universidade
de aveiro

Computer Systems Forensic Analysis
AFSC

Open Source INTelligence

Artur Varanda

School Year 2023-2024

History

Classical sources of information and intelligence

HUMINT Human Intelligence – intelligence gathered by means of interpersonal contact, typical activities consist of interrogations and conversations with persons having access to information

SIGINT Signals Intelligence – intelligence-gathering by interception of signals, whether communications between people or from electronic signals not directly used in communication

SIGINT (Signals Intelligence) can be subdivided into:

COMINT Communications Intelligence – deals with messages or voice information derived from the interception of communications between people

ELINT Electronic signals intelligence – intelligence-gathering by use of electronic sensors

FISINT Foreign instrumentation signals intelligence – telemetry, tracking systems, video data links, and arms control

IMINT Imagery intelligence – collects information via satellite and aerial photography

MASINT Measurement and Signature Intelligence – analysis of data obtained from sensing instruments for the purpose of identifying any distinctive features associated with the source, emitter or sender, to facilitate the latter's measurement and identification. This often includes radar intelligence, acoustic intelligence, nuclear intelligence, and chemical and biological intelligence

Evolution of OSINT - Open Sources Intelligence

- Since the 1930s, at the University of Princeton, the Foreign Broadcast Information Service (FBIS) collected information serving as an intelligence source in World War II and the Cold War;

- OSINT terminology was introduced by Americans in the mid-1970s and in Europe in the 1980s given the importance of criminal analysts in combating crime, including organized crime;

- In 2005, after the failure of September 11, CIA opened an Open-Source Center (OSC). However, it was discredited, because it was believed that non confidential information was not valuable.

Information Sources

Information sources - Classification

Origin

- Primary information source is a place where original information resides (usually true)
- Secondary information source is a document or record that relates or discusses information originally presented in primary information sources (needs to be verified)
- Tertiary information sources are made up of lists or summaries of primary or secondary sources, such as bibliographies, lists of readings, or articles on research

Information sources - Classification

Authority

- Closed Source Information
- Open Source Information

Closed source of information

- A closed source involves obtaining a judicial authorization or formal explicit or implicit authorization through the delegation of powers
- Is outside the legal limits of collecting information on its own initiative
  - to maintain information privacy
  - to avoid the inadmissibility of its use
  - to avoid civil, disciplinary and criminal liability over the expert or his organization

Closed source of information – Portuguese examples

Institute of Registries and Notaries (IRN)

- civil identification
- car registration
- collective entities
- criminal record
- land registry
- visa applications
- trademark registration
- registration of religious entities
- foundations and associations
- judicial power of attorney certificates
- marriages, divorces and sex changes
- nationality attributions and losses
- wills and public scriptures
- *etc*

Closed source of information – Portuguese examples

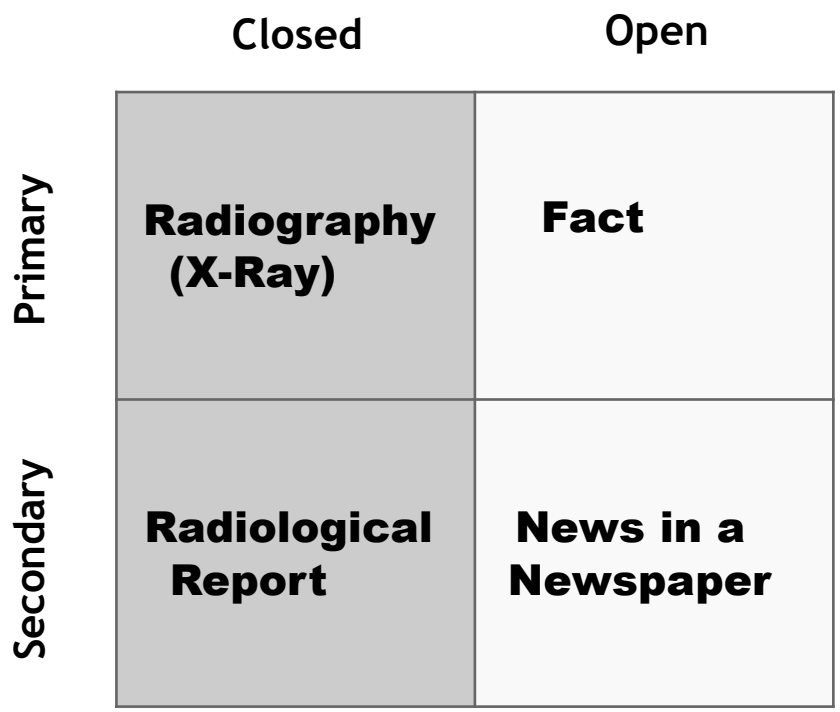
Tax Authority (AT)

- tax registration
- taxable assets
- household
- taxed IRS
- billing
- taxed IUC, *etc*

Institute of Mobility and Transports (IMTT)

- driver identification
- automotive inspections
- automotive booklets
- *etc*

Open vs. Closed sources



Legal Authorization Required

Open source information is said to be open if:

- it is fully accessible by third parties
- can be of individual or collective origin
- can be collected and processed automatically or manually

Open source categories

- **Traditional Media** newspapers, magazines, radio, and television
- **Internet** communities and user-generated content (social networks, video sharing networks, wikis, chats and blogs)
- **Public Data** official data from governmental and other organizations, conferences, speeches, from companies
- **Observation and Reporting** data collected by specialized citizens
- **Pictures, Videos and Sound** maps, satellite images
- **Professional and Academic** greyliterature, reports and articles

Advantages

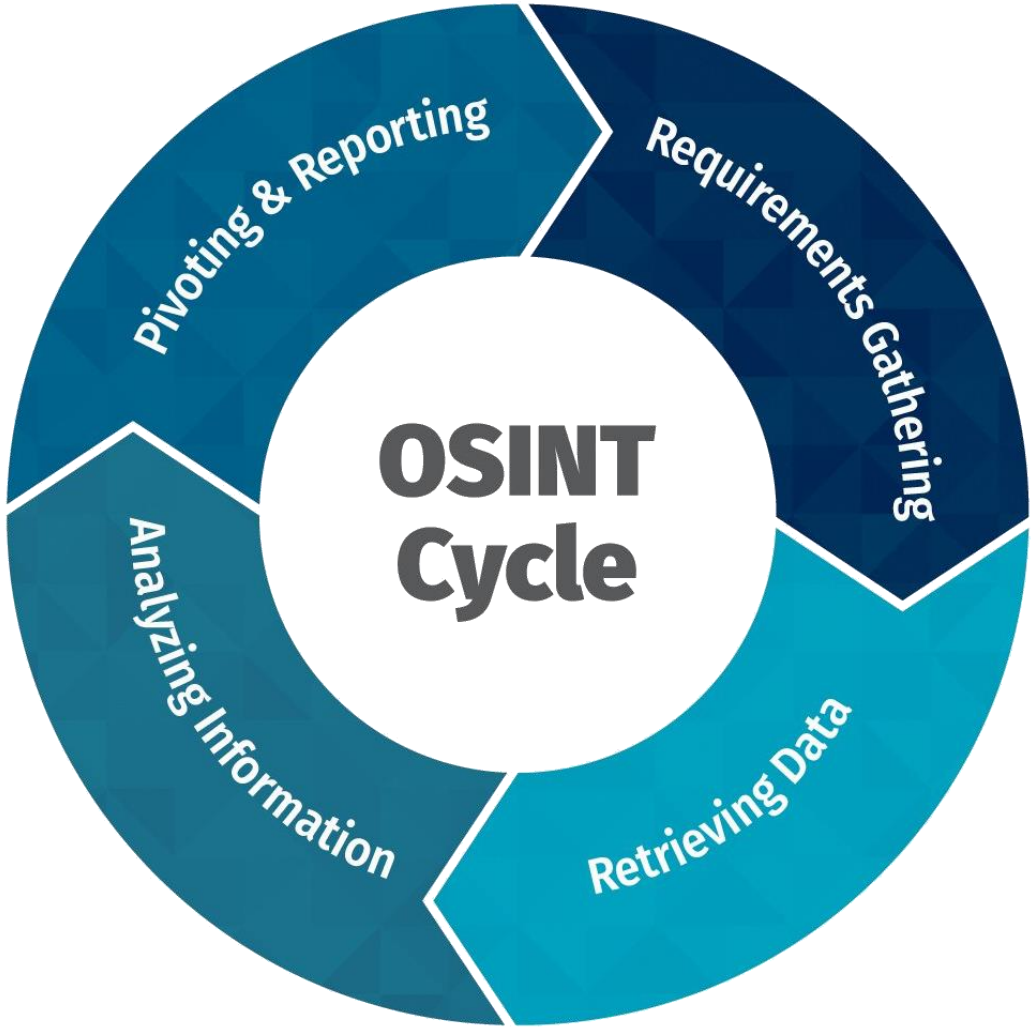
- open sources are less costly and quick to access
- open sources provide information beyond closed sources and in greater quantities
- open sources can reduce the need for the production of classified intelligence and complement the latter
- does not compromises the purposes of the investigation
- open sources are a contextual element for classified intelligence operations

Disadvantages

- open source information does not replace classified information due to its intentionally secretive nature
- can be subject to misinformation or arbitrariness and therefore generally require a validation process
- need experts in various domains, this is more relevant if foreign languages are involved
- the amount of information available on the Internet is usually excessive, this implies the use of specific tools

Information to Intelligence Cycle

Information to Intelligence Cycle



Advantages of Open Sources at the strategic level

- can help identify strengths, threats, risks and opportunities for sustainability and long-term *e. g.* growth in the number of refugees in Europe
- important in the cultural, demographic and geographical components in assessing opportunities and alliances
- essential to contextualize internal strategic information by comparison and benchmarking

Advantages of Open Sources at the operational level

- important to define and subsequently contextualize operations at the human and geographical level, such as distribution

  e. g. Google Maps, Bing Maps

- very important in coordinating joint commercial or other international operations where classified information does not exist

  e. g. hotel information, habits, *etc*

Advantages of Open Sources at the tactical level

- enabling current and imminent information is an important resource for addressing medium-term priorities and assessing the individual reliability of forces
  e. g. new products, new EUincentives, *etc*

- can provide good information about the human context for medium term operations
  e. g. current demographic, cultural, *etc*

Advantages of Open Sources at the technical level

- it's an important resource for assessing the effectiveness of forces at the technical level *e. g.* information about computer resources, level of automation, *etc*
- it allows to collect information about the systems, transport, communications and even financial context

Skills of the Analyst

Discipline in collecting due to excessive information

Time	Task	Description
15 minutes	Requirements definition	Ensure an understanding of commander's intent
30 minutes	Internet Collection	Use search tools, rapidly identify top ten sites and review
15 minutes	Resources' Table	Create Resources' Table for future use and for customer's reference
60 minutes	Commercial Collection	Use fee sources, identify top 20 items for exploitation
60 minutes	Analysis	Read, understand, evaluate, and structure collected information
60 minutes	Production	Carefully create an analytical summary, table of contents, and slides

4 hours – Total time to produce an open source analysis report using only internal sources

Source: NATO OSINT manual

OSINT process

1. **know who knows** have in-depth knowledge of the available sources' characteristics
2. **know what's what** ability to evaluate and assess the validity, scope, degree of accuracy and timeliness of the requirements
3. **know what's hot** ability to distinguish what is important and relevant
4. **know who's who** ability to distinguish between facts and speculation and avoid cognitive bias from the sources

Source: NATO OSINT manual

Cognitive deviations (bias)

- systematic error in thinking that affects the decisions and judgments that people make
- individuals create their own *subjective social reality* from their perception of the input
- may lead to perceptual distortion, inaccurate judgment, illogical interpretation, or what is broadly called irrationality

Important

The analyst should be able to avoid cognitive bias from the sources and from his own education, origin, religion, culture and profession

Open Source Possibilities

Traditional Media Sources

- allow to establish chronologies
- contextualize other sources of information
- update other information
- requires validation

Internet Sources

- provide personal information
- allow in-depth research
- can be worked automatically
- allows data mining and pattern identification
- requires validation

Public Data Sources

- provide specific information about organizations
- may provide important strategic information
- important at the strategic level
- does not require validation

Report Analysis Sources

- provide very specific information about organizations
- fill in gaps from other sources
- requires specific knowledge
- does not require validation

Professional and Academic Sources

- provide important technical and scientific information
- essential at the technical level
- allow validation of other sources
- requires specific knowledge

Automated Processing

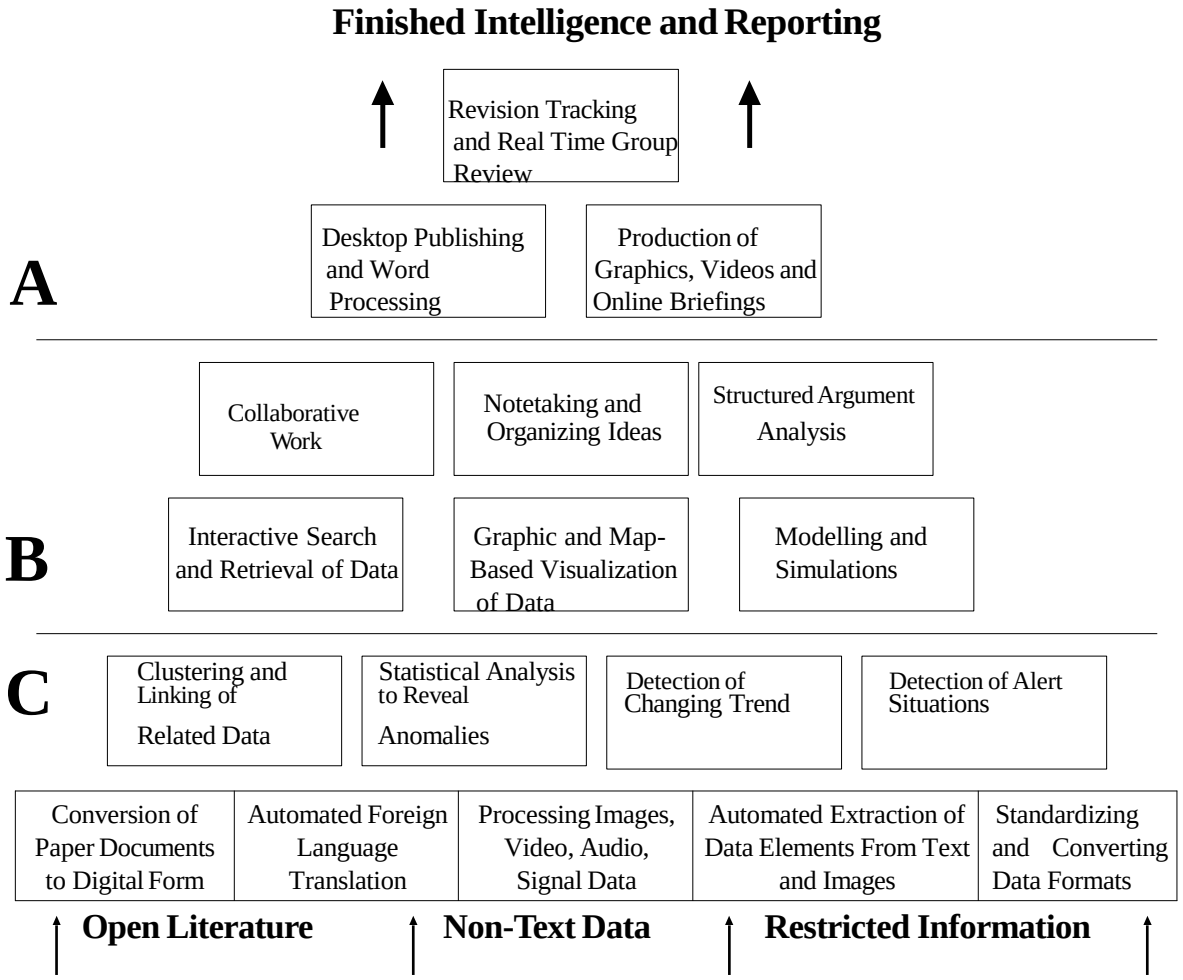
OSINT Software Hierarchy

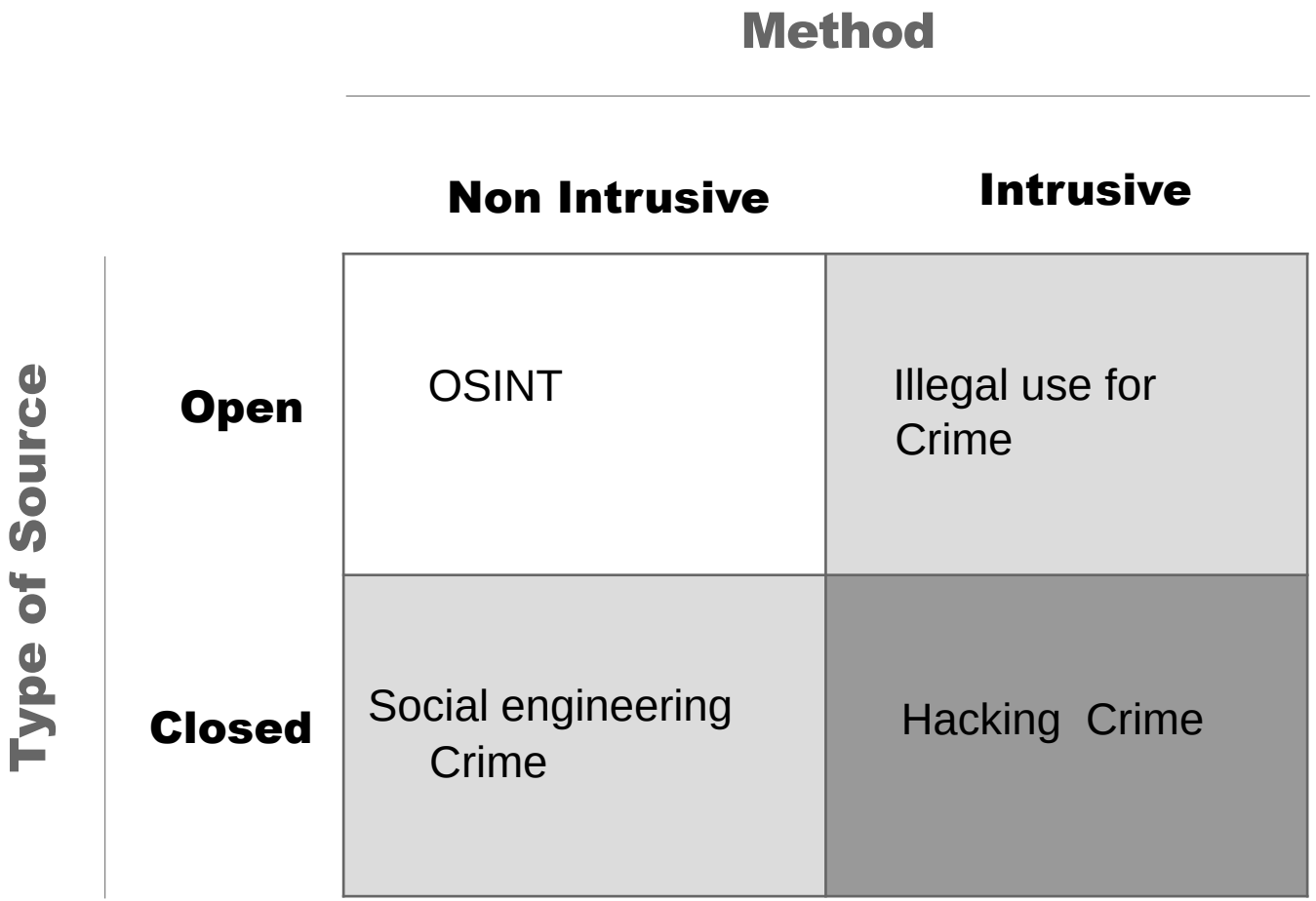
Software functionalities for a optimal OSINT toolkit:

A Dissemination – publishing and production management functionalities

B Analysis – combines collaborative work tools with data visualization and manipulation tools with thinking tools

C Automated pre-processing

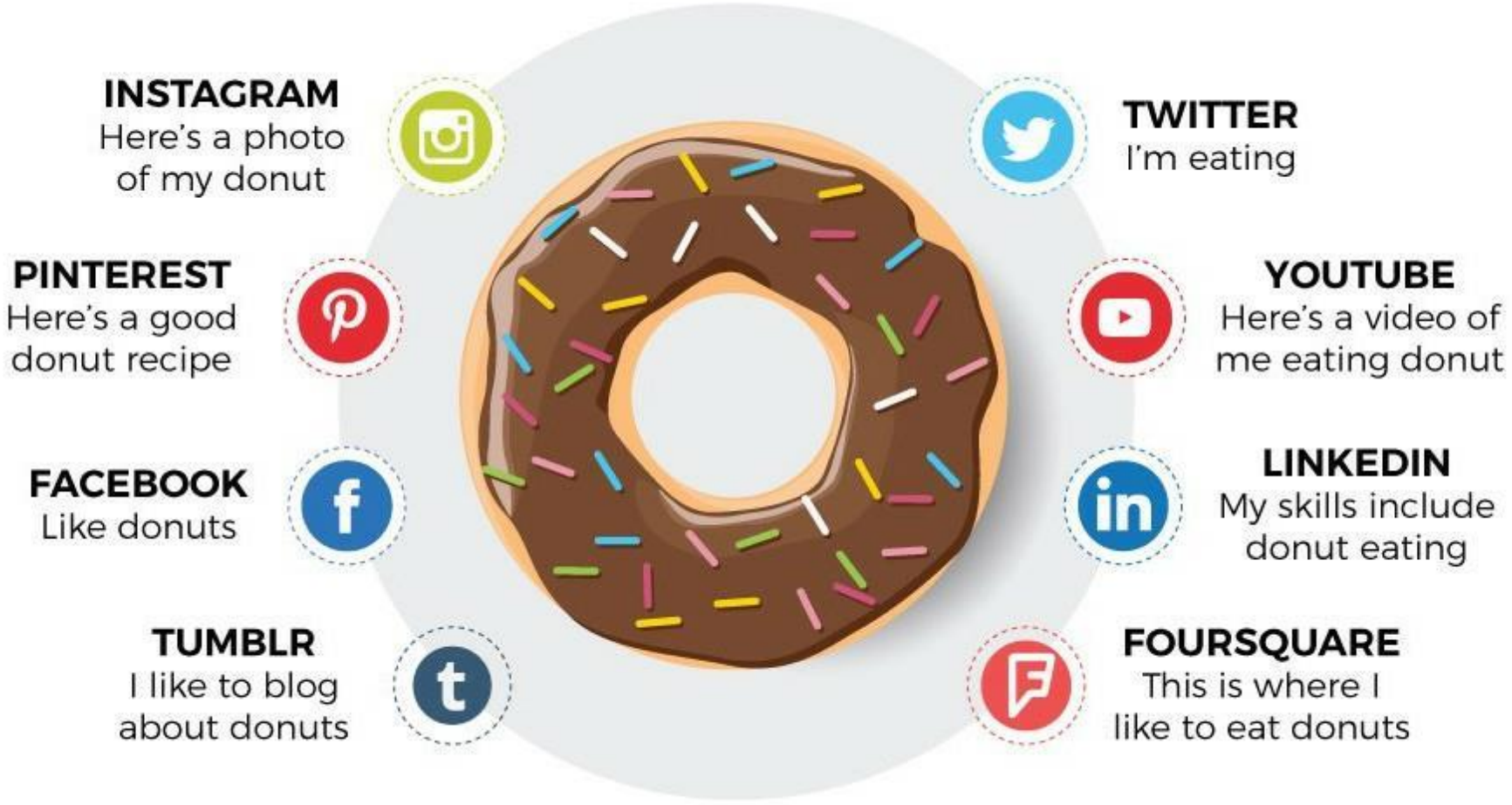



Method
Non Intrusive
Intrusive
Type of Source
Open
Closed
OSINT
Illegal use for Crime
Social engineering Crime
Hacking Crime

Resources

- crawlers
- analytics
- social media searches
  - Facebook, Twitter, YouTube, Instagram, FourSquare, *etc*
- digital footprint, anonymization, IP address as ID
- Application Programming Interface (API)
- darkweb

Social media

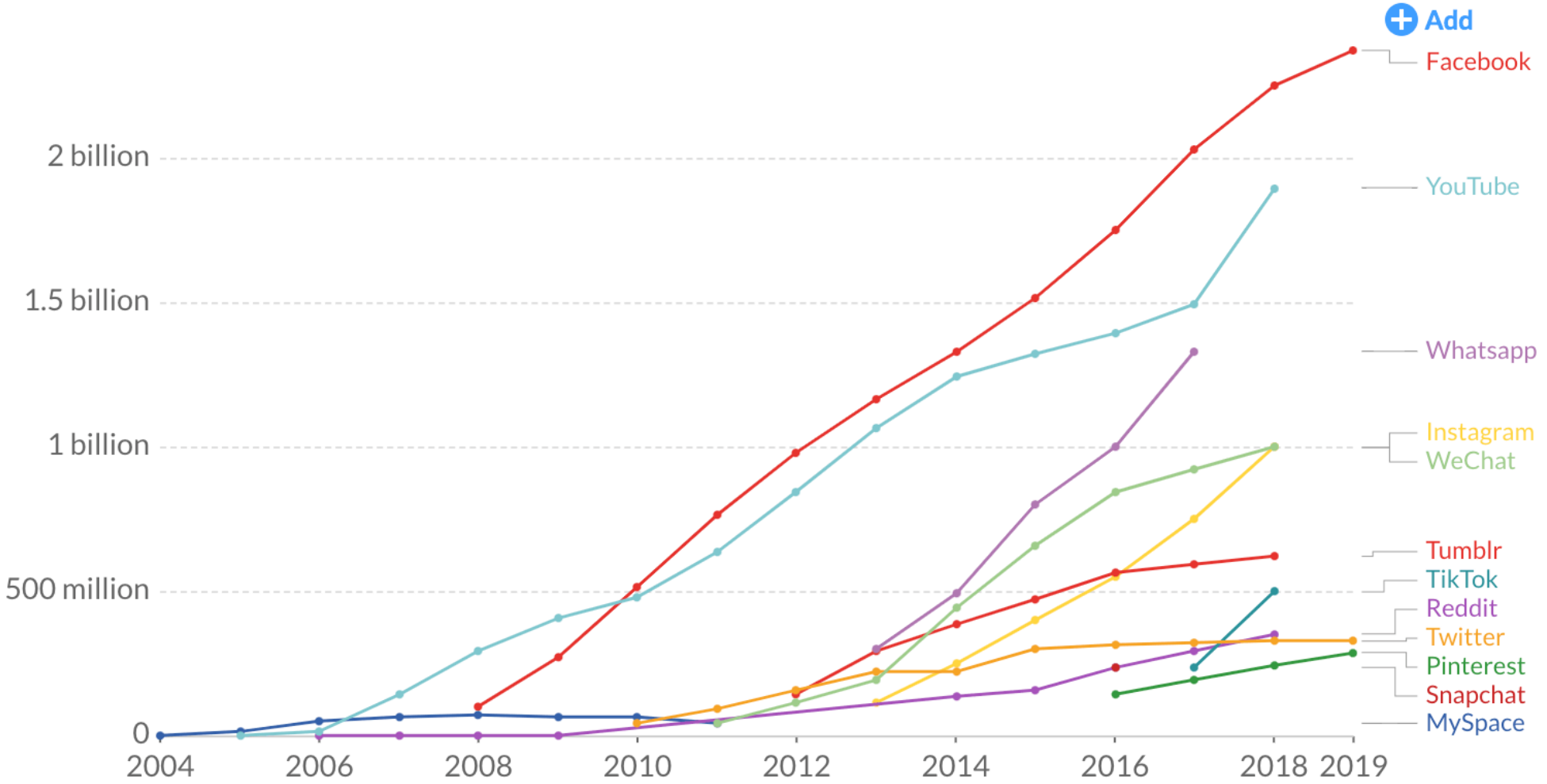


Facebook

Number of people using social media platforms

Estimates correspond to monthly active users (MAUs). Facebook, for example, measures MAUs as users that have logged in during the past 30 days. See source for more details.

Our World in Data

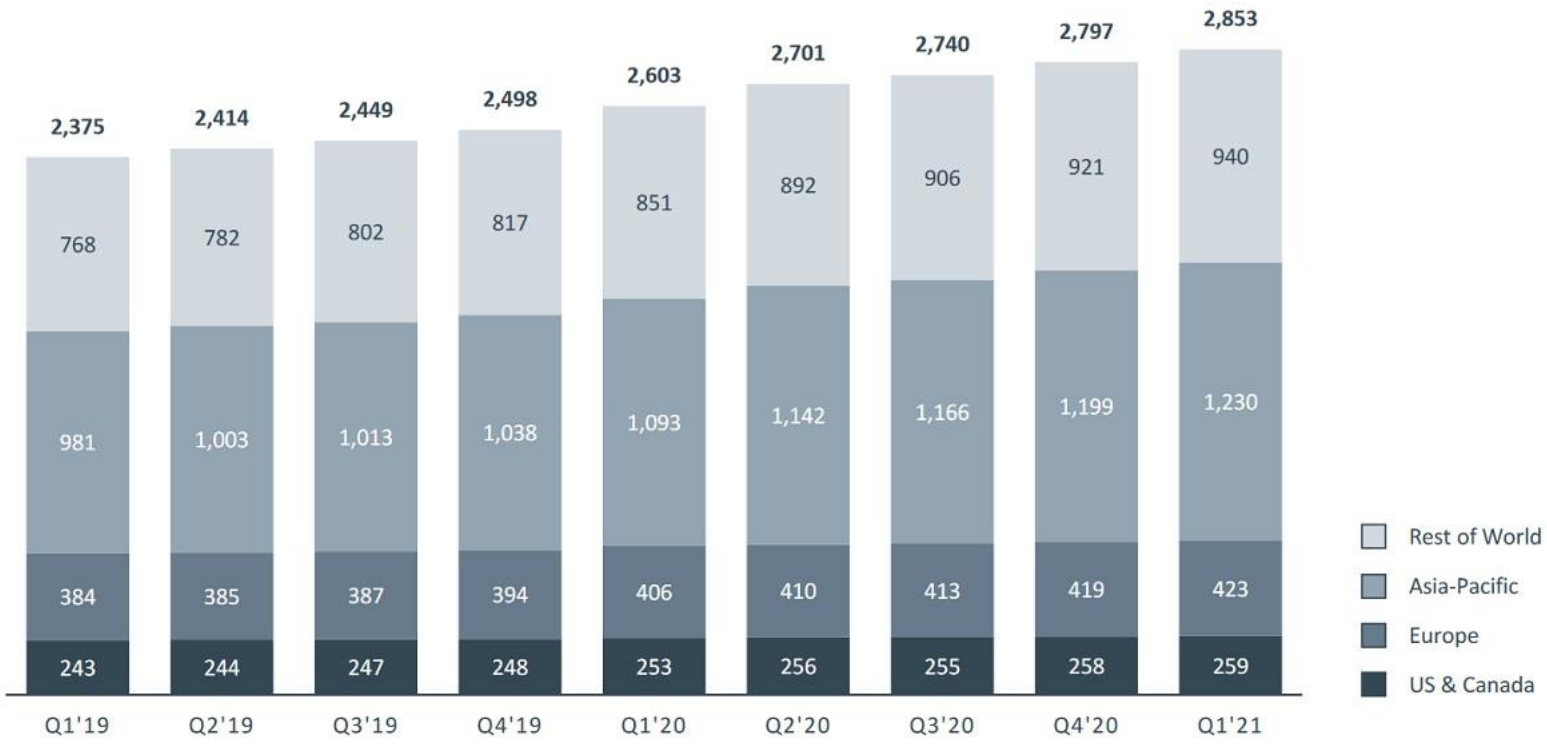


Source: Statista and TNW (2019)

Facebook Monthly Active Users (MAUs)

FACEBOOK

In Millions



Please see Facebook's most recent quarterly or annual report filed with the SEC for definitions of user activity used to determine the number of our Facebook DAUs and MAUs. The numbers for DAUs and MAUs do not include users on Instagram, WhatsApp, or our other products unless they would otherwise qualify as DAUs or MAUs, respectively, based on their other activities on Facebook.

- more than 2300 million Monthly Active Users (MAUs) worldwide
- countries with more users: United States, India and Brazil

The Graph API

Is the interface that allows any computer program to query Facebook data

- Nodes represent “things”: users, photos, pages, comments, *etc*
- Links represent relationships between “things”
- Fields represent “things” attributes: birthday, name, *etc*

Graph API Documentation

- Facebook for developers
  https://developers.facebook.com/
- Graph API Explorer
  https://developers.facebook.com/tools/explorer/

Graph API examples

- curl -i -X GET \
  "https://graph.facebook.com/your-facebook-user-id?fields=name&access_token=your-access-token"

- curl -i -X GET \
  "https://graph.facebook.com/your-facebook-user-id/photos?access_token=your-access-token"

- curl -i -X POST \
  "https://graph.facebook.com/your-facebook-user-id?email=you@your-email.com&access_token=your-access-token"

Facebook Graph API Fields

- id
- about
- address
- age_range
- bio
- birthday
- currency
- devices
- education
- first_name
- gender
- hometown
- inspirational_people
- interested_in
- languages
- location
- meeting_for
- middle_name
- name
- political
- relationship_status
- religion
- significant_other
- sports
- quotes
- timezone
- work
- website
- cover
- . . .

Facebook Graph API Links

- accounts
- activities
- albums
- books
- events
- friendlists
- games
- groups
- likes
- movies
- music
- statuses
- television
- videos
- family
- friends
- mutualfriends
- subscribers
- subscribedto
- …

Twitter

Two ways to use Twitter API

- REST API – provides access to read and write data on Twitter
- Streaming API – provides access to read large – scale and real-time Twitter data globally

Twitter API documentation

https://developer.twitter.com/en/docs

Twitter API example

https://api.twitter.com/2/users/84092683?user.fields=location"
-H "Authorization: Bearer $ACCESS_TOKEN

Twitter API – REST interface

GET https://api.twitter.com/1.1/users/show.json?screen_name=[user_name]

1. created_at
2. default_profile_image
3. description
4. expanded_url
5. favourites_count
6. followers_count
7. friends_count
8. id
9. lang
10. location
11. name
12. profile_background_image
13. screen_name
14. statuses_count
15. time_zone
16. verified
17. geo
18. in_reply_to
19. in_reply_to_status_id
20. status

Twitter Online Tools

Description	URL
Twitter advanced search	https://twitter.com/search-advanced
Twitter location search (enter GPS)	https://twitter.com/search?q=geocode%3A40.6306263%2C-8.6588713%2C1km&src=typed_query
Twitter time search (keyword + date)	https://twitter.com/search?q=Universidade%20de%20Aveiro%20since%3A2021-12-01%20until%3A2021-12-12&src=typed_query
Twitter analytics	https://foller.me
OmniSci (mapped Tweets)	https://www.omnisci.com/demos/tweetmap
OMTM (mapped Tweets)	https://onemilliontweetmap.com
Twitter analytics	https://www.twitonomy.com
Identifies fake accounts	https://www.socialbakers.com/feature/fake-influencers-detection
Analysis	https://fedica.com/

Google

Functionalities

Description	Query or tool url
Calculator and converter	"how much is 20% of 135" "cos(3x)+sin(x),cos(7x)+sin(x)" "1 light year to au" "what is the volume of a cylinder with radius 4cm and height 8cm"

Advanced search	https://www.google.pt/advanced_search
Translator	https://translate.google.com
Image search	https://images.google.com
News	https://news.google.com
Google Drive	https://drive.google.com
Book search	https://books.google.com
Video search	https://www.google.com/videohp
Academia	https://scholar.google.com
Finance	https://www.google.com/finance/
Trends	https://trends.google.com
Ngrams	https://books.google.com/ngrams

Google Hacking – Possibilities

- more assertive research
- less false positives
- access to devices connected to the network configured by default
- access to contente not available through normal search
- *etc*

Database of examples: https://www.exploit-db.com/google-hacking-database

Google Hacking – Examples

- intext:”password" site:www.domain.com (example)
- "contrato" filetype:pdf (example)
- "text" filetype:doc | filetype:docx (example)
- allintext:"*.@gmail.com" OR "password" OR "username" filetype:xlsx (example)
- inurl:view/view.shtml (example)
- parent directory Index of mp3 (example)
  - mp4,mov,jpg,jpeg,…

Google Hacking – Examples

- ext:txt intext:@yahoo.com intext:password (example)
- password console-password ext:cfg –git (example)
- intitle:"index of " "*.passwords.txt“ (example)
- inurl:login.txt filetype:txt (example)
- Index of /backup (example)
- inurl:/intranet/login.php (example)

Google Hacking – Examples

- "parent directory " /appz/ -xxx -html -htm -php -shtml -opendivx -md5 -md5sums
- "parent directory " DVDRip -xxx -html -htm -php -shtml -opendivx -md5 -md5sums
- "parent directory " Xvid -xxx -html -htm -php -shtml -opendivx -md5 -md5sums
- "parent directory " Gamez -xxx -html -htm -php -shtml -opendivx -md5 -md5sums
- "parent directory " MP3 -xxx -html -htm -php -shtml -opendivx -md5 -md5sums
- "parent directory " Name of Singer or album -xxx -html -htm -php -shtml -opendivx -md5 -md5sums

Other Open Sources

Other open sources - Georeferencing

Google Maps

Google Earth

Bing Maps

- Allow searches by address
- Work by layers
- Preferred XML as information exchange format
- Allow detailed view
- Area measurement tools
- Allow interactive presentations

Foursquare

- allows to search public places and goers
- allows to search people and places visited
- https://pt.foursquare.com/city-guide

YouTube

- Allows you to search for videos by themes
- Has a system of recommendations and alerts
- Allows you to search for users/commenters
- Allows you to view the entire comment history
- https://www.youtube.com

Utilities(1)

- "Whois" online

  https://whois.domaintools.com/

- Way Back Machine – Internet Archive

  https://web.archive.org/

- Google Cache

  https://cachedview.com/

- Email Header Analyzers

  https://toolbox.googleapps.com/apps/messageheader

  https://www.whatismyip.com/email-header-analyzer

  https://www.gaijin.at/en/tools/e-mail-header-analyzer

Utilities(2)

- Profile pictures

  https://www.picuki.com
  https://gramhir.com

- Reverse image

  https://tineye.com
  https://www.google.com/imghp?hl=en

- Fake photos

  https://fotoforensics.com

DarkNet

Visible Web

versus

Deep Web

Visible Web

- also called the Surface Web, Indexed Web, Indexable Web or Lightnet
- portion of the World Wide Web that is available to the general public
- search able with standard web search engines

Deep Web

- has contentes that can not be found or directly accessed via search engines
- sites that are purpose fully designed to keep search crawlers out
- however, doesn’t requires special browsers
- instead a direct link to access is needed

Dark Net

versus

Dark Web

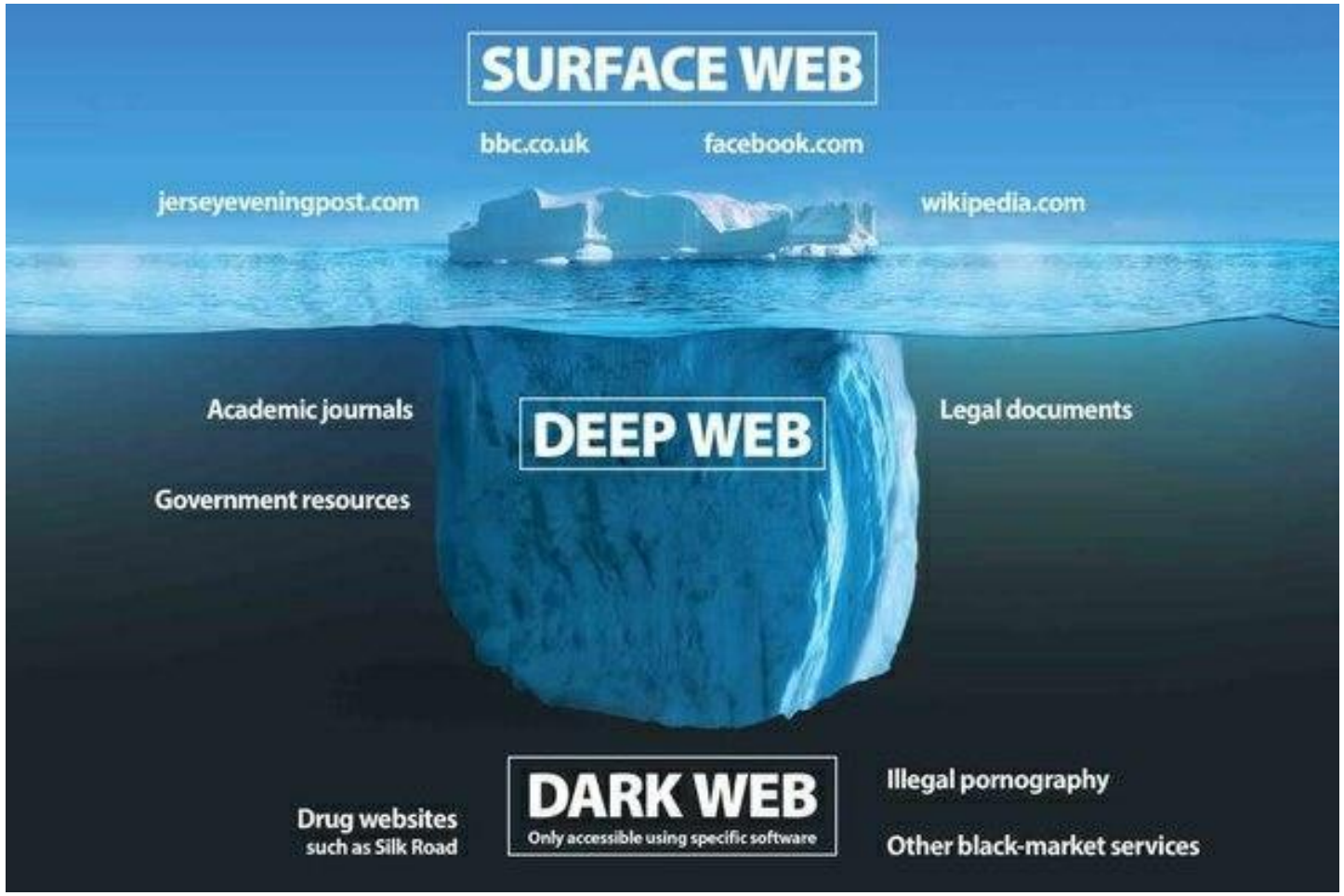
Dark Net

- is na overlay network (a network built on top of the Internet)
- was designed specifically for anonymity
- *e.g.* Tor, I2P, Freenet, DN42 *etc*

Dark Web

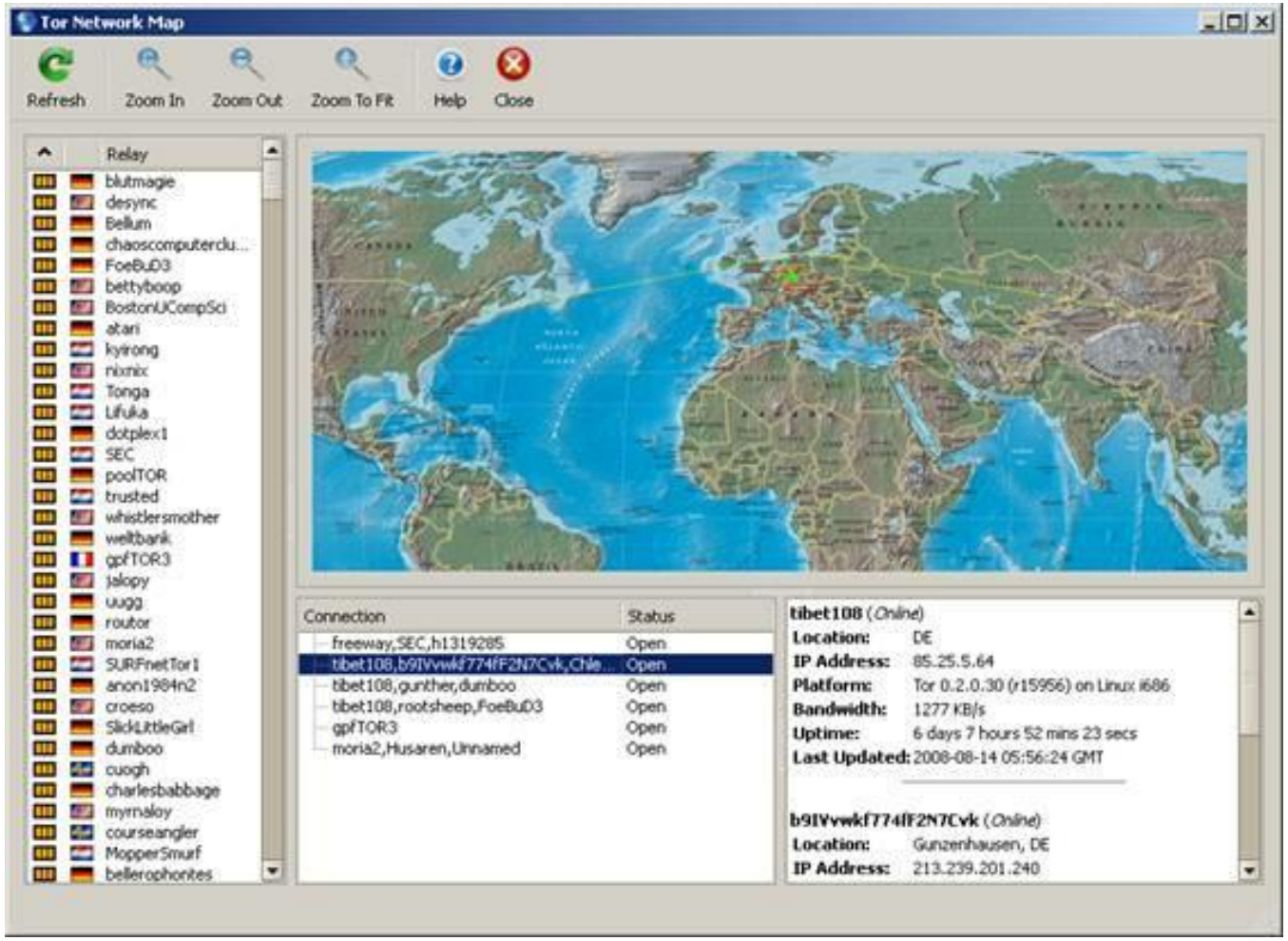
- Refers to websites on a Dark Net
- Market Places, Dark Net Markets, *etc*

Perspective

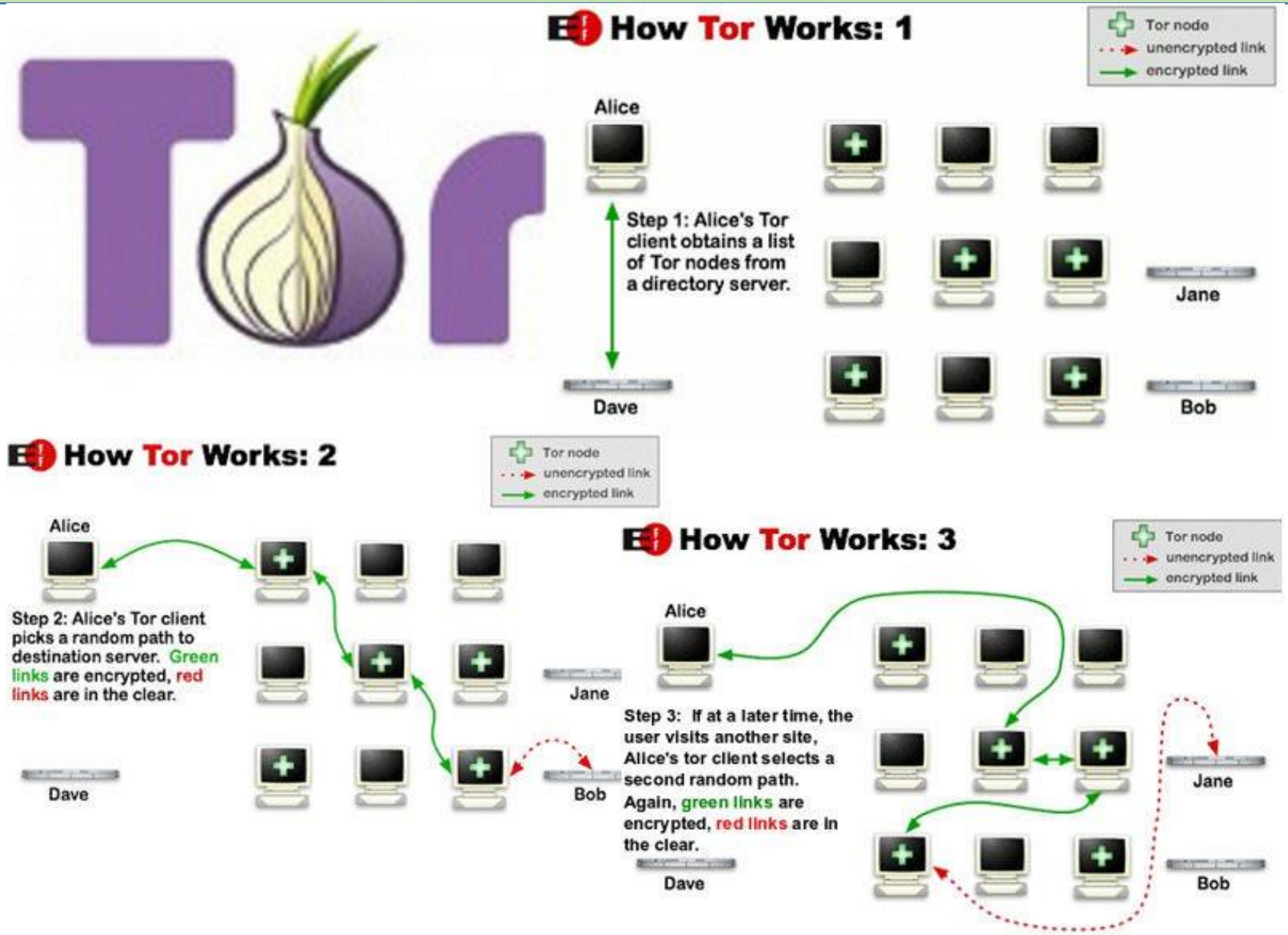


The Onion Router (TOR)

https://www.torproject.org

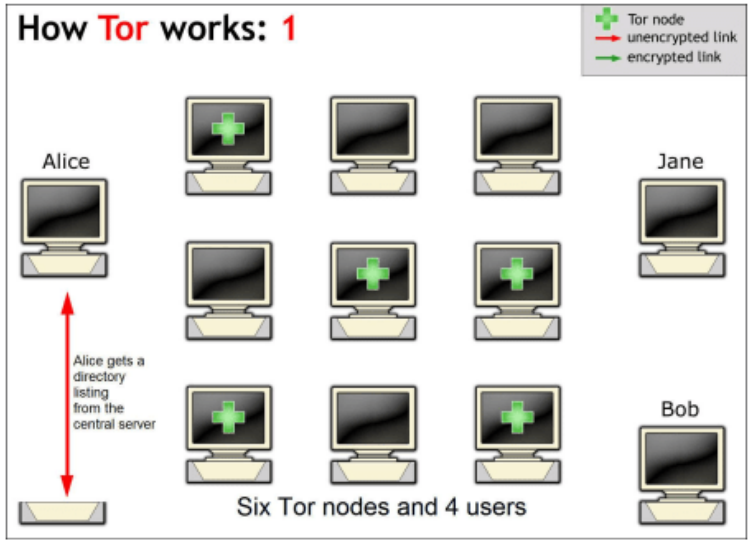


TOR Connections

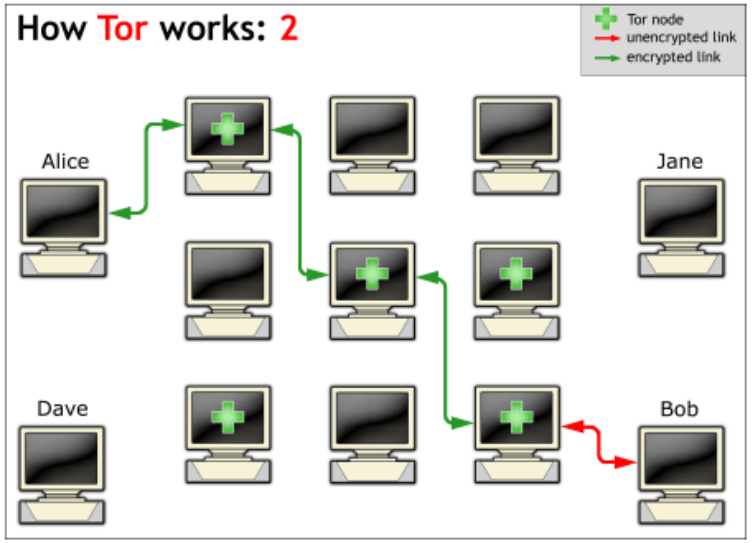


TOR Connections

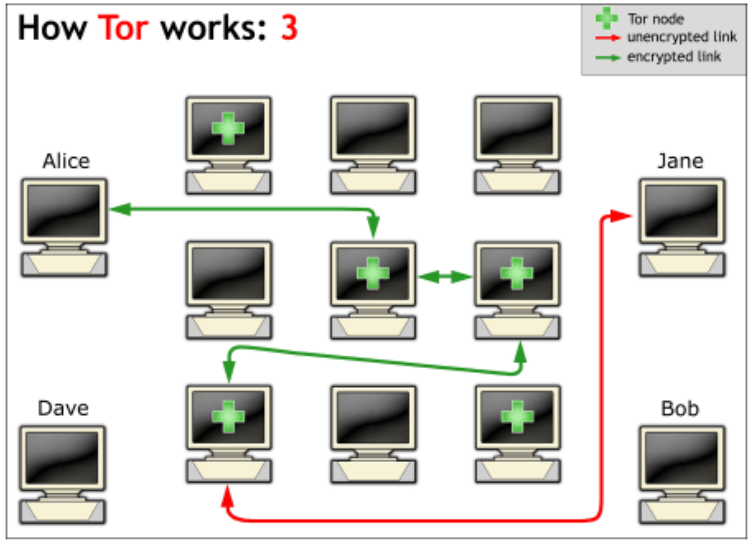
Connection set up



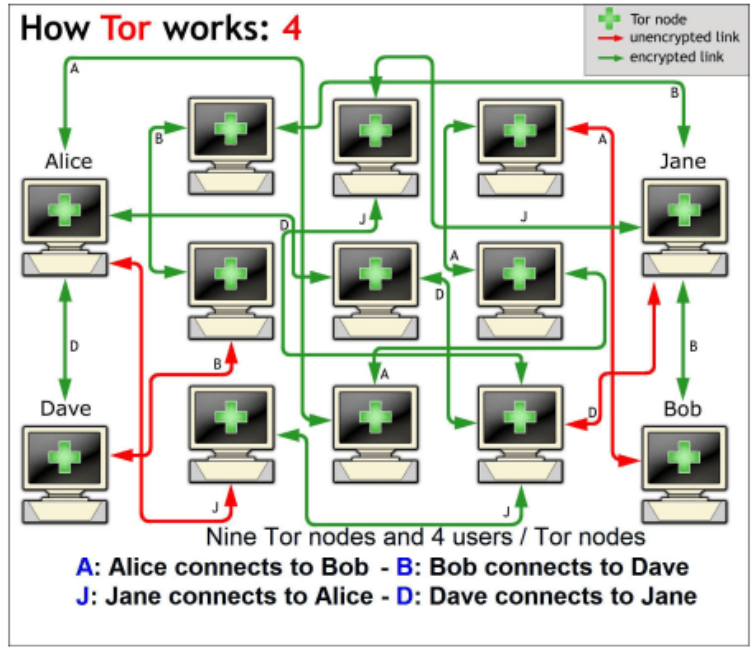
Connection



Connection Timeout - entry node change



A real scenario - multi purpose node



Other open sources – Dark Web

Useful Links:

The Hidden Wiki
http://zqktlwiuavvvqqt4ybvgvi7tyo4hjl5xgfuvpdf6otjiycgwqbym2qad.onion/wiki/index.php/Main_Page

OnionLinks
http://jaz45aabn5vkemy4jkg4mi4syheisqn2wn2n4fsuitpccdackjwxplad.onion

TorLinks
http://torlinksge6enmcyyuxjpjkoouw4oorgdgeo7ftnq3zodj7g2zxi3kyd.onion

Ahmia
http://juhanurmihxlp77nkq76byazcldy2hlmovfu2epvl5ankdibsot4csyd.onion

References

References

Websites

- https://www.inteltechniques.com/links.html
- http://rr.reuser.biz/
- https://www.exploit-db.com/google-hacking-database

References

Books

- Michael Bazzell,“Open Source Intelligence Techniques: Resources for Searching and Analyzing Online Information”,6th edition, 2018 ISBN-13978-1984201577

- NATO,“NATO Open Source Intelligence Handbook”, 2001

- The Open Source Manual for the Polícia Judiciária, by António Jorge Filipe Fonseca (Phd) 2015

- Mário Antunes, Baltazar Rodrigues, “Introdução à Cibersegurança: A Internet, os Aspetos Legais e a Análise Digital Forense”, 1st edition, 2018 ISBN-13978-9727228614

QUESTIONS?