



# Open Source Intelligence Methods and Tools

A Practical Guide to Online Intelligence

—

Nihad A. Hassan  
Rami Hijazi

**Apress®**

# **Open Source Intelligence Methods and Tools**

**A Practical Guide to Online  
Intelligence**

**Nihad A. Hassan  
Rami Hijazi**

**Apress®**

# ***Open Source Intelligence Methods and Tools: A Practical Guide to Online Intelligence***

Nihad A. Hassan  
New York, USA

Rami Hijazi  
Mississauga, Ontario, Canada

ISBN-13 (pbk): 978-1-4842-3212-5  
<https://doi.org/10.1007/978-1-4842-3213-2>

ISBN-13 (electronic): 978-1-4842-3213-2

Library of Congress Control Number: 2018948821

Copyright © 2018 by Nihad A. Hassan, Rami Hijazi

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

Trademarked names, logos, and images may appear in this book. Rather than use a trademark symbol with every occurrence of a trademarked name, logo, or image we use the names, logos, and images only in an editorial fashion and to the benefit of the trademark owner, with no intention of infringement of the trademark.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Managing Director, Apress Media LLC: Welmoed Spahr  
Acquisitions Editor: Susan McDermott  
Development Editor: Laura Berendson  
Coordinating Editor: Rita Fernando

Cover designed by eStudioCalamar

Cover image designed by Freepik ([www.freepik.com](http://www.freepik.com))

Distributed to the book trade worldwide by Springer Science+Business Media New York, 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail [orders-ny@springer-sbm.com](mailto:orders-ny@springer-sbm.com), or visit [www.springeronline.com](http://www.springeronline.com). Apress Media, LLC is a California LLC and the sole member (owner) is Springer Science + Business Media Finance Inc (SSBM Finance Inc). SSBM Finance Inc is a **Delaware** corporation.

For information on translations, please e-mail [rights@apress.com](mailto:rights@apress.com), or visit [www.apress.com/rights-permissions](http://www.apress.com/rights-permissions).

Apress titles may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Print and eBook Bulk Sales web page at [www.apress.com/bulk-sales](http://www.apress.com/bulk-sales).

Any source code or other supplementary material referenced by the author in this book is available to readers on GitHub via the book's product page, located at [www.apress.com/9781484232125](http://www.apress.com/9781484232125). For more detailed information, please visit [www.apress.com/source-code](http://www.apress.com/source-code).

Printed on acid-free paper

*To my mom, Samiha, thank you for everything.  
Without you, I'm nothing.*

*—Nihad A. Hassan*

# Table of Contents

**About the Authors..... xiii**

**About the Technical Reviewer .....xv**

**Acknowledgments .....xvii**

**Introduction .....xix**

  

**Chapter 1: The Evolution of Open Source Intelligence..... 1**

    Open Source Information Categories ..... 3

    OSINT Types ..... 5

    Digital Data Volume..... 5

    OSINT Organizations ..... 6

        Government Organizations ..... 7

        Private Sector ..... 7

        Gray Literature Vendors ..... 8

    Parties Interested in OSINT Information..... 10

        Government ..... 10

        International Organizations ..... 11

        Law Enforcement Agencies ..... 11

        Business Corporations..... 12

        Penetration Testers and Black Hat Hackers/Criminal Organizations ..... 12

        Privacy-Conscious People ..... 13

        Terrorist Organizations ..... 13

    Information Gathering Types ..... 14

        Passive Collection..... 14

        Semipassive ..... 14

        Active Collection..... 15

TABLE OF CONTENTS

Benefits of OSINT ..... 15

Challenges of Open Source Intelligence ..... 16

Legal and Ethical Constraints ..... 17

Summary..... 18

Notes..... 19

**Chapter 2: Introduction To Online Threats and Countermeasures ..... 21**

Online Threats ..... 22

    Malware..... 22

    Black Hat Hackers ..... 23

    Pharming ..... 23

    Phishing..... 24

    Ransomware..... 27

    Adware and Spyware..... 28

    Trojan..... 29

    Virus ..... 29

    Worms ..... 29

    Scareware ..... 29

    Rootkits ..... 30

    Juice Jacking ..... 30

    Wi-Fi Eavesdropping..... 30

Security Software ..... 31

    Antivirus ..... 31

    Firewall..... 32

    Anti-malware..... 33

Securing the Operating System ..... 33

    Hardening the Windows OS..... 34

    Staying Private in Windows 10..... 39

    Destroying Digital Traces..... 41

General Privacy Settings ..... 45

    Covering Your Laptop Camera ..... 45

    Avoiding Pirated Software ..... 45

Handling Digital Files Metadata.....	46
Physically Securing Computing Devices.....	50
Online Tracking Techniques.....	52
Tracking Through IP Address .....	52
Cookies.....	55
Digital Fingerprinting.....	57
HTML5 .....	58
Checking Your Digital Footprint .....	58
Secure Online Browsing.....	59
Configuring Firefox to Become More Private.....	59
Secure Online Communication.....	64
VPN.....	65
Proxies.....	66
DNS Leak Test.....	67
Online Anonymity .....	69
Using the TOR Network.....	69
Using the Tails OS and Other Security OSs.....	76
Sharing Files Securely.....	77
Making Anonymous Payments .....	79
Encryption Techniques .....	81
Securing Your Passwords.....	81
Encrypting Your Hard Drive/USB Sticks .....	82
Cloud Storage Security.....	82
Secure E-mail Communications .....	83
Virtualization Technology .....	86
Android and iOS Emulator .....	88
Essential Prerequisites.....	88
Drawing Software and Data Visualization.....	89
Free Translation Services .....	92
Final Tips .....	92
Summary.....	94

TABLE OF CONTENTS

**Chapter 3: The Underground Internet ..... 95**

    Layers of the Internet..... 96

    Darknet Users ..... 103

    Accessing the Darknet..... 104

        Security Checks When Accessing the Darknet..... 104

        Accessing the Darknet from Within the Surface Web..... 106

    Using Tor ..... 107

    Using the Tails OS ..... 109

        Warning When Using the Tails OS..... 114

    Searching the Tor Network ..... 115

    Other Anonymity Networks ..... 116

        I2P ..... 117

        Freenet ..... 123

    Going Forward..... 123

    Summary..... 124

    Notes..... 125

**Chapter 4: Search Engine Techniques ..... 127**

    Keywords Discovery and Research..... 129

    Using Search Engines to Locate Information ..... 130

        Google ..... 130

        Bing ..... 138

        Privacy-Oriented Search Engines ..... 140

        Other Search Engines ..... 141

        Business Search Sites..... 142

        Metadata Search Engines..... 147

        Code Search ..... 150

        FTP Search Engines..... 151

        Automated Search Tools..... 152

        Internet Of Things (IoT) Device Search Engines ..... 153

    Web Directories..... 154



Translation Services.....	156
Website History and Website Capture.....	158
Website Monitoring Services .....	160
RSS Feed .....	162
News Search.....	163
Customize Google News.....	164
News Websites .....	166
Fake News Detection.....	166
Searching for Digital Files.....	170
Document Search .....	170
Image.....	183
Video.....	191
File Extension and File Signature List .....	196
Productivity Tools.....	196
Summary.....	201
Notes.....	201
<b>Chapter 5: Social Media Intelligence.....</b>	<b>203</b>
What Is Social Media Intelligence? .....	205
Social Media Content Types.....	206
Classifications of Social Media Platforms .....	208
Popular Social Networking Sites .....	210
Investigating Social Media Sites .....	211
Facebook.....	211
Twitter .....	231
Google+ .....	241
LinkedIn.....	247
General Resources for Locating Information on Social Media Sites .....	253
Other Social Media Platforms .....	254
Pastebin Sites .....	255

TABLE OF CONTENTS

Social Media Psychological Analysis ..... 256

    Tone Analyzer ..... 257

    Watson Tone Analyzer ..... 257

    Facebook and Twitter Prediction ..... 258

    Fake Sport ..... 258

    Review Meta ..... 258

    TweetGenie ..... 258

Summary..... 258

Notes..... 259

**Chapter 6: People Search Engines and Public Records ..... 261**

    What Is a People Search Engine? ..... 261

    What Are Public Records? ..... 262

    Example of Public Records ..... 263

    Searching for Personal Details..... 264

        General People Search ..... 264

        Online Registries ..... 268

        Vital Records ..... 269

        Criminal and Court Search..... 272

        Property Records..... 273

        Tax and Financial Records..... 274

        Social Security Number Search..... 275

        Username Check ..... 275

        E-mail Search and Investigation..... 275

        Data Compromised Repository Websites..... 277

        Phone Number Search..... 279

        Employee Profiles and Job Websites..... 280

        Dating Website Search ..... 281

        Other Public Records..... 283

Summary..... 284

Notes..... 284

<b>Chapter 7: Online Maps .....</b>	<b>285</b>
The Basics of Geolocation Tracking .....	285
How to Find the GPS Coordinates of Any Location on a Map .....	286
How to Find the Geocode Coordinates from a Mailing Address.....	288
General Geospatial Research Tools .....	288
Commercial Satellites .....	294
Date/Time Around the World .....	294
Location-Based Social Media.....	295
YouTube .....	295
Facebook .....	296
Twitter .....	298
Other Social Media Platforms.....	302
Conducting Location Searches on Social Media Using Automated Tools .....	303
Country Profile Information.....	304
Transport Tracking.....	304
Air Movements .....	305
Maritime Movements.....	307
Vehicles and Railway.....	309
Package Tracking.....	310
Webcams .....	311
Digital File Metadata .....	312
Summary.....	312
<b>Chapter 8: Technical Footprinting.....</b>	<b>313</b>
Investigate the Target Website .....	314
Investigate the Robots.txt File .....	316
Mirror the Target Website .....	317
Extract the Links.....	317
Check the Target Website's Backlinks .....	318
Monitor Website Updates.....	318
Check the Website's Archived Contents .....	318

TABLE OF CONTENTS

Identify the Technologies Used..... 319

Web Scraping Tools ..... 322

Investigate the Target Website’s File Metadata..... 324

Website Certification Search ..... 325

Website Statistics and Analytics Tools ..... 325

Website Reputation Checker Tools ..... 326

Passive Technical Reconnaissance Activities ..... 327

    WHOIS Lookup..... 327

    Subdomain Discovery..... 329

    DNS Reconnaissance..... 332

IP Address Tracking..... 337

Summary..... 339

**Chapter 9: What’s Next? ..... 341**

    Where Will OSINT Go Next? ..... 341

    OSINT Process..... 343

    Final Words ..... 344

**Index..... 345**

# About the Authors

**Nihad A. Hassan** is an independent information security consultant, digital forensics and cybersecurity expert, online blogger, and book author. He has been actively conducting research on different areas of information security for more than a decade and has developed numerous cybersecurity education courses and technical guides. He has completed several technical security consulting engagements involving security architectures, penetration testing, computer crime investigation, and cyber open source intelligence (OSINT). Nihad has authored four books and scores of information security articles for various global publications. He also enjoys being involved in security training, education, and motivation. His current work focuses on digital forensics, anti-forensics techniques, digital privacy, and cyber OSINT. He covers different information security topics and related matters on his security blog at [www.DarknessGate.com](http://www.DarknessGate.com) and recently launched a dedicated site for open source intelligence resources at [www.OSINT.link](http://www.OSINT.link). Nihad has a bachelor's of science honors degree in computer science from the University of Greenwich in the United Kingdom.

Nihad can be followed on Twitter (@DarknessGate), and you can connect to him via LinkedIn at <https://www.linkedin.com/in/darknessgate>.

**Rami Hijazi** has a master's degree in information technology (information security) from the University of Liverpool. He currently works at MERICLER Inc., an education and corporate training firm in Toronto, Canada. Rami is an experienced IT professional who lectures on a wide array of topics, including object-oriented programming, Java, e-commerce, agile development, database design, and data handling analysis. Rami also works as information security consultant, where he is involved in designing encryption systems and wireless networks, detecting intrusions and tracking data breaches, and giving planning and development advice for IT departments concerning contingency planning.

# About the Technical Reviewer

**Reem Naddar** has a bachelor's of science degree in mathematics from Dalhousie University and has been in the data analytics industry since 2006. She has substantial experience in designing and executing solutions that address complex business problems involving large-scale data warehousing, real-time analytics, software architecture, and reporting solutions. She employs leading-edge tools and techniques when implementing fast and efficient data acquisition including Big Data processing used by global practitioners.

Reem has worked for major corporations and chartered banks in Canada both as a contractor and as a permanent staff member. She is fond of open source intelligence (OSINT) projects where she adopts different frameworks and processes to capture, transform, analyze, and store terabytes of structured and unstructured data gathered from publicly available sources.

# Acknowledgments

I start by thanking God for giving me the gift to write and convert my ideas into something useful. Without God's blessing, I would not be able to achieve anything.

I want to thank the ladies at Apress: Susan, Rita, and Laura. I was pleased to work with you again and very much appreciate your valuable feedback and encouragement.

Specifically, to book acquisitions editor Susan McDermott, thank you for believing in my book's idea and for your honest encouragement before and during the writing process. To book project editor Rita Fernando, you were very supportive during the writing process. You made authoring this book a joyful journey. To book development editor Laura Berendson, thank you very much for your diligent and professional work in producing this book.

I also want to thank all the Apress staff who worked behind the scenes to make this book possible and ready for launch. I hope you will continue your excellent work in creating highly valued computing books. Your work is greatly appreciated.

—Nihad A. Hassan

# Introduction

*Open Source Intelligence Methods and Tools* focuses on building a deep understanding of how to exploit open source intelligence (OSINT) techniques, methods, and tools to acquire information from publicly available online sources to support intelligence analysis. The harvested data can be used in different scenarios such as financial, crime, and terrorism investigations as well as in more regular tasks such as analyzing business competitors, running background checks, and acquiring intelligence about individuals and other entities. This book will also improve your skills in acquiring information online from the surface web, the deep web, and the darknet.

Many estimates show that 90 percent of useful information acquired by intelligence services comes from public sources (in other words, OSINT sources). Social media sites open up numerous opportunities for investigations because of the vast amount of useful information located in one place. For example, you can get a great deal of personal information about any person worldwide by just checking their Facebook page. This book will show you how to conduct advanced social media investigations to access content believed to be private, use advanced search engines queries to return accurate results, search historical deleted versions of websites, track individuals online using public record databases and people-searching tools, locate information buried in the deep web, access and navigate the dark web, collect intelligence from the dark web, view multiple historic satellite images and street views of any location, search geolocation information within popular social media sites, and more. In short, you will learn how to use a plethora of techniques, tools, and free online services to gather intelligence about any target online.

OSINT-gathering activities should be conducted secretly to avoid revealing the searcher's identity. Therefore, this book will teach you how to conceal your digital identity and become anonymous online. You will learn how to exchange data secretly across hostile environments like the Internet and how to communicate with your peers privately and anonymously. You will also learn how to check your digital footprint and discover what kind of digital traces you are leaving behind and how to delete them.



*Open Source Intelligence Methods and Tools* is an indispensable guide for anyone responsible for collecting online content from public data, and it is a must-have reference for any casual Internet user who wants to dig deeper into the Internet to see what information it contains.

# Target Audience

The following types of people will benefit from this book:

- Penetration testers
- Digital forensics investigators
- Intelligence services
- Military personnel
- Law enforcement
- UN agencies and nonprofit organizations
- For-profit enterprises
- Risk management professionals
- Journalists
- Academic researchers
- University students
- End users who want to learn how to exploit Internet resources effectively

# What the Book Is Not

This book is not about the history of open source intelligence, and it does not discuss at length the legal issues of personal reconnaissance online. We will not talk about policies and regulations that govern different countries or business organizations. Although some of these issues are discussed briefly in Chapter 1, the main aim of this book is to create a guidebook to support all types of investigations. You can read the chapters in any order because each chapter is considered an isolated unit that discusses the chapter subject's comprehensively.

# Summary of Contents

Here is a brief description of each chapter's contents:

- Chapter 1, “The Evolution of Open Source Intelligence”: In this chapter, we introduce you to the term OSINT and explain how it has evolved over time. We introduce the different parties interested in exploiting publicly available data and the benefits gained from doing so. We include some technical information about online gathering techniques and the challenges involved, as well as the legal aspects when harvesting data from publicly available sources.
- Chapter 2, “Introduction To Online Threats and Countermeasures”: In this chapter, we teach you everything you need to know to stay safe when going online. This knowledge is essential when conducting advanced searches online to avoid being tracked since using advanced search operators and other OSINT search techniques will attract attention online and make your connection a target for interception by different outside parties.
- Chapter 3, “The Underground Internet”: This chapter is devoted to uncovering the secrets of the invisible web, which contains both the darknet and the deep web. This knowledge is essential as the underground net contains a wealth of valuable information that any cybersecurity professional should know how to access.
- Chapter 4, “Search Engine Techniques”: In this chapter, we show you how to use advanced search techniques using typical search engines such as Google and Bing to find anything online. We also cover other specialized search engines for images, video, news, web directories, files, and FTP.
- Chapter 5, “Social Media Intelligence”: In this chapter, we show you how to use a wide array of tools and techniques to gather intelligence about a specific person or entity from social media sites. For instance, using Facebook you can gather intelligence about people worldwide. Other major tech companies like Google and Microsoft own huge databases of information about their users. A great amount of information is published publicly on these sites, and this chapter

teaches you how to search for people, including their relationships, names, addresses, and communications (and interactions) with others on social sites, to formulate a complete profile about your target.

- Chapter 6, “People Search Engines and Public Records”: Here we list specific search engines and other public resources to search for people’s names and get details around them. You will learn to use different reverse search criteria to find people online such as birth records, mail addresses, résumés, dating websites, e-mails, phone numbers, previous breached usernames, and more. We also cover government resources such as vital records, tax records, criminal information, and other public sources you can use to gain intelligence about people and entities.
- Chapter 7, “Online Maps”: This chapter covers how to use Google Maps and other free geolocation services to investigate the geolocation information acquired about target people.
- Chapter 8, “Technical Footprinting”: This chapter covers how to gather technical information about a target website and network system in passive mode to support your OSINT intelligence.
- Chapter 9, “What’s Next?”: This chapter covers the OSINT process and its future trends.

## Book Companion Website

In this book, we list hundreds of online services that help OSINT gatherers to collect and analyze information. We all know about the ever-changing nature of the Web, though; new sites launch and others close down daily, so some links might not work by the time you read this. To prevent this hassle and to avoid making part of this book useless after publishing it, we have created a dedicated website where we offer a digital list of all the links mentioned in this book in addition to many more resources that just wouldn’t fit in the printed version. We will do our best to keep this site updated and continually work to add new useful OSINT content that reflects improvements in the field. Dead links will get deleted or updated, so the content of this book will remain current for many years to come.

See [www.OSINT.link](http://www.OSINT.link).

## Comments and Questions

To comment or ask technical questions about this book, send an e-mail to [nihad@protonmail.com](mailto:nihad@protonmail.com). For additional references about the subject, computer security tools, tutorials, and other related matters, check out the author's blog at [www.DarknessGate.com](http://www.DarknessGate.com).

## CHAPTER 1

# The Evolution of Open Source Intelligence

Since the end of the Cold War, global societies have become more open, and the revolution of the Internet and its widespread use have turned the world into a small village. Unleashing the Internet network to billions of people worldwide to communicate and exchange digital data has shifted the entire world into what is now an information age. This transformation to the digital age brought huge benefits to our society; however, the speed and scope of the transformation have also triggered different kinds of risks. For instance, cybercriminals, terrorist groups, oppressive regimes, and all kinds of malicious actors are using the Internet effectively to conduct their crimes. Juniper Research predicts that cybercrime will cost businesses more than \$2 trillion by 2019,<sup>1</sup> so these risks encourage governments to invest in the development of open source intelligence (OSINT) tools and techniques to counter current and future cybersecurity challenges.

OSINT refers to all the information that is publicly available. There is no specific date on when the term OSINT was first proposed; however, a relative term has probably been used for hundreds of years to describe the act of gathering intelligence through exploiting publicly available resources.

The United States is still leading the world in the intelligence arena, with vast resources dedicated by the U.S. government to its intelligence agencies that enable it to build sophisticated surveillance programs to harvest and analyze a large volume of data covering all the major spoken languages. This makes our discussion of OSINT history largely dependent on U.S. history, although during the Cold War many countries also developed OSINT capabilities to gain intelligence. Still, no other country has reached the level of the U.S. programs.

The U.S. Department of Defense (DoD) defines OSINT as follows:

“Open-source intelligence (OSINT) is an intelligence that is produced from publicly available information and is collected, exploited, and disseminated in a timely manner to an appropriate audience for the purpose of addressing a specific intelligence requirement.”<sup>ii</sup>

In modern times, OSINT was introduced during World War II as an intelligence tool when the United States established the Foreign Broadcast Information Service (FBIS) to monitor publicly available information that related to supporting its troop operations at that time. This all happened before the U.S. intelligence community even existed.

After the end of World War II, the FBIS has continued its work in exploiting OSINT sources globally, until the September 11, 2001, terror attacks on the United States. This drew attention to the importance of creating an independent OSINT agency to intensify exploiting these resources to protect national security. This is what was suggested by the 9/11 Commission, which called for the creation of a specialized agency for gathering OSINT.<sup>iii</sup> In 2005, the WMD Commission, which was formed to measure the effectiveness of the intelligence community to respond to threats raised by weapons of mass destruction (WMD) and other related threats of the 21st century, suggested the creation of an Open Source Directorate within the Central Intelligence Agency (CIA).<sup>iv</sup>

Following these recommendations and other debates, the Director of National Intelligence (DNI) announced the creation of the National Intelligence Open Source Center (OSC). The main tasks of the OSC are to collect information available from both online and offline public sources, which was previously done by the FBIS. Later, the Intelligence Reform and Terrorism Prevention Act, which was proposed to reform the intelligence activities of the U.S. government, merged the FBIS and other related research entities into one body. This organization is now called the Open Source Enterprise and is managed by the CIA.

OSINT sources are distinguished from other forms of intelligence because they must be legally accessible by the public without breaching any copyright or privacy laws. That's why they are considered “publicly available.” This distinction makes the ability to gather OSINT sources applicable to more than just security services. For example, businesses can benefit from exploiting these resources to gain intelligence about their competitors.

**Note!** During the search for OSINT sources, classified information that is not protected properly can appear. This includes leaked documents, such as those published by WikiLeaks. This type of information is called NOSINT, as opposed to OSINT. Intelligence usually considers all sources regardless of their legal accessibility.

---

In addition to its significant importance to the intelligence community, OSINT gathering is less expensive and less risky than traditional spying activities. Unlike other intelligence sources that may require using spy satellite images or secret agents to collect information, all you need to gather OSINT online resources is a computer and an Internet connection. And, of course, you need the required searching skills.

As technology proliferates and the volume of available data increases, government departments, nongovernmental organization (NGO) organizations, and business corporations are starting to rely to a large extent on OSINT rather than private and classified information. This book will teach you how to exploit OSINT sources to search for and gather information online. In this chapter, we will describe the term OSINT, discuss the types of OSIN, and talk about different parties' benefits from using OSINT and their motivations, as well as trends and challenges for the future. In later chapters, we will cover how to use a plethora of tools and techniques to acquire data from publicly available sources.

## Open Source Information Categories

There are different kinds of information that you may encounter when conducting OSINT analysis. According to the *NATO Open Source Intelligence Handbook V1.2* published in 2001, there are four categories of open information and intelligence.

- *Open source data (OSD)*: This is generic data coming from a primary source. Examples include satellite images, telephone call data and metadata, datasets, survey data, photographs, and audio or video recordings that have recorded an event.
- *Open source information (OSINF)*: This is generic data that has undergone some filtering first to meet a specific criterion or need; this data can also be called a secondary source. Examples include books about a specific subject, articles, dissertations, artworks, and interviews.

**Note!** The set of sources legally available to the public through specific channels is called *gray literature*. These sources include books, journals, dissertations, technical reports, and internal documents of commercial enterprises, commercial imagery, and any information that is controlled by its producer. Gray literature is a major element of OSINF and can be obtained legally by acquiring the permission of its copyright holder or by paying for it (for example, through subscriptions agencies, commercial bookstores, and so on).

---

- *Open source intelligence (OSINT)*: This includes all the information that has been discovered, filtered, and designated to meet a specific need or purpose. This information can be used directly in any intelligence context. OSINT can be defined in a nutshell as the output of open source material processing.
- *Validated OSINT (OSINT-V)*: This is OSINT with a high degree of certainty; the data should be confirmed (verified) using a non-OSINT source or from a highly reputable OSINT source. This is essential, as some outside adversaries may spread inaccurate OSINT information with the intent to mislead OSINT analysis. A good example of this is when a TV station broadcasts live the arrival of a president to another country; such information is OSINT, but it has a large degree of certainty.

As you saw, OSD and OSINF comprise the main sources (primary and secondary) of information that OSINT uses to drive its results.

Another issue you need to understand within the OSINT context is the difference between data, information, and knowledge. The three terms are usually used interchangeably; however, each one has a different meaning, although the three do interact with each other.

- *Data*: This is a set of facts describing something without further explanation or analysis. For example, “The price of gold per ounce is \$1,212.”
- *Information*: This is a kind of data that has been interpreted properly to give a useful meaning within a specific context. For example, “The price of gold per ounce has fallen from \$1,212 to \$1,196 within one week.”



- *Knowledge*: This is a combination of information, experience, and insight that has been learned or inferred after some experimentation. Knowledge describes what your brain has recorded in the past, and these records can help you to make better decisions about the future when facing similar contexts. For example, “When the price of gold falls more than 5 percent, this means the price of oil will fall too.”

## OSINT Types

OSINT includes all publicly accessible sources of information. This information can be found either online or offline, including in the following places:

- The Internet, which includes the following and more: forums, blogs, social networking sites, video-sharing sites like YouTube.com, wikis, Whois records of registered domain names, metadata and digital files, dark web resources, geolocation data, IP addresses, people search engines, and anything that can be found online
- Traditional mass media (e.g., television, radio, newspapers, books, magazines)
- Specialized journals, academic publications, dissertations, conference proceedings, company profiles, annual reports, company news, employee profiles, and résumés
- Photos and videos including metadata
- Geospatial information (e.g., maps and commercial imagery products)

## Digital Data Volume

As you already saw, OSINT encompasses not only online sources. Paper editions of public sources must also get investigated thoroughly as part of any OSINT-gathering process; however, online sources comprise the largest segment of OSINT.

Today we live in an information age, and publishers as well as corporations, universities, and other suppliers of OSINT sources are shifting their business processes to digital formats. The number of users on social media sites will also continue to increase, and the number of Internet of Things (IoT) devices will intensify in the future, leading to a huge increase in the volume of digital data coming from the billions of sensors and machines worldwide. In other words, most OSINT sources in the future will be online sources.

---

**Note!** Gartner estimates that 20.4 billion IoT devices will be in use by 2020.<sup>v</sup>

---

The volume of digital data is exploding rapidly. According to IDC Research,<sup>vi</sup> by the year 2020, the total amount of digital data created worldwide will reach 44 zettabytes, and the number will increase faster within five years to reach 180 zettabytes in 2025.

By 2020, the Gartner research group estimates that an average person will spend time interacting with automated bots more than with their spouse, and of course all these interactions will be digital. Another estimate says that in 2021, 20 percent of all activities a human do will involve using a service from at least one of the giant IT companies (Google, Apple, Facebook, Amazon). Not to mention, most people will prefer to use voice commands to interact with their computing devices over typing.

These figures should give you an idea about what the near future will look like in the digital age. The volume of digital data along with the increased number of people using the Internet to do their jobs will make online sources the primary source of OSINT for both governments and business corporations in the future.

## OSINT Organizations

Some specialized organizations provide OSINT services. Some of them are government based, and others are private companies that offer their services to different parties such as government agencies and business corporations on a subscription basis. In this section, we will mention the main OSINT organizations worldwide.

## Government Organizations

Government organizations working in OSINT analysis are still considered the best because of the resources available from their governments to do their jobs. The two most famous government agencies that do OSINT globally are the Open Source Center in the United States and BBC Monitoring in Great Britain.

### Open Source Center

We already talked about the Open Source Center (OSC); it is the largest OSINT organization and has vast resources to do its job. OSC works closely with other local intelligence agencies in the United States and offers its services to U.S. government intelligence agencies.

### BBC Monitoring

BBC Monitoring (<https://monitoring.bbc.co.uk/login>) is a department within the British Broadcasting Corporation (BBC) that monitors foreign media worldwide. It has a similar role as the Open Source Center in the United States, with the main difference being that it does not belong to British Intelligence. BBC Monitoring is funded from its stakeholders in addition to many commercial and governmental entities around the world. It was first established in 1939 and has offices in different countries around the globe. It actively monitors TV, radio broadcast, print media, Internet, and emerging trends from 150 countries in more than 70 languages. BBC Monitoring is directed by the BBC and offers its services on a subscription basis to interested parties such as commercial organizations and UK official bodies.

## Private Sector

You should not underestimate the private sector when looking at who supplies OSINT information; many private corporations have developed advanced programs and techniques to gather data from public sources for commercial gain. Indeed, most private OSINT corporations partner with government agencies to supply them with such information. In this section, we will mention the main ones around the globe.

## Jane's Information Group

Jane's Information Group (<http://www.janes.com>) is a British company founded in 1898. Jane's is a leading provider that specializes in military, terrorism, state stability, serious and organized crime, proliferation and procurement intelligence, aerospace, and transportation subjects. It publishes many journals and books related to security matters in addition to its OSINT sources that track and predict security matters in 190 states and 30 territories.

## Economist Intelligence Unit

The Economist Intelligence Unit (<https://www.eiu.com/home.aspx>) is the business intelligence, research, and analysis division of the British Economist Group. The main domain of the Economist Intelligence Unit is its business and financial forecasts; it offers a monthly report in addition to a country economic forecast for the coming five years with a comprehensive view about current trends on economic and political issues.

## Oxford Analytica

Oxford Analytica (<http://www.oxan.com>) is a relatively small OSINT firm compared with the previous two. Oxford Analytica specializes in geopolitics and macroeconomics subjects. It has a global macro expert network to advise its clients on the best practices of strategy and performance when accessing complex markets. Its expert networks contain more than 1400 experts. Most of them are scholars on their subject, senior faculty members in top universities, and high-profile specialists in their sector.

## Gray Literature Vendors

We already talked about gray literature as part of OSINT data. However, this type of data deserves to have its own reference when talking about the main sources of information used in OSINT gathering because of its great intelligence value.

Gray literature is mainly produced by the world's publishing companies. It includes books, journals, newspapers, and anything published publicly. However, there is another type of gray literature called *gray information* that has different acquisition requirements.

Usually the terms *gray literature* and *gray information* are used interchangeably. However, in the intelligence arena, they are slightly different. Gray literature refers to all publications that can be obtained from traditional bookstore channels, while gray information refers to other publications that cannot be obtained from traditional routes. Hence, gray information has its own channels, and it may be difficult to identify and acquire it. Gray information includes the following and more: academic papers, preprints, proceedings, conference and discussion papers, research reports, marketing reports, technical specifications and standards, dissertations, theses, trade publications, memoranda, government reports and documents not published commercially, translations, newsletters, market surveys, trip reports, and festival agendas.

Gray literature can be divided into three main kinds.

- *White*: This includes anything published publicly for sale through traditional bookstore channels. The publication should have an ISBN or ISSN and can be obtained directly from its seller. Books, journals, and newspapers fall in this category.
- *Ephemeral*: This type is short-lived. Examples include flight schedules, draft versions, copies of invoices, advertisements, posters, tickets, business cards, and anything that is self-published.
- *Gray*: This contains a mix of the previously mentioned two types.

Generally, gray literature can be obtained by paying subscription fees for such content or through buying books, journals, magazines, and other publications directly from bookstores. To acquire more hidden gray information, you have to use other specialized services. The following are some of them.

## Factiva

Factiva (<http://new.dowjones.com/products/factiva>) is a global news database with licensed content. It harvests data from more than 33,000 premium sources, and many of these sources (74 percent) are licensed and cannot be found freely online. Factiva collects sources in 28 languages in addition to its unique service of being able to provide access to resources that have not been published yet by their creators.

## LexisNexis

LexisNexis (<https://www.lexisnexis.com/en-us/gateway.page>) is currently owned by RELX Group (formerly Reed Elsevier). It originally focused on providing high-quality legal and journalistic documents, but it has expanded its coverage to include more services such as media monitoring tools, supply management tools, sales intelligence solutions, market intelligence tools, and risk solutions that analyze public and industry-specific content to predict risk and improve decision-making.

The following are other companies that specialize in gathering online intelligence from both public and private sources:

- InsideView (<https://www.insideview.com>)
- NewsEdge ([www.newsedge.com](http://www.newsedge.com))
- Semantic Visions ([www.semantic-visions.com](http://www.semantic-visions.com))
- DigitalGlobe ([www.digitalglobe.com](http://www.digitalglobe.com))

## Parties Interested in OSINT Information

OSINT can be beneficial for different actors. In this section, we will list them and explain what motivates each one to search for OSINT resources.

### Government

Government bodies, especially military departments, are considered the largest consumer of OSINT sources. The huge technological developments and widespread use of the Internet worldwide have made governments a huge consumer for OSINT intelligence. Governments need OSINT sources for different purposes such as national security, counterterrorism, cybertracking of terrorists, understanding domestic and foreign public views on different subjects, supplying policy makers with required information to influence their internal and external policy, and exploiting foreign media like TV to get instant translations of different events happening outside.

Intelligence agencies combine legally accessible information with their secretly acquired intelligence (for example, using spy satellite images, electronic listening stations, and spies) to answer a specific question or to predict the future. Those people have the required resources (money and equipment) to capture and analyze huge