSINT;
d. as for the MOOCs obtained from the queries, it can be identified that most of them are centralized in Udemy and are oriented to cyber security and defense;
e. a common problem identified in this research is metadata. According to [44–46], metadata sounds like one of the most boring things for people; for that reason, they don't care if resources are correctly described and written. This situation generates little information about resources;
f. Most OSINT "educational" resources published on the web are focused on how the topic is defined, what kind of tools and gadgets you have to use, how you have to set up these tools, and how you can use these tools in a specific field. These topics correspond to the firsts stages in the OSINT process. However, topics referring to how to analyses data obtained from OSINT tools, and making decisions based on these analyses, do not have enough resources published on the web.

Additionally, from the results obtained, it is, therefore, evident to generate educational resources and MOOCs complementing the training processes in which the OSINT topic is introduced. However, it is important to document the existing general resources with a metadata structure defined for this purpose, which provides the necessary elements to make them accessible and reusable in the educational environment.

4.5. Other Queried Repositories

Among the sources consulted, those shown in Table 3 present OSINT resources. However, they do not offer APIs or any type of complementary metadata.

This query identifies Course Hero as a potential repository of OSINT resources classified into courses, study documents, study guides, videos, questionnaires and troubleshooting books. However, Course Hero is a proprietary repository in which material from schools and universities around the world can be found. Therefore, this does not limit it to be an open-source tool specifically.

Table 3. Exploration in other Repositories.

Repositories	Resources
Open Yale Courses	0
Carnegie Mellon University	One Resource. Topic: Defense. Type: Paper
Japanese Industrial Standards Committee	0
Open Michigan	0
National Center for Curriculum Development in non-proprietary systems	0
CREA Resources	0
Inter-American Development Bank	0
Eduteka–Icesi University	0
MIT Open Courseware	One Resource. Topic: Open Source Collection. Type: Conference Guide
Polytechnic University of Madrid Open Courseware	0
Europeana	0
Course Hero	1959 Resources

4.6. Mapping the OSINT Application Fields

Briefly, concerning the approaches developed on OSINT, the following results are generalized:

- regarding the literature review, applications of OSINT in several knowledge domains are identified. One of the most worked areas focuses on security and cybercrime. Research such as that carried out by [47–49] are examples of this type of application;
- according to the outcomes collected through web scraping, research in OSINT has been advancing in fields such as security analysis, focusing on topics such as web analysis, video analysis, and link analysis;
- as for OSINT publications, a higher level of publications was identified on topics related to security, data management, and defense;
- concerning the educational resources, Udemy and Course Hero are the major MOOC providers. The MOOCs production is focusing on OSINT life-cycle, cybersecurity, hacking, terrorism, among other fields;
- regarding OSINT trends, the security analytics segment is expected to garner a significant share for 2026. Numerous benefits provided by security analytics have been fueling the growth of this market [48].

These results show that both scientific publications and resource production are aligned with OSINT market trends. However, in addition to the need to convert OSINT materials into educational resources, there is a need to focus on transforming OSINT into a robust and self-managed solution [10].

4.7. Ratio of Total Resources vs. OSINT Resources

As shown in Figures 15 and 16, the participation of OSINT resources in the resource repositories does not represent 1% of their publications. In such figures, it can be seen that even in repositories with a high level of publication of resources such as NewsBank, only 0.00012% of OSINT resources are published. The repository with the highest number of published OSINT resources pertain to Sage Knowledge with 0.14% of its resources.

This shows that, although OSINT has been generating a great impact on subjects such as security and defense, the scope of its publications is not yet robust enough on the academic and scientific media. This situation can be derived, for example, from the criticality or confidentiality of information handled, from the domains in which they are applied (terrorist profiling, military objectives, etc.), or from the caution of experiences with regard to their application and use, which prevents making the process executed and the results obtained public.

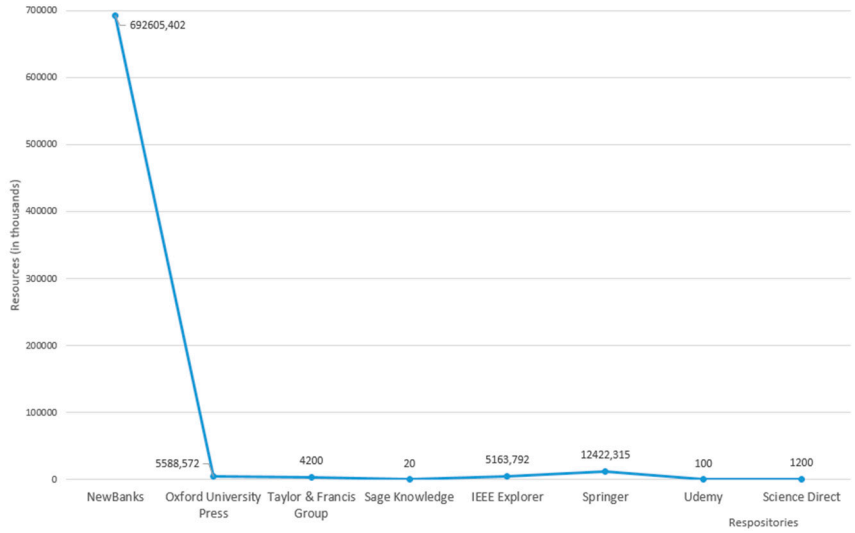

Figure 15. Total Resources published in repositories.

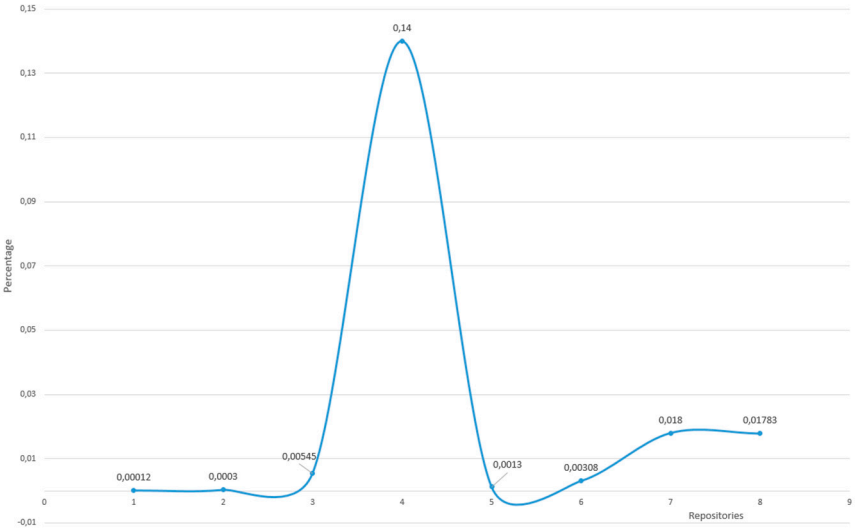

Figure 16. Percentage of OSINT resources in repositories.

Overall, the results of this study show an alignment with the factors contributing to the growth of OSINT industry [50–52]. Thanks to the accumulation of information currently increasing as well as the opening of a growing number of sources, OSINT has been able to make almost real time analysis, presenting a remarkable development in the areas of defense, national security, and public security, both in virtual and physical environments. On the other hand, regarding the limitations foreseen for the growth of the open-source intelligence market, the lack of investments and experience in open source analysis can be clearly identified, along with data quality problems. Finally, it can also be recognized that in the global market, North America has the highest share as far as OSINT products is concerned, and its growth rate is based on the demand for cloud-based security solutions. These OSINT market reports trends—projected to 2026—consolidate not only the results obtained in this research about the areas of growth and approach that has been given to it, but also its growth projections, thanks to the benefits that this technology offers for the security and defense processes.

However, despite the growing spread of OSINT as a training topic, as well as the existence of blogs and sites dedicated to explaining both its application process, and the use of the tools supporting this technology, along with the existence of videos, talks about its uses, different tools and approaches

to its professionalization, the review carried out does not identify any studies or statistics that could allow analyzing the production and growth of OSINT resources.

5. Conclusions

This study shows that the interest in OSINT has been growing, taking into account its benefits for performing intelligence processes that allow to increasingly generate reactions to risks and threats in real time, by making the most of the amount of data and sources available on the web. This type of intelligence has been playing a preponderant role in sensitive areas, facing the existing world conditions such as security and defense in their different modalities and strategies. As a result, it has projected itself for 2026 as one of the markets with an increasing flow of money for its use. Similarly, in addition to the growth of the existing ones, the application of OSINT is projected as future strategies to continue supporting processes of analysis of markets, services and products, along with other intelligence schemes.

Regarding the academic scientific dissemination of OSINT resources, although it is currently growing, it does not reflect a high degree of participation within the repositories and databases, being those non-profit and free use sources the ones with the greatest presence of OSINT resources. On the other hand, both the publications and the Latin American repositories present an even lower production and publication rate, which makes it possible to visualize a field of action available to address this type of technology in private or public projects.

As for the shortcomings that the diffusion of resources of this type of technology allows to identify, the following can be considered:

- the need to document the metadata of existing resources in a more thorough way, which allows for more timely information for those seeking to exploit such resources;
- given the growing foray of OSINT in training processes either as a subject or as a global body of knowledge, it is therefore necessary to design educational resources that provide not only a clear and timely metadata structure but also improve their accessibility and reuse, in addition to being didactically and pedagogically contextualized;
- along with creating these educational resources, it is also necessary to index such OSINT resources in educational resource repositories, apart from being able to have an open license for them since the few discussing the subject are exclusive and linked to training processes for a fee;
- as for the information that people publish, often without knowledge of its public availability, it is, therefore, essential to propose strategies that allow people to become aware of the sensitivity of the information they publish;
- open data policies are presented as another challenge to be addressed, given the tendency to open data to meet requirements, rather than having a real approach to openness;
- given the increasing openness of public data, it is necessary to design applications that allow the ordinary citizen to be able to exploit and analyze the information coming from this type of sources;
- finally, and considering the growing availability of both information and tools to capture such information, it is important to train people who interact with OSINT in how to analyze data and provide better results for decision-making processes. Similarly, it is vital to strengthen competencies on how to refine the quality of information coming from multiple open sources in order to provide better quality analysis and intelligence.

In general terms, a promising future for OSINT can be considered in different fields of action. However, a greater focus is required in the last stages of the open-source intelligence process in order to be able to use this knowledge in more assertive ways and with better quality. As future work, the design of a data model that allows to complement the description of educational resources can be projected. This could contribute to the accessibility and the reuse of the same resources with the purpose of adding value in the results and intelligence processes carried out by this type of technology in domains such as education.

Author Contributions: Conceptualization, J.F.H.-C., P.A.G.-G. and S.S.-A.; methodology, J.F.H.-C.; software, J.F.H.-C., P.A.G.-G. and S.S.-A.; validation, J.F.H.-C., P.A.G.-G. and S.S.-A.; formal analysis, J.F.H.-C.; writing—original draft preparation, J.F.H.-C.; writing—review and editing, J.F.H.-C., P.A.G.-G. and S.S.-A.; supervision, P.A.G.-G. and S.S.-A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: This work was developed within the doctoral research project framework on Linked Data at the Universidad Distrital Francisco José de Caldas. In the same way, Linked Data and open sources are also being worked as research topics of the GIIRA Research Group.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Queried Sources

Information retrieval using APIs or web scraping techniques. The sources queried by these strategies are shown in Table A1.

Table A1. Sources queried using Application Programming Interfaces (APIs) or web scraping techniques.

Platform	Observations
ARXIV	Online archive for the publication of scientific articles in the fields of mathematics, physics, computer science, and quantitative biology. API: http://export.arxiv.org/api. Format responding: XML
ScienceDirect	Online scientific and medical research database. API: https://api.elsevier.com/. Format Responding: XML Security: API key
Google Scholar	Search engine specialized in scientific-academic content and bibliography. API: None. Web Scraping must be done with PyScholar.
Udemy	Online learning platform. API: https://www.udemy.com/api-2.0/. Format responding: JSON. Security: Basic Author, Secret, and ID, supplied by UDEMY
YouTube API v3	Video sharing website. API: https://developers.google.com/youtube/v3/. Format responding: JSON

Information retrieval using web interfaces from sources. The sources queried by these strategies are shown below.

- Booklick: https://booklick.co/
- NewsBank: https://www.newsbank.com/
- PasaLaPagina: https://pasalapagina.com/
- Oxford University Press: https://global.oup.com/?cc=co
- Taylor and Francis Group: https://taylorandfrancis.com/
- Springer Nature: https://www.springernature.com/gp
- Sage Journals: https://journals.sagepub.com/
- Sage Knowledge; https://sk.sagepub.com/
- IEEE Explore: https://ieeexplore.ieee.org/Xplore/home.jsp
- MathScinet: https://mathscinet.ams.org/mathscinet
- ProQuest: https://about.proquest.com/
- Springer Link: https://link.springer.com/
- Scopus: https://www.scopus.com/home.uri
- Zbmath: https://zbmath.org/
- Britannica Academic: https://academic.eb.com/
- Thomson Reuters: https://www.thomsonreuters.com/en.html

- Ebsco Host: https://www.ebsco.com/es-es/productos/plataforma-ebscohost
- Naxos Music Library: https://www.naxosmusiclibrary.com/home.asp?rurl=%2Fdefault%2Easp
- Naxos Sheet Music: http://sheetmusiclib.com/login.aspx
- Britannica Image Quest: https://quest.eb.com/failedlogin?target=%2F
- Britannica Enciclopedia Moderna: https://moderna.eb.com/
- Digitalia Films Library: https://www.digitaliafilmlibrary.com/
- Digitalia Hispánica: http://www.digitaliapublishing.com/
- HighWire Press, Stanford University: https://www.highwirepress.com/
- Latindex: https://www.latindex.org/latindex/inicio
- SciELO: https://scielo.org/es/

MOOCs information retrieval. The sources queried by these strategies are shown below.

- The Open University: http://www.open.ac.uk/
- Iversity Open Course Platform: https://un.iversity.org/
- Alison—global online learning community: https://alison.com/es
- OpenLearning—University of New South Wales: https://www.openlearning.com/unswmoocs/
- Coursera: https://es.coursera.org/
- edX: https://www.edx.org/
- Udemy: https://www.udemy.com/
- SANS Institute: https://www.sans.org/

Information retrieval from other repositories. The sources queried by these strategies are shown below.

- Open Yale Courses: https://oyc.yale.edu/
- Carnegie Mellon University: https://www.cmu.edu/
- Japanese Industrial Standards Committee: https://www.jisc.ac.uk/
- Open Michigan: https://open.umich.edu/
- National Center for Curriculum Development in non-proprietary systems: https://cedec.intef.es/recursos/
- CREA Resources: https://emtic.educarex.es/proyectocrea
- Banco Interamericano de Desarrollo: https://www.iadb.org/es
- Eduteka—Universidad Icesi: http://eduteka.icesi.edu.co/
- MIT Open Courseware: https://ocw.mit.edu/index.htm
- Universidad Politécnica de Madrid Open Courseware: http://ocw.upm.es/
- Europeana: http://www.europana.eu/es
- Course Hero: https://www.coursehero.com/

Appendix B. Queried Sources without OSINT Resources

Springer Nature, MathScinet, Zbmath, Pasalapagina, Naxos Music Library, Naxos Sheet Music, Britannica Academic, Britannica Image Quest, Britannica Enciclopedia Moderna, Digitalia Films Library, Digitalia Hispánica, Highwire, Latindex, y SciELO.

Glossary

Big data	it refers to extremely large data sets that require a scalable architecture for efficient storage, manipulation, and analysis (University of Wisconsin).
Deep web search	this search is not indexed by popular search engines. Users require login or credentials to discover and access a specific service.
Intelligence	process of collecting, processing, and analyzing raw data from different sources and transform them into information to address a need, make decisions on, or to be used in a context.
Intelligence processes	also called the Intelligence Cycle. It refers to the process of tasking, collecting, processing, analyzing, and disseminating intelligence. The cycles consist of six steps: Requirements, planning & direction, collection, processing, analysis & production, dissemination, and feedback. (US Naval War College)
JSON	according to Request for Comments (RFC) 8259, JavaScript Object Notation (JSON) is a lightweight, text-based, language-independent data interchange format.
Knowledge domain	it refers to the knowledge related to a common topic, or process, for instance. In other words, a field of study, a branch of knowledge, or discipline on which research is developed.
Open source	openly sources such as news broadcasts, public repositories, social media, traditional mass media, conference proceedings, for instance, that provide publicly available materials. In these sources, you can find data in several formats: text, video, image, and audio.
OSINT ecosystem	it refers to the major players (technological tools, application domains, open sources, people, resources, for instance) as well as main processes that interact in the OSINT life-cycle.
OSINT educational resources	it refers to courses, textbooks, streaming videos, and any other educational materials that have been made available for educational purposes and are used to support access, teaching, learning, and research purposes on OSINT.
OSINT publications	OSINT research, educational and academic publications, such as research papers, journals, conference proceedings, etc.
OSINT resources	it refers to videos, documents, sites, and any other materials about OSINT, published on the web and not necessarily made for educational purposes.
OSINT subareas	it refers to the codification of the key areas or sectors of OSINT application. This areas are identified in different kinds of literature. OSINT has found a big deal of use in the fields and sectors like Government, Defense, Cyber Security, among others.
OSINT technologies	set of tools, techniques or specialized software which allow intelligence labors over open information sources, focusing on specific kind of information or type of data format (emails, documents, domain, etc.)
Surface web search	kind of search available to the general public using a search engine and simple keywords.
Web scraping	it is a set of methods useful to extract data from a website by identifying patterns, transforming the information into structured data for later analysis

Abbreviations

The following abbreviations are used in this manuscript:

API	Application Programming Interface
DOAJ	Directory of Open Access Journals
EBSCO	Elton B. Stephens Company
Get	It is a HTTP methods and is used to request data from a specified resource.
IEEE	The Institute of Electrical and Electronics Engineers, Inc.
Infosec	information security
JSON	JavaScript Object Notation
MOOCs	Massive Online Open Courses
NASA	National Aeronautics and Space Administration
OER	Open Educational Resources
OS	Operating System
OSINT	Open-source intelligence
PDF	Portable Document Format
RFC	Request for Comments. Formal document from the Internet Engineering Task Force (IETF)
VPN	virtual private network

References

1. Pune, M.; Open Source Intelligence (OSINT). Market Research Report-Global Forecast to 2023—Market Analysis, Scope, Stake, Progress, Trends and Forecast to 2023. Market Research Future. 2020. Available online: https://www.marketresearchfuture.com/reports/open-source-intelligence-market-4545 (accessed on 28 May 2020).
2. Pastorino, C. Técnicas y Herramientas OSINT Para la Investigación en Internet. Welivesecurity by ESET. 2019. Available online: https://www.welivesecurity.com/la-es/2019/10/07/tecnicas-herramientas-osint-investigacion-internet/ (accessed on 20 March 2020).
3. Passi, H. Top. 10 Popular Open Source Intelligence (OSINT) Tools. GreyCampus. 2018. Available online: https://www.greycampus.com/blog/information-security/top-open-source-intelligence-tools (accessed on 28 March 2020).
4. Portillo, I.; Nykiel, W. From Zero to OSINT Hero. Universidad de Alcalá de Henares. 2019. Available online: https://es.slideshare.net/WiktorNykielLION/from-zero-to-osint-hero-universidad-de-alcal-de-henares-ivan-portillo-morales-y-wiktor-nykiel (accessed on 28 May 2020).
5. Pastor-Galindo, J.; Nespoli, P.; Gomez Marmo, F.; Martinez Perez, G. OSINT Is the Next Internet Goldmine: Spain as an Unexplored Territory. V Jornadas Nacionales de Investigación en Ciberseguridad. 2019. Available online: https://www.researchgate.net/publication/333703698_OSINT_is_the_next_Internet_goldmine_Spain_as_an_unexplored_territory (accessed on 25 March 2020).
6. Norton, R. Guide to Open Source Intelligence. *Intell. J. US Intell. Stud.* **2011**, *18*, 65–67. Available online: https://www.afio.com/publications/Norton_Open_Source_in_AFIO_INTEL_WinterSpring2011.pdf (accessed on 1 April 2020).
7. NATO. *Open Source Intelligence Handbook*; North Atlantic Treaty Organization: Brussels, Belgium, 2001. Available online: http://www.oss.net/dynamaster/file_archive/030201/ca5fb66734f540fbb4f8f6ef759b258c/NATO%20OSINT%20Handbook%20v1.2%20-%20Jan%202002.pdf (accessed on 2 March 2020).
8. Korkisch, F. *NATO Gets Better Intelligence*; Center for Foreign and Defense Policy: Vienna, Austria, 2010. Available online: https://natowatch.org/sites/default/files/NATO_Gets_Better_Intell_April_PDP_0.pdf (accessed on 29 June 2020).
9. LISA Institute. Osint (Inteligencia de Fuentes Abiertas): Tipos, Métodos y Salidas Profesionales. 2020. Available online: https://www.lisainstitute.com/blogs/blog/osint-inteligencia-fuentes-abiertas (accessed on 28 June 2020).
10. Pastor-Galindo, J.; Nespoli, P.; Gomez Marmol, F.; Martinez Perez, G. The not yet exploited goldmine of OSINT: Opportunities, open challenges and future trends. *IEEE Access* **2020**, *8*, 10282–10304. [CrossRef]
11. Lee, S.; Shon, T. Open source intelligence base cyber threat inspection framework for critical infrastructures. In Proceedings of the 2016 Future Technologies Conference (FTC), San Francisco, CA, USA, 6–7 December 2016; pp. 1030–1033. [CrossRef]
12. Yeboah-Ofori, A.; Brimicombe, A. Cyber Intelligence and OSINT: Developing Mitigation Techniques against Cybercrime Threats on Social Media. *Int. J. Cyber Secur. Digit. Forensics.* **2018**, *7*, 87–98. [CrossRef]
13. Backfried, G.; Schmidt, C.; Pfeiffer, M.; Quirchmayr, G.; Glanzer, M.; Rainer, K. Open Source Intelligence in Disaster Management. In Proceedings of the 2012 European Intelligence and Security Informatics Conference, Odense, Denmark, 22–24 August 2012; pp. 254–258. [CrossRef]
14. Eijkman, Q.; Weggemans, D. Open Source Intelligence and Privacy Dilemmas: Is it Time to Reassess State Accountability? *Sec. Hum. Rights* **2013**, *23*, 285–296. [CrossRef]
15. Best, C. Challenges in Open Source Intelligence. In Proceedings of the 2011 European Intelligence and Security Informatics Conference, Athens, Greece, 12–14 September 2011; pp. 58–62. [CrossRef]
16. Carcaño, F. What Is OSINT and What Are Open Sources? FCD Intelligence. 2018. Available online: https://www.fcd-intelligence.com/2018/09/que-es-osint-y-que-son-fuentes-abiertas/ (accessed on 5 May 2020).
17. Hassan, N. An Introduction to Open Source Intelligence (OSINT) Gathering. 2018. Available online: https://www.secjuice.com/introduction-to-open-source-intelligence-osint/ (accessed on 1 July 2020).
18. Khera, V. An Introduction to Open Source Intelligence (OSINT). Cyber Security Magazine. 2020. Available online: https://cybersecurity-magazine.com/an-introduction-to-open-source-intelligence-osint/ (accessed on 30 June 2020).
19. Richelson, J. *The US Intelligence Community*, 7th ed.; Routledge: England, UK, 2016.

20. Williams, H.; Blum, I. Defining Second Generation Open Source Intelligence (OSINT) for the Defense Enterprise. RAND Corporation. 2018. Available online: https://www.rand.org/pubs/research_reports/RR1964.html (accessed on 10 March 2020).
21. Recorded Future. What Is Open Source Intelligence and How Is it Used? Recorded Future. 2019. Available online: https://www.recordedfuture.com/open-source-intelligence-definition/ (accessed on 20 May 2020).
22. Bielska, A.; Anderson, N.; Benetis, V.; Viehman, C. *Open Source Intelligence Tools and Resources Handbook*; I-Intelligence: Zurich, Switzerland, 2018. Available online: https://www.i-intelligence.eu/wp-content/uploads/2018/06/OSINT_Handbook_June-2018_Final.pdf (accessed on 22 May 2020).
23. Hock, R. Internet Tools and Resources for Open-Source Intelligence. Online Strategies. 2020. Available online: http://www.onstrat.com/osint/ (accessed on 2 July 2020).
24. Awesome Open Source. The Top. 146 Osint Open Source Projects. 2020. Available online: https://awesomeopensource.com/projects/osint (accessed on 19 June 2020).
25. Kemp, S. Digital 2020 Global Digital Overview. 2020. Available online: https://datareportal.com/reports/digital-2020-global-digital-overview (accessed on 10 July 2020).
26. Mejía, J. Estadísticas de Redes Sociales 2020: Usuarios de Facebook, Instagram, YouTube, LinkedIn, Twitter, Tiktok y Otros. 2020. Available online: https://www.juancmejia.com/marketing-digital/estadisticas-de-redes-sociales-usuarios-de-facebook-instagram-linkedin-twitter-whatsapp-y-otros-infografia/ (accessed on 25 March 2020).
27. Martínez, A. OSINT—La Información es Poder. Incibe-Cert. 2014. Available online: https://www.incibe-cert.es/blog/osint-la-informacion-es-poder (accessed on 1 July 2020).
28. World Wide Web Foundation. The Open Data Barometer. 2020. Available online: https://opendatabarometer.org/?_year=2017&indicator=ODB (accessed on 10 July 2020).
29. European Data Portal. Open Data Maturity Report 2019. 2019. Available online: https://www.europeandataportal.eu/sites/default/files/open_data_maturity_report_2019.pdf (accessed on 1 May 2020).
30. CIS. Cybersecurity Spotlight—The Surface Web, Dark Web, and Deep Web. Center of Internet Security. 2019. Available online: https://www.cisecurity.org/spotlight/cybersecurity-spotlight-the-surface-web-dark-web-and-deep-web/ (accessed on 20 June 2020).
31. Quinney, A. Surface Web vs Deep Web vs Dark Web. Service Care Solutions. 2016. Available online: https://www.servicecare.org.uk/news/surface-web-vs-deep-web-vs-dark-web-61792715468 (accessed on 20 June 2020).
32. Pinto, R.; Hernández, M.; Pinzón, C.; Díaz, D.; García, J. Inteligencia de fuentes abierta (OSINT) para operaciones de ciberseguridad. Aplicación de OSINT en un contexto colombiano y análisis de sentimientos. *Rev. Vínculos Cienc. Tecnol. Soc.* **2018**, *15*, 195–214. [CrossRef]
33. Montejo-Ráez, A.; Martínez-Cámara, E.; Martín-Valdivia, M.; Urena-López, L. Ranked Word Net graph for Sentiment Polarity Classification in Twitter. *Comput. Speech Lang.* **2014**, *28*, 93–107. [CrossRef]
34. Acquisti, A.; Fong, C. An Experiment in Hiring Discrimination via Online Social Networks. 2015. Available online: http://dx.doi.org/10.2139/ssrn.2031979 (accessed on 10 June 2020).
35. Central Intelligence Agency. INTellingence: Open Source Intelligence. 2018. Available online: https://www.cia.gov/news-information/featured-story-archive/2010-featured-story-archive/open-source-intelligence.html (accessed on 5 June 2020).
36. ReportLinked. Global Open Source Intelligence Market Is Expected to Grow with a CAGR of 23.2% over the Forecast Period from 2020–2026. GlobeNewsWire. 2020. Available online: https://www.globenewswire.com/news-release/2020/06/25/2053377/0/en/Global-open-source-intelligence-market-is-expected-to-grow-with-a-CAGR-of-23-2-over-the-forecast-period-from-2020-2026.html (accessed on 25 July 2020).
37. iCrowd Newswire. Open Source Intelligence (OSINT) Market 2020 Global Analysis by Size, Share, Developments, Key Players, Opportunities, Future Prospects and Forecast 2023. 2020. Available online: https://reportedtimes.com/open-source-intelligence-osint-market-2020-global-analysis-by-size-share-developments-key-players-opportunities-future-prospects-and-forecast-2023/ (accessed on 2 August 2020).
38. Petersen, K.; Feldt, R.; Mujtaba, S.; Mattsson, M. Systematic mapping studies in software engineering. In Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering, Bary, Italy, 26–27 June 2008; pp. 68–77. Available online: http://www.robertfeldt.net/publications/petersen_ease08_sysmap_studies_in_se.pdf (accessed on 2 August 2020).

39. Xiao, L. Clustering research based on feature selection in the behavior analysis of MOOC users. *J. Inf. Hiding Multimed. Signal Process.* **2019**, *10*, 147–155. Available online: http://bit.kuas.edu.tw/~{}jihmsp/2019/vol1/JIH-MSP-2019-01-17.pdf (accessed on 22 July 2020).
40. Mozilla MDN Web Docs. Cadenas del User Agent de Gecko. 2019. Available online: https://developer.mozilla.org/es/docs/Cadenas_del_User_Agent_de_Gecko (accessed on 2 March 2020).
41. MYIP.MS. User Agent Mozilla/5.0. 2020. Available online: https://myip.ms/view/comp_browsers/2345/Firefox_72.html (accessed on 2 May 2020).
42. Martín-Martín, A.; Orduna-Malea, E.; Thelwall, M.; Delgado López-Cózara, E. Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *J. Informetr.* **2018**, *12*, 1160–1177. [CrossRef]
43. European Union Agency for Law Enforcement Training. OSINT. 2019. Available online: https://www.cepol.europa.eu/tags/osint (accessed on 15 May 2020).
44. Deahi, D. Metadata Is the Biggest Little Problem Plaguing the Music Industry. The Verge. 2019. Available online: https://www.theverge.com/2019/5/29/18531476/music-industry-song-royalties-metadata-credit-problems (accessed on 31 May 2020).
45. Swoyer, S. It's Official: Metadata Management Is a Strategic Problem. UpSide Where Data Means Business. 2016. Available online: https://tdwi.org/articles/2016/11/02/metadata-management-is-a-strategic-problem.aspx (accessed on 30 June 2020).
46. Daminion. The Most Common Issues with Metadata. 2011. Available online: https://daminion.net/articles/tips/the-most-common-issues-with-metadata/ (accessed on 30 June 2020).
47. Nouh, M.; Nurse, J.; Webb, H.; Goldsmith, M. Cybercrime investigators are users too! Understanding the socio-technical challenges faced by law enforcement. In Proceedings of the 2019 Workshop Usable Security, San Diego, CA, USA, 24 February 2019; Available online: https://www.ndss-symposium.org/wp-content/uploads/2019/02/usec2019_02-3_Nouh_paper.pdf (accessed on 2 August 2020).
48. Dawson, M.; Lieble, M.; Adeboje, A. Open source intelligence: Performing data mining and link analysis to track terrorist activities. In *Information Technology—New Generations*; Springer: Cham, Switzerland, 2018; Volume 558, pp. 1–11. Available online: https://link.springer.com/chapter/10.1007%2F978-3-319-54978-1_22 (accessed on 20 July 2020).
49. Ashcroft, M.; Fisher, A.; Kaati, L.; Omer, E.; Prucha, N. Detecting Jihadist messages on twitter. In Proceedings of the 2015 European Intelligence and Security Informatics Conference, Manchester, UK, 7–9 September 2015; pp. 161–164. Available online: https://ieeexplore.ieee.org/document/7379742 (accessed on 17 July 2020).
50. MarketWatch. OSINT Market- What Are the Main Factors That Contributing Towards Industry Growth? Press Release. 2020. Available online: https://www.marketwatch.com/press-release/osint-market--what-are-the-main-factors-that-contributing-towards-industry-growth-2020-08-07?tesla=y (accessed on 25 June 2020).
51. Allied Market Research. Open Source Intelligence Market Statistics: 2027. 2020. Available online: https://www.alliedmarketresearch.com/open-source-intelligence-market (accessed on 17 June 2020).
52. AP News. Open Source Intelligence (OSINT) Market 2020 Global Size, Share, Industry Growth, Opportunities, Development Status, Competitive Landscape, and Regional Forecast 2023. 2020. Available online: https://apnews.com/a47934d9ca6f0b27605a2e93ecad283c (accessed on 22 June 2020).