

OSINT COLLECTION METHODOLOGIES

Date: November 20, 2023
Revision Date: N/A

Document Number: 20231120-01
Authored By: Practitioner Committee

BACKGROUND: The Practitioner Committee of the OSINT Foundation published a definitions document on November 28, 2022, entitled *OSINT Definitions* – document number 20221128-01, and a *Mediums of Publicly Available Information (PAI)* document on March 04, 2023 – document number 20230304-01. The legal and adopted definition of open-source intelligence (OSINT) is “*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.*”¹ The Practitioner Committee determined that delineating the OSINT methodologies was a critical baseline for the discipline.

APPLICABILITY: The following is applicable for OSINT practitioners in and supporting the U.S. Intelligence Community. The information contained herein is instructive, but not necessarily applicable to non-U.S. Intelligence Community activities.

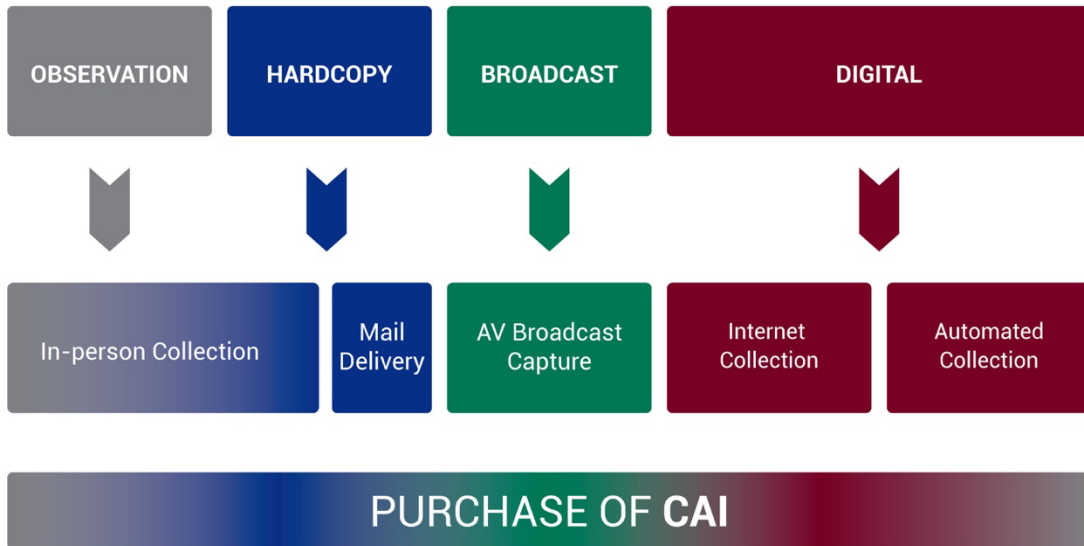
ASSUMPTIONS: Collection is one part of the overall Tasking, Collection, Processing, Exploitation, and Dissemination (TCPED) process within the OSINT discipline. Data, once collected, nearly always requires some degree of processing and exploitation, to include translation, formatting, or data engineering, to place it into a useful format for analysis.

At this time, it is too early to assess the benefits and challenges of artificial intelligence (AI) as a capability. Current generative AI tools appear largely to benefit the processing and exploitation of PAI. The use of AI as a collection methodology is not yet mature.

For the purposes of this document, the terms purchase and acquisition are synonyms.

¹ Public Law 109-163, Section 931 on January 6, 2006.

PAI CATEGORIES



OSINT COLLECTION METHODOLOGIES

Copyright 2023, OSINT Foundation, Incorporated, free for use with proper attribution.

METHODOLOGIES

The order of the following methodologies follows the previously articulated *Mediums of PAI* order² and should not be construed as being listed in order of priority or prevalence.

In-Person Collection: An OSINT collection methodology that involves observation³ or direct gathering of PAI from public events and spaces. Information may be acquired while attending public meetings, forums, workshops, lectures, exhibits, parades, conferences, trade shows, conventions, press conferences, and other public events. The information collected may come in various forms, including observation or meeting notes, reports, policy statements, rules and regulations, schedules, questionnaires, slides, surveys, plans, bulletins, brochures, flyers, and circulars. An individual collecting in-person information may retrieve the documents in hard copy or take notes, photographs, videos, and audio recordings of the PAI.

² See OSINT Foundation, *Mediums of PAI* document 20230304-01a, dated October 16, 2023.

³ Ibid

Mail-Order Delivery: A collection methodology that involves the collection or acquisition of hardcopy PAI materials, including books, journals, newspapers, maps, graphics, reports, policy documents, rules and regulations, schedules, questionnaires, slides, surveys, plans, bulletins, brochures, flyers, circulars, and other formally published materials. Mail-order delivery of digital materials is possible but requires use of digital media, such as CDs, DVDs, or flash storage devices.

Audio/Video Broadcast Capture: A collection methodology that involves collection of audio and video content from non-internet broadcast sources, including television, radio, ham radio, and loudspeaker⁴. Audio/video collection may require compression tools (in a format accessible by customers) for lengthy digital captures. Examples include monitoring international news and television channels, radio broadcasts, or public speeches and announcements.

Internet Collection: The process of gathering PAI from various online and internet-collected sources for OSINT purposes. Collectors may need to make use of specialized tradecraft and technology to obfuscate the interests, identity, and search techniques of the collector, referred to collectively as managed attribution (MA). Practitioners may collect PAI from sources within the surface, deep, or dark webs.

- The surface web includes the small portion of the World Wide Web that is both easily accessible and easily searchable on search engines. Examples of surface web sources could include social media platforms, online forums and blogs, business websites, and public databases.
- The deep web includes the large portion of the World Wide Web that is accessible but is not easily searchable on search engines. Examples of deep web sources could include databases, login-protected websites, dynamic content, private networks and social media forums, and any other sites the search engines do not index.
- The dark web includes a small, hidden portion of the Internet that can only be accessed using specialized software and is not easily searchable on search engines. Examples of dark web sources could include illicit marketplaces, underground forums, illicit file-sharing sites, or discrete communications platforms.

Surface Web

Easily searchable and navigable

Deep Web

Not easily searchable but still navigable

Dark Web

Not searchable and not navigable without specialized software

⁴ Ibid

Automated Collection: A collection methodology that uses technical applications to collect PAI while minimizing human involvement in the collection activity. This approach allows for rapid and comprehensive data gathering, as well as the ability to process large volumes of information, but may suffer from issues with the accuracy and reliability of collected data, adversary denial and deception, potential biases in the algorithms, and compliance with intelligence oversight and privacy regulations. Automated collection may also involve MA.

Examples of Automated Collection for OSINT purposes include data scraping, where software automatically extracts data from websites or other online sources; data downloading, which involves automated retrieval of dynamic files or datasets from online repositories; and obtaining data automatically through application program interfaces (APIs) or other data-sharing mechanisms. Social media monitoring is another example of automated collection, where tools are used to track, analyze, and extract valuable insights from user-generated content, discussions, and trends across various social media platforms which may or may not have associated geolocation information.

Purchase of Commercially Available Information (CAI): A collection methodology that involves acquiring information, including intelligence, from a variety of commercial sources. OSINT organizations must leverage existing authorities when acquiring CAI that supports their assigned missions, functions, and tasks. OSINT organizations must establish information handling practices in line with statutory and policy data governance frameworks appropriate to their mission area and based upon their authority for access or collection, the type of data sought, the collection mechanism, and the purpose and use of the data.

CAI excludes proprietary information not available for purchase by the public.⁵ Examples of the purchase of CAI range from simple acquisitions like books or maps from a bookstore, to online subscriptions for periodicals, to the acquisition of specific data sets or services.

ADDITIONAL CLARIFICATION:

These methodologies listed above may not be mutually exclusive; an OSINT practitioner may combine multiple collection methods to achieve a comprehensive understanding of a subject. Leveraging diverse data sources and approaches can enhance the accuracy and reliability of the collected information, providing a more complete picture that overcomes the inherent limitations of any single approach.

⁵ See OSINT Foundation, *OSINT Definitions* document 20221128-01, dated November 28, 2022.

Gray Literature: Gray literature⁶ can be collected using several OSINT collection methodologies, but is most closely associated with the Purchase of CAI, in-person collection, mail delivery, and internet collection methodologies. These sources can be valuable for OSINT practitioners as they may contain unique insights and information that may not be readily available through other means.

IoT Telemetry: An emerging area of interest in OSINT collection is telemetry data from devices and machines.

- Internet of Things (IoT) refers to the network of interconnected devices featuring sensors, software, processing capabilities, and other technologies that communicate with each other and the cloud. The cloud, in this context, refers to servers accessed over the internet that host software and databases in data centers located worldwide.
- IoT telemetry refers to the data generated by these connected devices (such as urban traffic monitoring data or data from coastal Automatic Identification System receivers) that are available for free or by purchase or subscription from commercial vendors. There are many emerging use cases for publicly available IoT telemetry data in OSINT, particularly in terms of automated collection, for gaining unique insights into patterns of life, operational environments, and real-time events.
- IoT telemetry may be available for internet collection, automated collection, and purchase of CAI.

Bilateral Exchanges: U.S. OSINT organizations often establish and develop OSINT sharing relationships with counterpart organizations in foreign governments. Foreign partner investments in OSINT are often greater and, in many cases, more mature in tradecraft, source development, language capability, and cultural understanding than typically found in U.S. intelligence organizations. The U.S. has a great deal to gain by engaging with partners as peers and sharing PAI/CAI data in multi-lateral constructs, subject to terms and conditions of licenses or subscriptions.

⁶ Ibid