Introduction to OSINT

CRAAP Methodology

- Currency: the timeliness of the information.
- Relevance: the importance of the information for your needs.
- Authority: the source of the information.
- Accuracy: the reliability, truthfulness, and correctness of the content
- Purpose: the reason the information exists.

- *CRAAP applied to a website (Exercise)*

OSINT

So, What is Open-Source Intelligence (OSINT)?

Information that is freely available

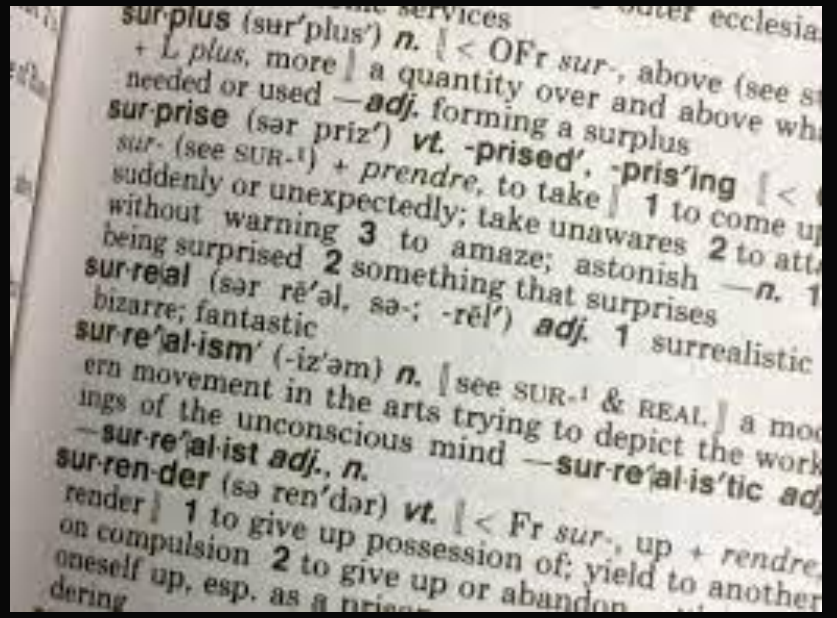
AND

can be used to support policy, plans, or operations

What is OSINT?

<u>According to U.S. public law</u>, Open-Source intelligence:

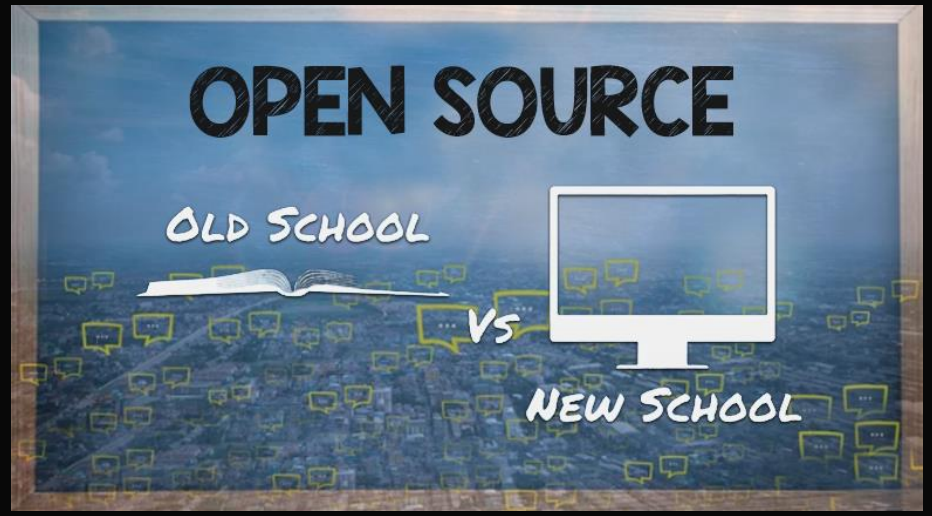
- Is produced from publicly available information
- Is collected, analyzed, and disseminated in a timely manner to an appropriate audience
- Addresses a specific intelligence requirement



OSINT HISTORY

- Pre WWII, the Research and Analysis Branch of OSS was dedicated to OSINT.
- Collected print and radio worldwide to gain intel about US' adversaries that could reveal their intentions and movements.
    - No digital assistance; still OSINT.

- After WWII, the intelligence community stepped back from OSINT,
    - humINT (human intelligence) and sigINT (signals intelligence) preferred.
    - Cold War
- "Green Revolution" in Iran in 2009 sparked its comeback.



Who Uses OSINT?

- ❖ Governments,
- ❖ Law Enforcement,
- ❖ Military
- ❖ International NGOs,
- ❖ Businesses,
- ❖ Cybersecurity & Cybercrime Groups,
- ❖ Privacy Conscious People & Organizations,
- ❖ Terrorist Groups

"By some estimates, more than 80% of what a U.S. president or military commander needs to know comes from OSINT, and not from foreign agents, spy satellites or expensive eavesdropping platforms."

- Rise of Open-Source Intelligence Tests U.S. Spies. WSJ. Dec. 11, 2022

OSINT is Used to Investigate:

- **Cyber-enabled threats**
  - Credit card fraud
  - Money laundering
  - Counterfeiting
  - Theft and gift card fraud
  - Workplace harassment
  - Insider threats

- **Physical security threats**
  - VIP-targeted doxxing and harassment (SWATTING)
  - Travel risk management
  - Event monitoring
  - Crises like terrorism and natural disasters

How is OSINT Used Commercially

- Gaining Competitive Edge
- Executive / Employee / Vendor Screening
- Know Your Customer
- Market Surveys
- IP Protection
- Loss Prevention
- Data Leak Detection
- Fraud Investigations
- Crisis Response
- Executive Protection
- Event Security
- …

"So they do some work, long story short, about 22 hours later through that very building, three JDAMS take that entire building out. Through social media. It was a post on social media. Bombs on target in 22 hours.

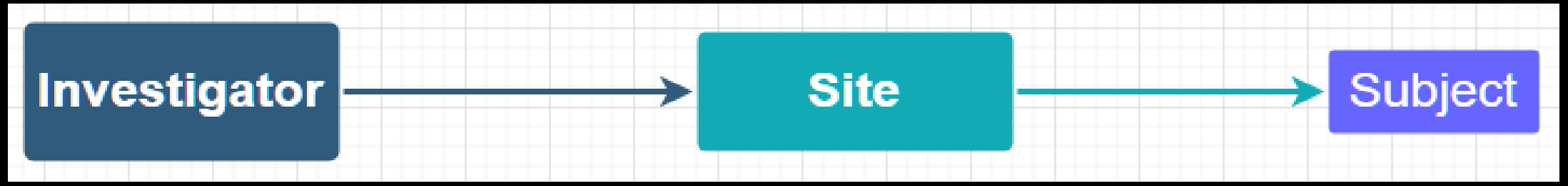
- Gen. Hawk Carlisle, commander of Air Combat Command, June 1, 2015

Types of OSINT

Passive OSINT

Access publicly available resources using non-technical means

- Example: materials published by a 3rd-party (databases, newspapers, etc.)
- Example: collecting the public posts of a social media user
- Goal: Subject does not see your activity



Semi-Passive OSINT

Sends traffic directly to servers to gather information about them

- Example: Visiting the org's website
- Example: Looking at metadata in published docs
- Example: Using “Builtwith” to identify the tool stack behind a website

Goal: Blend in with the regular traffic. Subject can see the activity, but cannot attribute it to a particular source

Active OSINT

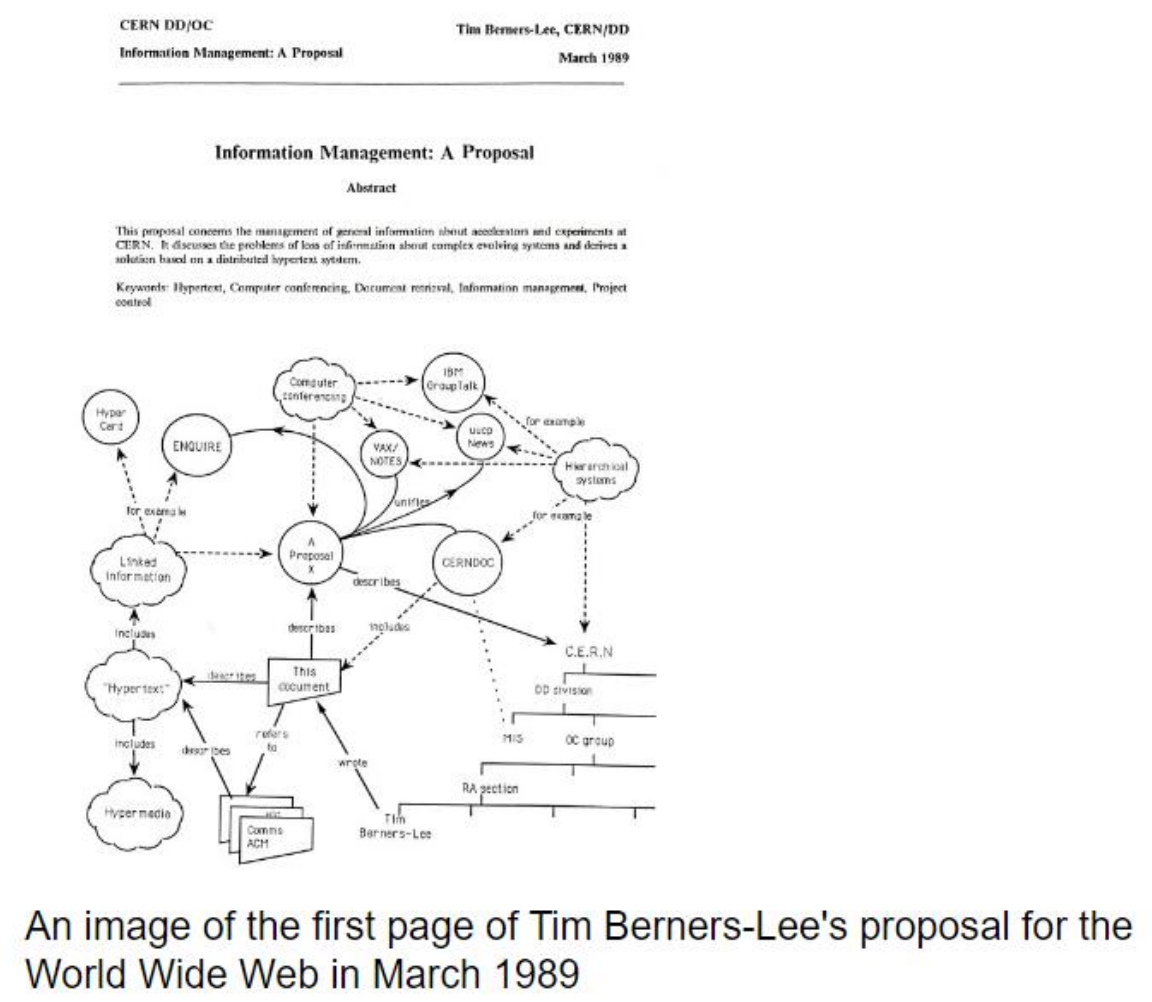
Interact directly with the server to gather info

- Example: a port vulnerability scan using netcat or nmap
- Example: Searching for "unpublished" servers or directories
- Example: Some Maltego transforms
- Example: Entering the physical building
- ***Average users don't do this – it will stand out in the logs and draw attention to the researcher and their actions.***

(Spoiler: He gets caught. Eventually.)

The Internet

Proposing the Web..



CERN DD/OC — Tim Berners-Lee, CERN/DD

Information Management: A Proposal — March 1989

Information Management: A Proposal

Abstract

This proposal concerns the management of general information about accelerators and experiments at CERN. It discusses the problems of loss of information about complex evolving systems and derives a solution based on a distributed hypertext system.

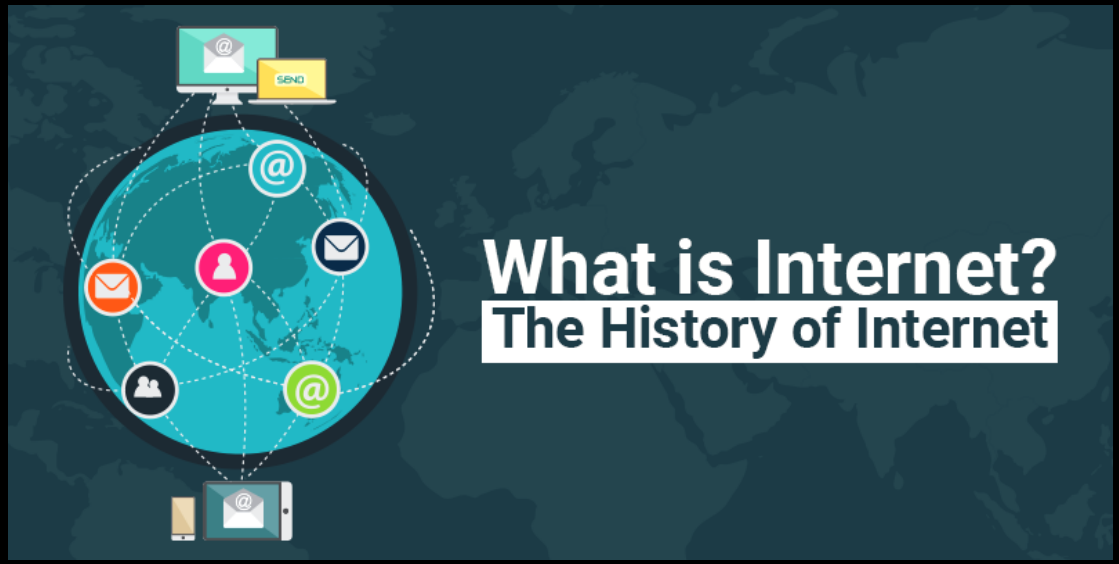
Keywords: Hypertext, Computer conferencing, Document retrieval, Information management, Project control

An image of the first page of Tim Berners-Lee's proposal for the World Wide Web in March 1989

..but first..

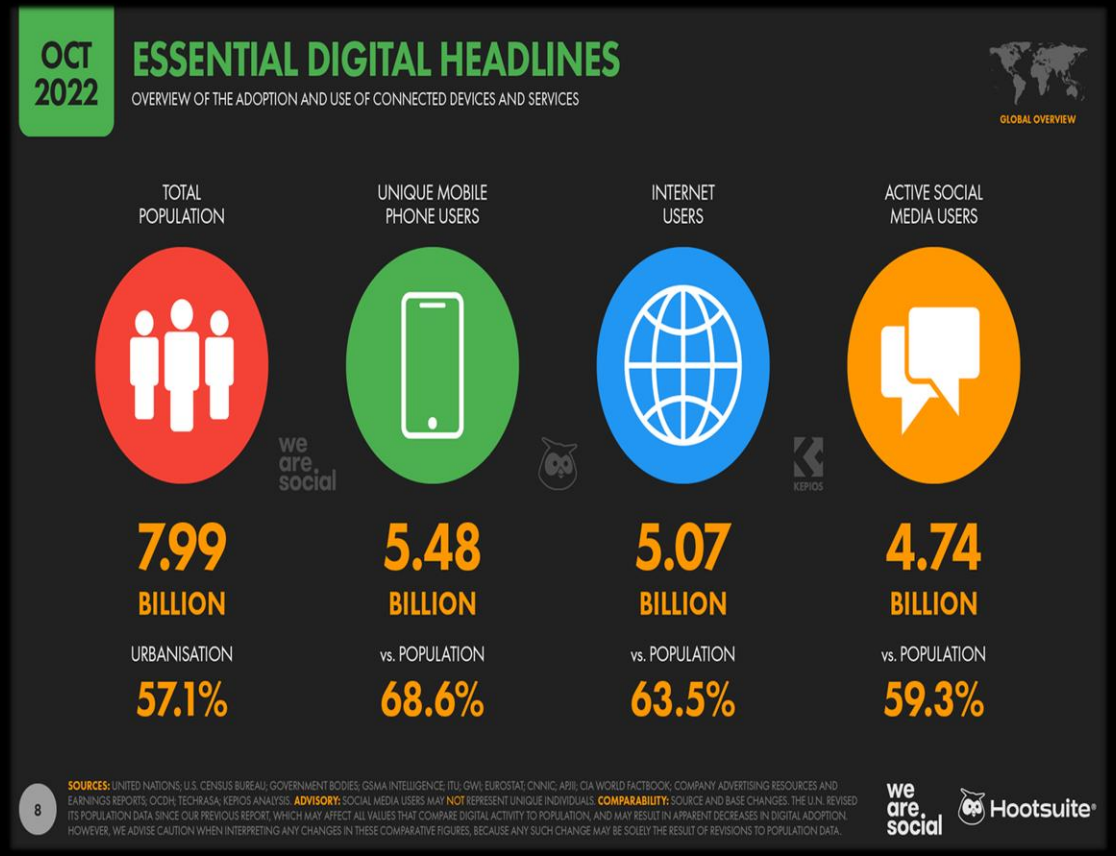
Timeline

1957 – USSR launches Sputnik, so US creates ARPA
1962 – J.C.R. Licklider proposes "intergalactic network" of linked computers in case of Soviet attack on government phone systems
1969 – ARPAnet's 1st four computers get connected.
1972 – 1st demo of "electronic mail"
1973 – ARPAnet goes global. England & Norway connect
1979 – Birth of Social Media: Usenet starts.
1990 – Tim Berners-Lee develops HTML at CERN
1991 – First webcam – Engineers hook up a camera facing the coffeepot so they can see when it's empty
1993 – The web gets graphical: Marc Andreesen develops Mosaic
1994 – US privatizes Internet's backbone, Netscape, Yahoo & Internet Explorer born. First spam goes out - (2 lawyers send an ad for green card services to almost every Usenet group, regardless of relevance.)
1998: Google launches
2001: Wikipedia launches
2004/5: Facebook and YouTube launch
2006: Twitter and iPhone are born



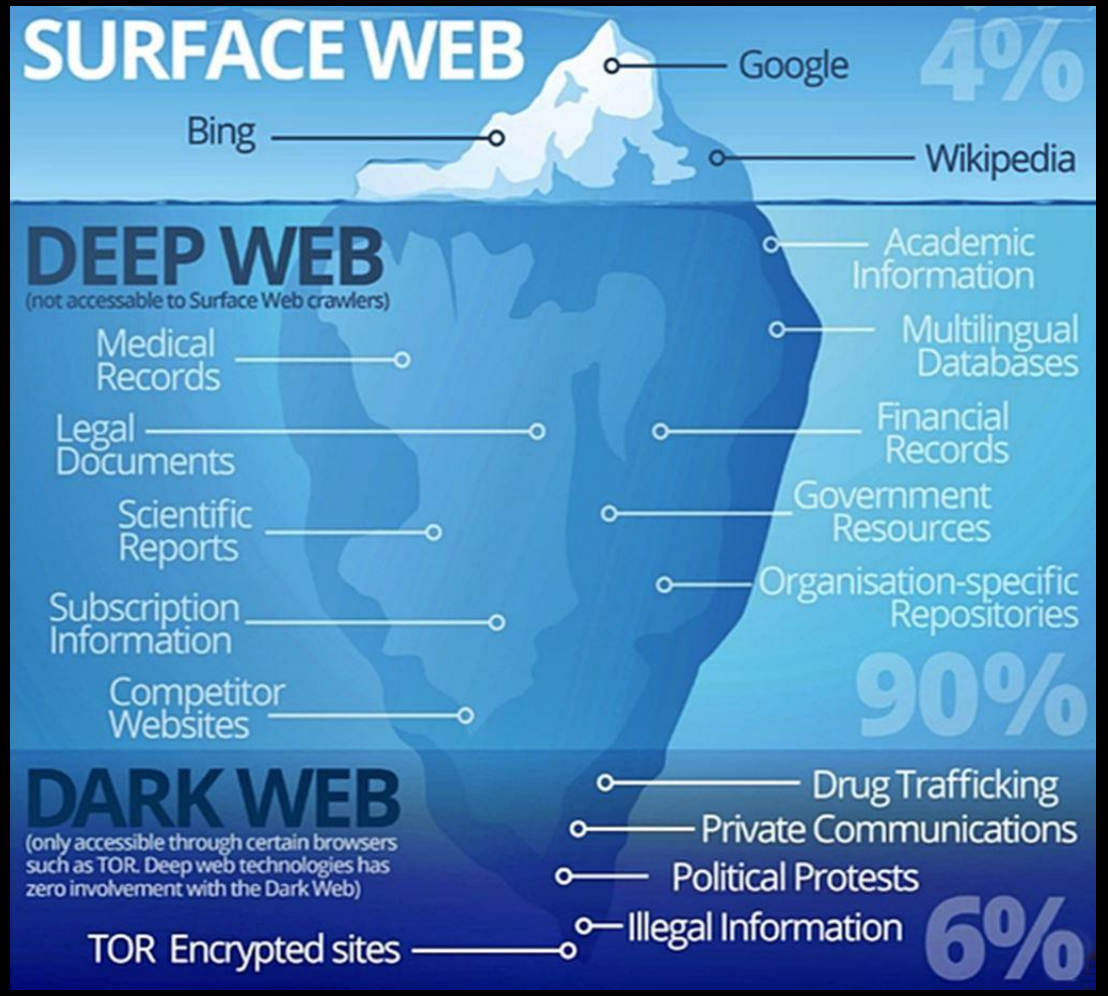
Now the World is Online…

- A total of 5.07 billion people around the world use the internet today – equivalent to 63.5 percent of the world's total population.
- Internet users continue to grow too, with the latest data indicating that the world's connected population grew by more than 170 million in the 12 months to October 2022.
- There are now fewer than 3 billion people who remain "unconnected" to the internet, with the majority of these people located in Southern and Eastern Asia and in Africa.



Surface/Deep/Dark Web

- Google only sees the Surface.
- Over 90% of the Internet lives in the "**deep web**" - That's 5 million **terabytes** of data, or 5 trillion **megabytes**.
- Google's indexed 4% of it - **Over 96% of the web is unindexed**
- The net grows exponentially – Google is now focusing on organization of info, as opposed to indexing every page.



Surface Web

- Also called the "clear web"
- The part of the web that's indexed, accessible through:
  - Duckduckgo
  - Yahoo
  - Google.ly (Libya)
  - Yandex
  - Baidu
- Websites
- Ecommerce
- Blogs

Search Terms & Boolean Operators

AND	**Searches for all the search terms you specify**	**Amazon AND Rainforest for websites that include both terms**
OR	Searches for one term or another	**Putin OR Russia** if you'd like results on either term but not necessarily both
" "	Group phrases with quotation marks for results that only have those words together	"Vladimir Putin"

DO Capitalize Boolean Operators
Combine them! "Vladimir Putin" AND Russia

("Vlad" OR "Vladimir" OR "Vladmir") AND Putin --> helpful if the correct spelling is unknown

Dorking, or "Advanced Search Techniques"

"How to cut your search time in half"

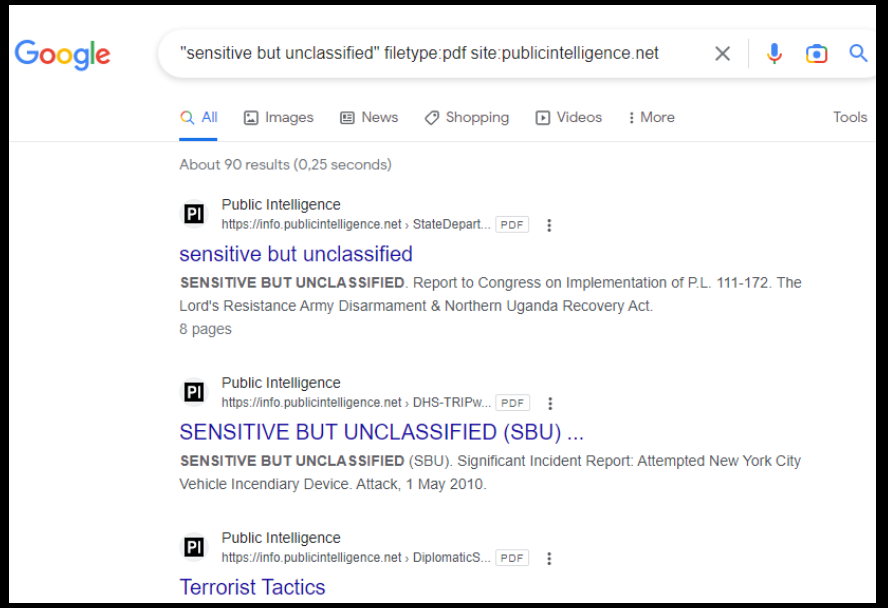
Short List of Dorks		
Dork	**Purpose**	**Example**
site:	Focuses search on the site	site:theguardian.com Putin
cache:	Looks for a cached version of a page	Cache:theguardian.com/world/vladimir-putin
ext:	Finds files on a subject with a particular extension - .doc, .pdf, .xls, xlxs and more	ext:pdf Putin OR site:brookings.edu ext:pdf Putin
filetype:	Same as above – sometimes different results	filetype:pdf Putin OR site:brookings.edu filetype:pdf Putin

Dorking, or "Advanced Search Techniques"

"How to cut your search time in half"

A Model Dork: "____" filtetype:___ site:___

Example: "sensitive but unclassified" filetype:pdf site:publicintelligence.net



Deep web

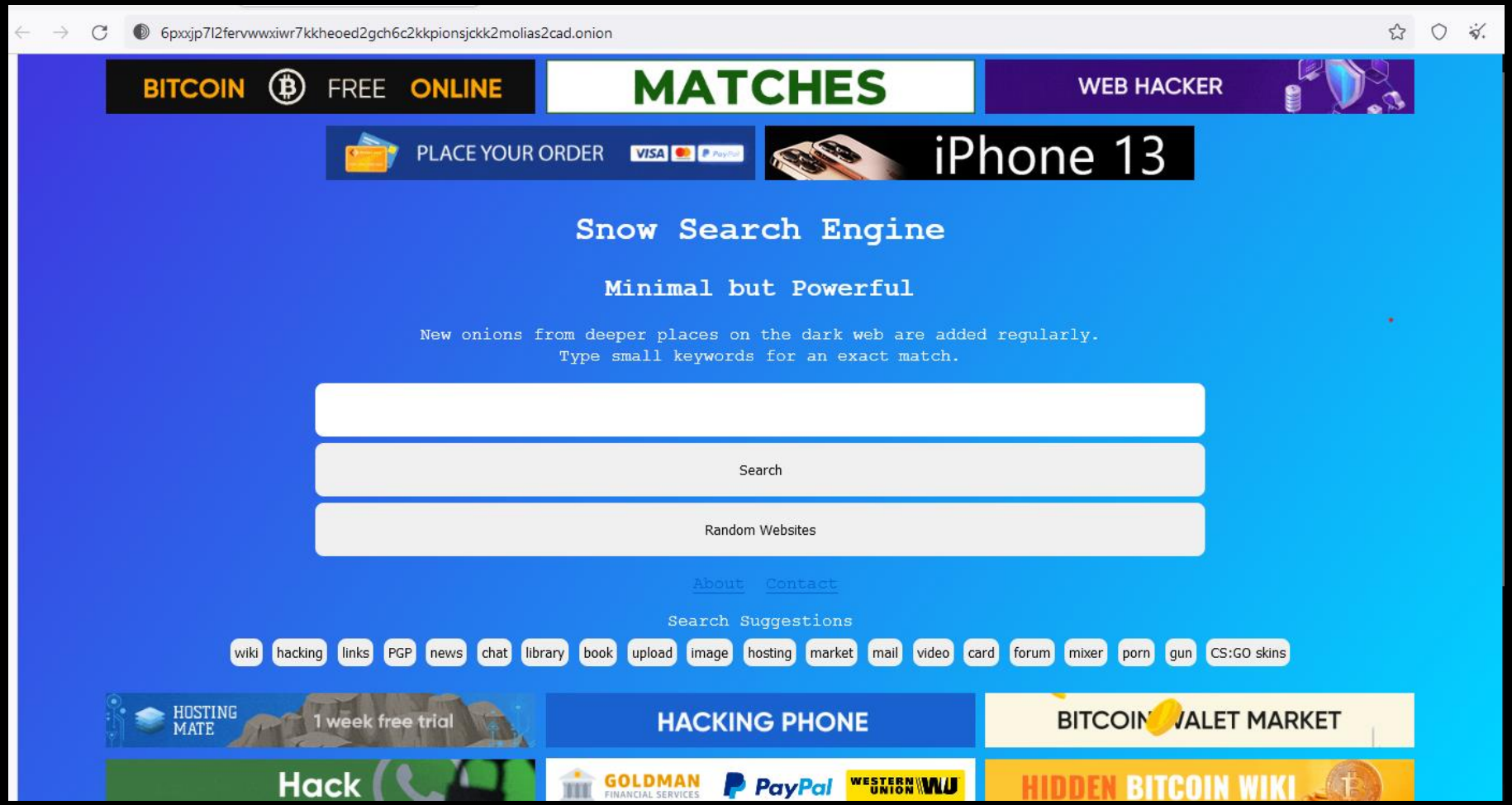
- It is legitimate. It is noncriminal. You use it every day.
  - Email messages, chat, private content on social media, government resources, scientific reports, legal records, academic databases, and other internet accessible content
  - Typically, sites that require a password or other security access to pass.
- Google may be able to get the researcher to a "front door" - i.e., a search interface, but may not be able to search the database itself.
- Other tools may help:
  - Data.gov
  - WorldCat
  - Internet Archive
  - Shodan

Deep Web Search Strategy

Look for the hidden doorway to the content – not the content itself

- Databases
- Guides, portals and directories
- Associations, foundations, or fan organizations
- University research centers and research institutes –
- Think tanks
- Experts:
  - In the community
  - Through local news archives
  - Scholarly publications (Google Scholar, Web of Science)
- Government & Law Directories/Lists

Searching the Dark Web on Tor



But, also:


THIS HIDDEN SITE HAS BEEN SEIZED
by the Federal Bureau of Investigation,
as part of a coordinated law enforcement action taken against the NetWalker Ransomware.
DEPARTMENT OF JUSTICE
FEDERAL BUREAU OF INVESTIGATION
НСлС
The action has been taken in coordination with
the United States Attorney's Office for the Middle District of Florida and
the Computer Crime and Intellectual Property Section of the Department of Justice,
with substantial assistance from the Bulgarian National Investigation Service
and General Directorate Combating Organized Crime.

Dark Web?

- Anonymity. They are not all illegal or nefarious.
  - Protect users' identities (within certain parameters and with precautions)
  - Legitimate use:
    - Anonymous information exchange, whistleblowers, general privacy, some newspapers have .onion sites
  - Illegitimate use:
    - Black markets, drugs, arms, stolen or counterfeit goods, data, human trafficking, CSAM, and more
- Tools: VPN, Tor (for .onion sites)
- Self-contained networks (peer-to-peer): I2P, Freenet, Zeronet

Dark Web Search Strategy

- Access:
  - Tor (for anonymity and access) an VPN (for privacy) together
  - A tool such as Authentic8, which has a built-in VM and VPN
- Searching from the Clearweb
  - Ahmia.fi
  - Dark.fail - (Is a .onion site online)
- Use common sense – do not trust anyone or anything
- Do not download anything into your regular work VM
- Do use a sock puppet (a fake online identity) if you must login somewhere or speak to someone
- Use a new/burner email account

Using OSINT to support HUMINT

OSINT provides background context, inspires questions, and helps verify Subject answers.

<u>Pre-Interview questions</u>:

- What are the current public activities of the Subject?
- What is known about the Subject?
- Who in the Subject's network can corroborate info?
- Who are Key Players in the Subject's sphere?
- Does Social Media/Press/Other Media support or contradict the Subject's account?

Table Game

The Subject was in Hollywood, Florida on June 7th, 2015, and during the interview, made comments about the weather.

What was the temperature at 1 PM that day in Fahrenheit?

Answer: we are looking for the number of temperature in Fahrenheit that it was at 1 PM on June 7th, 2015.

Questions?