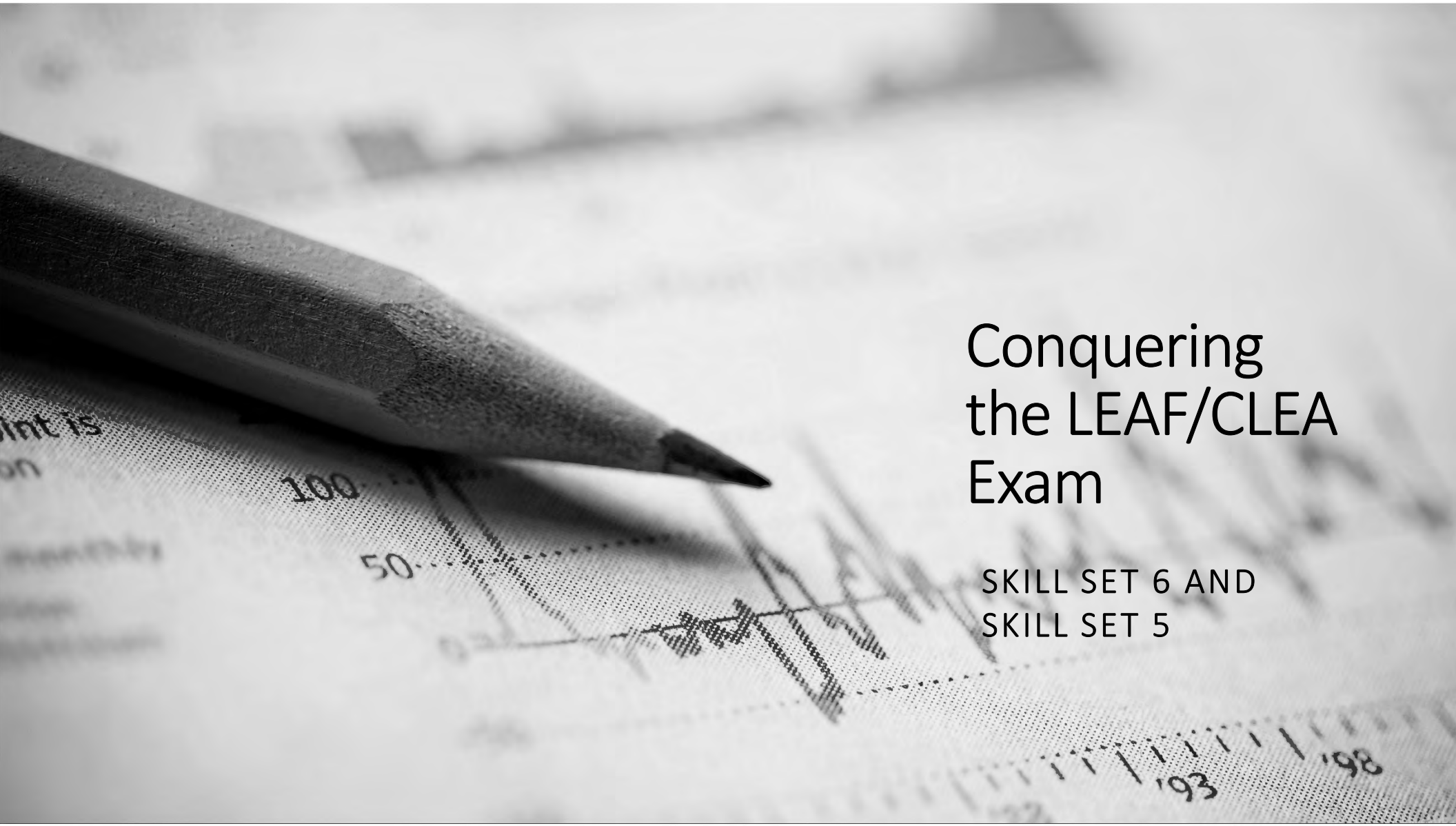


Conquering the LEAF/CLEA Exam

SKILL SET 6 AND SKILL SET 5

About the Instructor/Course

- Instructor – Jenny Zawitz Jennifer.Zawitz@gmail.com
- CLEA Study Guide: https://iaca.net/wp-content/uploads/2021/06/CLEA-Skill-Sets_Study-Resources-051821.pdf
- LEAF Study Guide: https://iaca.net/wp-content/uploads/2021/06/en_LEAF-Core-Competencies_Study-Resources.pdf
- Exploring Crime Analysis: Readings on Essential Skills (3rd Edition) - IACA
- Each month will cover a different section of the study guide
- Intended as a supplement NOT a substitute for the texts and the Essential Skills classes
  - This course will help you focus your studying, but the courses and text will provide the actual understanding you need to pass the tests

Data Management

SKILL SET 6, CHAPTER 6

Data Management is

- Readily accessible, timely and complete, accurate data
- Analysts should be able to perform flexible queries and extract data for analysis in a variety of tools.
- Data table allows for this – quick scanning and summary of elements relevant to a question

Detective	Total Cases	No. Closed	No. Open
Peralta	42	38	4
Diaz	38	37	1
Santiago	50	48	2
Boyle	36	31	5
Scully	7	0	7
Hitchcock	6	0	6

VS

Peralta, Diaz, Santiago, and Boyle have had busy quarters with 42, 38, 50, and 36 cases assigned respectively.
Peralta has closed 38.
Diaz has closed 37.
Santiago has closed 48 and should not, as Peralta suggests, be renamed Detective Terrible Detective.
Boyle has closed 31 cases.
Scully and Hitchcock have not closed any cases between them with 7 and 6 cases open respectively.

Structured vs. Unstructured Data

- Refers to data in original form not information that has gone through the analysis process
- Structured data must be used for quantitative analysis
- Unstructured data is not easily aggregated or summarized by category – qualitative analysis.
  - Unstructured data more common – crime reports, CAD calls
- To structure unstructured data, we need to code it or assign variables to groups. Do this by creating a series of "fields" or categories for the data.
  - Structure/Unstructured data is a scale. Not binary.
  - Data gets more structured with more discrete fields. Higher specificity in fields, more structured.
  - Be as discrete as possible up to the point of loss of utility.
- Note that when data is structured, the collector is interpreting the data. May cause error.
  - Ex: time of larceny from vehicle may not be known but extrapolated from other surrounding larcenies.
- Be sure to preserve the data in unstructured form for official use (police reports)

Relational Databases

- In databases, the table is the major unit of organization. Tables predate computers.
- Weakness of stand-alone tables is that they organize in such a way as to only look at the topic in one way.
  - Ex: can create a table that organizes crime by criminal name or criminal name by crime but not both.
- Solution: relational database! (basically all of your RMS)
  - In these, tables store "topics" of data but the tables can be linked together to search complex questions
- Relational databases store each topic of data in a separate table with tables linked through key fields (case numbers, crime type, offender, etc.). Each field has a distinct name and data normalized to make sure no item of data is entered more than once (perfect world)
- Each table has a unique key or primary key that never duplicates between records (case number, SSN for their respective tables).

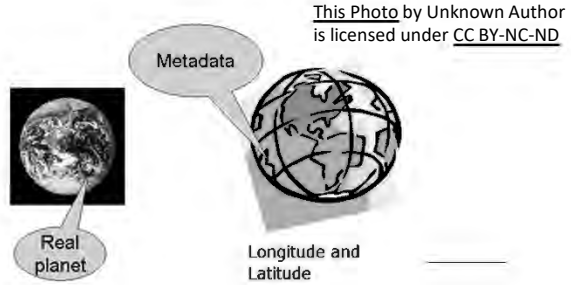
Relational Databases (cont.)

- Two basic types of relationships in database: one-to-one and one-to-many
- One-to-one relationship: each record in one table only has one related record in another table.
  - Uncommon – if data were related in this way, you could just use a single table.
  - Can be useful if certain fields might rarely be entered for most records. Ex: burglary m.o. table. Would only need this table to be connected to burglaries and not other crime, so burglary m.o. table can have a one-to-one relationship with burglaries.
- One-to-many relationship: entry in one table can be related to many records in an associated table. May be several relationships.
  - Ex: calls for service related to units dispatched, people related to addresses
- Relational databases managed by a Database Management System (DBMS)
  - Ex: Microsoft SQL Server, Oracle, MySQL, Microsoft Access

Police Records Management Systems

- CAD: Computer-Aided Dispatch
- RMS: Records Management System
- Each vendor does these differently, but generally RMS use an "incident" table to store dates, times, locations of incidents with related offenses, stolen property, and vehicles in associated tables.
- When using these systems, you may notice a number of different modules that group common tables together (Calls for Service module, Incident module, Person module, Property module, Evidence module, Arrest module, etc.).
- Library table: stores values used in lookups or drop-down menus. May use when connecting tables via Access to look up table names. May be hidden in architecture of database. Make sure you have access to this to know what tables to join for your queries.
- Systems should be evaluated with the analysts in mind – seek systems with open architecture and available data dictionary to facilitate accessibility and quality.

Metadata



- Data about data
- Data that describes aspects of the dataset, answering questions to properly analyze and interpret data.
- Ex: spatial data might come with metadata that describes its coordinate system and geocoding percentage.
- Often unstructured or semi-structured. Can be described in textual paragraphs rather than discrete fields.
- Metadata can also be found in emails or separate text files rather than attached to the primary file.
- Analysts should include a few paragraphs of metadata to assist with the interpretation of the data when it is shared with other analysts, external researchers, and the public.
- Describes the fields that were used and the methodology by which you pulled the data.

Data Sources

- Primary data source: collected directly from the source by the researcher or analyst
  - Ex: surveys, interviews, field notes from environmental observations
- Secondary data source: collected by someone else and made available to the analyst (most frequently used)
  - Ex: data in RMS, crime scene reports, calls for service data, news articles, arrest reports, traffic citations, crime bulletins from other agencies, probation data, motor vehicle registration, social media data, pawn data.
- Analyst coding data: gray area in between primary and secondary. May not be the ones to interview the victim or offender but may code the data that results from these primary sources in a way RMS might not anticipate
- Access level – does the analyst have access to the raw data or reported data. Raw data means you have access to the entirety of a set and can search it vs. reported where you can search one piece of data individually at a time.

Crime Analysis Data Management

- Generally – administrative and tactical data depends on data reported through the police records management system whereas crime intelligence and strategic analysis depend on data from external sources.

- Analysts should be able to do the following in a enterprise-level data management system:

1) Link to and Import data from a variety of sources
2) Join tables from disparate sources in common queries
3) Modify and clean records
4) Supplement data from other sources with analyst-collected or analyst-coded data
5) Export and link to any data analysis tools that the analyst may wish to use

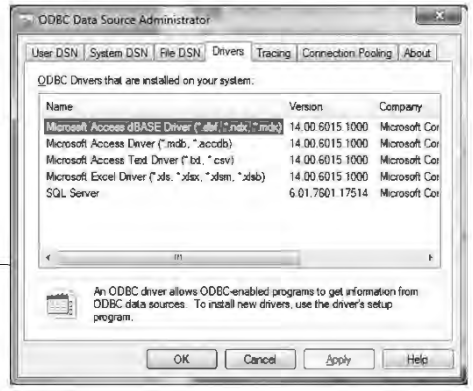
Common Mistakes in Data Management

- Refusal to enter data: need to take responsibility for your data. Some analysts don't want to do data entry and cleaning.
- Entering data in non-databases: waste of time to enter data into Word, PowerPoint where data is unsearchable and irretrievable.
- Entering data that can be calculated: if already have DOB, then don't need to store age as a separate field, if you enter a robbery, you don't need to have an entry for violent crime.
- Accepting "no" when it comes to raw data access or common analytical technologies: must have direct ODBC access to their CAD and RMS

Desktop Databases vs. Servers

	Desktop DBMS	**Server DBMS**
Examples	Access, FileMaker, Open Office's Base	SQL Server, Oracle, IBMs DB2
Strengths	• Low cost • Application-development tools and integrated reporting tools • Easier to learn • Large existing user base, training, technical assistance	• Faster • Larger storage • Less prone to crashes and corruption • More security features
Weaknesses	• Limited security options • Size limitations • Slower • Poor performance over a network	• Sharp learning curve, need significant support • Need separate programs to develop interfaces and reports

ODBC



- Open Database Connectivity (ODBC) crucial for querying
- Databases only as useful as the ability to ask questions of them.
- Microsoft introduced ODBC in the early 90s with Windows 95. Free technology.
- Act as a medium between DBMS and the tools you want to use to ask questions about the data. Allows applications to connect to databases.
- In policing, use ODBC to connect your RMS to Access, Excel, ArcGIS, Crystal Reports, and SPSS for querying, mapping, reporting, and other tasks.
- Connection typically facilitated by the system vendor – can be difficult because vendors want you to use the reports they set up.
- Set up in each computer's ODBC control panel and can either apply to anyone who uses the computer (system DSN – data source name) or a single user (user DSN).
- Once set up, can connect DMS to any application – Access, Crystal Reports, SPSS, etc.

ODBC (cont.)

With ODBC connection analysts can:

1) Link to the data and use querying and reporting tools from other applications. Original data is left alone but ODBC connections applies data querying, reporting, visualization, and mapping.

2) Link to data to supplement it with data in analyst-managed tables. Can link the original data to data that the analyst collects and codes in local tables. Can use this to replicate fields in the original data that have poor accuracy or completeness.

3) Copy the data out of the original system. Original data may be so poor that the analyst can't use it in raw form. Export and store in local tables where it can be cleaned and supplemented. Helpful for slow connections as well.
    1) Referred to as a "shadow RMS" – a copy of the original system, generally only accessible to the analyst which over time becomes more complete and accurate than the original system. May lead to questions about which statistics are the official agency position.

Data Quality – Types of Bad Data

Bad Data Type	Examples	Consequences
Missing Reports	Officer does not complete full report, call remains in CAD only	Missed patterns and series, less understanding of trends/hot spots, can't link individual to offense
Fields Not Filled In	Time from/time to not filled in, height/weight skipped	Inability to accurately analyze crimes for appropriate characteristics, can't find key records
Mis-coded fields	Crime type entered wrong	Incidents do not come up when conducting a search and are counted incorrectly in reports
Misspellings	Misspell street names	Inability to geocode/map data, can't find key records
Duplications	Same incident/person entered more than once	Double counting in reports, confusion as to which record is correct
Mismatch in fields	Street number too high for street, agg. assault that has no weapons/injury	Inability to analyst crime for important factors

Data Cleaning/Supplementation

- Two methods for cleaning: One Records, Many Fields or One Field, Many Records
  - One Record, Many Fields: look at one complete record at a time and scan for problem in all fields correcting as needed. Time consuming, part of a regular analyst routine.
  - One Field, Many Records: searching for a common problem in a single data field and cleaning all at once. Queries that compare street names to master list, DOBs in the future, duplicates of master name index.
- If issues occur in predictable ways, can automate the process.
- Data cleaning works for existing data, but is incomplete/inaccurate because sometimes data is left out (blank fields, fields outside of the RMS preselected fields).
- Need ability to supplement with linked or imported data with additional data that the analyst cares about – categorize individuals with mental illness, homelessness, substance abuse, sex offenders; include if crime was gang or gun involved.
- If supplement data, need to minimize duplication of data entry
- Expressions or calculated fields allow analysts to make new data out of existing data. Can use to determine offender's age based on dates either at the time of the crime or to date.

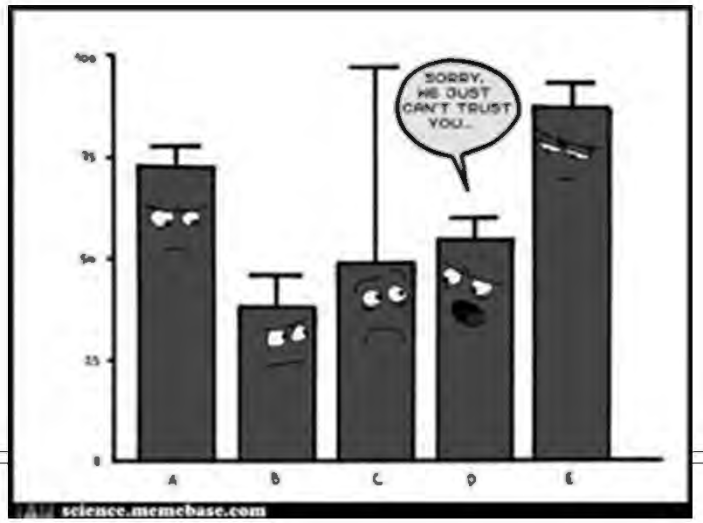
Structured Querying Language (SQL)

- Most common language when working with data, developed by IBM in the 1970s
- Uses a collection of plain-English clauses, conditions, expressions, and operators to query, insert, update, and delete data in a relational database.
- Access is a visual tool for writing SQL behind the scenes. Can switch to SQL view.

Types of Queries	**What They Do**	**Examples**
Queries that filter	SELECT, FROM, WHERE commands to show records that match criteria	List of traffic crashes that involve pedestrians from dates at location
Queries that aggregate	GROUP BY, COUNT, SUM, AVG that quickly summarize datasets. Useful in strategic analysis.	Count of types of weapons used in a year's armed robberies
Queries that cross tabulate	Compare aggregations across two or more factors. Good for statistical reports.	Count of how many offenses occurred in each year.

Data Visualization

- Graphical display of data to make it easier for an audience to make sense of it and find important patterns - Charting, mapping, graphs
- Dynamic visualization of data – mapping that includes dashboard software. Can create a variety of maps and charts that can be manipulated or changed in real time to show different things.
- Most useful if they draw from the same dataset as the rest of the analysis products
- Not analysis by itself – just a step on the way to analysis

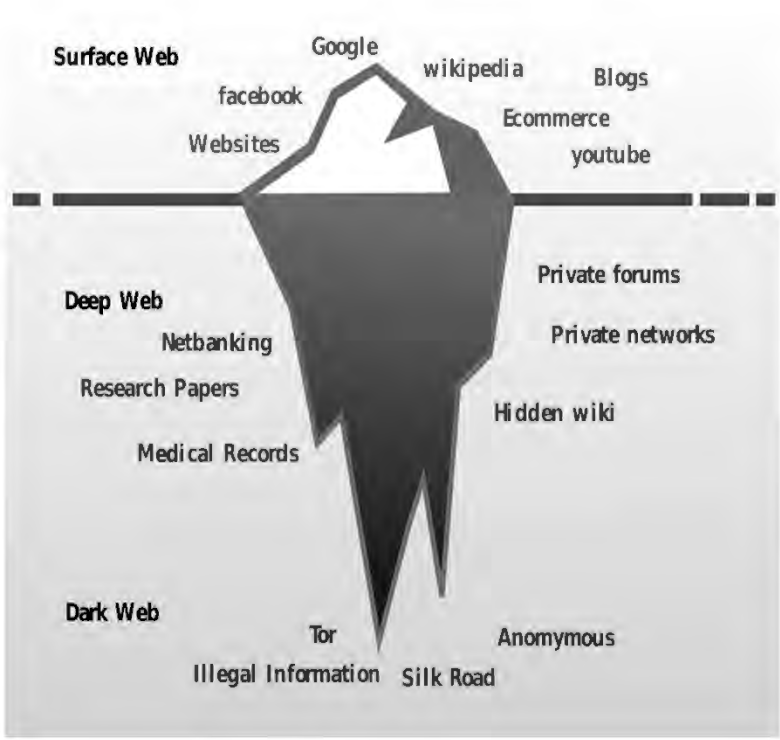


Internet Resources

SKILL SET 5/CHAPTER 5

Internet/World Wide Web

- Not exactly the same – WWW allows you to access many parts of the Internet and visualize information contained within
- Access internet using a browser
- Search engines are used to crawl the web, index sites, and make them discoverable
- Surface Web: parts of the web that can easily be found using search engines
- Deep Web: parts of the web that can be easily accessed but not indexed by search engines. Includes data sets and databases.
- Dark Web: small part of the deep web that is deliberately hidden from view and contains many things, including sites for criminal activities. The Onion Router (TOR) lets you browse anonymously but need to be careful.



History of the World Wide Web

- Evolved from Arpanet – developed in 1969 with funding from DoD
- DoD created a new and separate network in 1983, referred to as Internet by 1984
- WWW developed by Tim Berners-Lee when he created Hypertext Markup Language (HTML) in 1990
- WWW available to general public in 1991
- Berners-Lee also invented Hypertext Transfer Protocol (HTTP) which defines how messages are formatted and transmitted over the internet and the URI (Uniform Resource Identifier)
- URL (Uniform Resource Locator) is the most common URI – uses text strings to determine a network location on the WWW
- HTTPS = secure connection to a site

Domain Names

- Domain name is the name of a website
- Points person to a specific website being hosted at a specific location
- Regulated by Internet Corporation for Assigned Names and Numbers (ICANN) – non-profit that coordinates internet naming system and determines what domain names are allowed to be used
- TLD – Top Level Domain – is the last part of the domain name (.com, .net)
- .com = commercial sites used by anyone
- .org = organizational sites originally reserved for causes and non-profits
- .gov, .mil, and .edu = government, military, education sites
- Can also have a TLD for two digit country codes - .us for United States, .ca for Canada
- Can use whois. websites to figure out who registered a domain, but can be registered anonymously for a fee.

IP

- Hidden behind domain names are the Internet Protocol (IP) addresses
- When you connect with www, you're using a connection with an IP address to connect to another location with an IP address. You can see your IP by Googling "show my IP"
- Locard's Principle – every contact leaves a trace. Every time you visit a website, you leave your IP address, among other things. There are tools that allow you to be anonymous.
- Intranet: internal private network that uses the same protocols as the Internet. Used within organizations to share information internally.
- Extranet: intranets that allow restricted access by external users such as other agencies, vendors, contractors, and customers.
- VPN: Virtual Private Networks, restricted to specific users but extend outside of an organization's private network structure using public networks. Often encrypted and secure.

Search Engines and Useful Sites

- Search engines are generally designed to bring back the most popular page for the widest range of users that search for them.
- Industry built on Search Engine Optimization (SEO) that ranks sites based on a number of positive and negative factors
- Google has an advanced search page (www.google.com/advanced_search) that helps users search in a more specific way – includes things like time range which is when the site was first indexed by Google or location based searches
- Wayback Machine allows you to search for past versions of a website (https://archive.org/web/web.php)
- Google Dorking (search term) provides tutorials on how to run advanced searches on Google including Deep Web searches
- Google Alerts can be used to set up a search so that you get an email alert every time a website with your search term is newly indexed on Google

Useful Sites (cont.)

- Also using continuous information – RSS feeds (Rich Site Summary aka Really Simple Syndication) is a way for publishers to share updated content across the Web in a standardized format. RSS feeds let the user get aggregated updates on a search term in a single location.
  - Good for monitoring news sites and blogs without signing up for accounts
- SearchDiggity or Google Hacking Database are good tools for advanced searches
- Shodan (https://www.shodan.io) allows you to search for any internet connected devices (webcams, routers, security cameras, traffic lights, industrial control systems, etc.)
- Snopes is very good for identifying false stories and debunking them with citations
- Law Enforcement Online (LEO) (includes direct access to N-Dex, NGIC, and Internet Crimes Complaints Center) and Homeland Security Information Network (HSIN)
- Check your book for additional useful websites – public records, social media, government agencies, free training (YouTube, FEMA, DHS, FLETC), and free tools

Boolean Searches

Operator	Purpose	Example
" " (Quotes)	Search term for exact match	"Crime Analysis"
- (Dash or minus sign)	Before a word excludes that word	Crime analysis –CSI
OR (capitalized)	Put between words that you want different search words in conjunction with the same term	IACA Conference 2021 OR 2022
site:	Search within a specific site or domain	Crime analysis site:iaca.net
link:	Identify sites that link to your target website	link:iaca.net returns some sites that contain a link to IACA (not all)
related:	Sites that are similar to the site	related:iaca.net gives sites related to IACA
* (asterisk)	Wildcard placeholder (w/ or w/o quotes)	International * of Crime Analysts
cache:	What the site looked like the last time Google looked as the site (good if info recently removed)	cache:www.iaca.net
info:	Provides info on a specific website	info:iaca.net shows link and several options

Open Source Intelligence

- Information collected from publicly available sources and analyzed and distilled into something that supports a specific requirement
- Gathered to create OSINT (Open Source Intelligence) – know agency policies and procedures as well as local, state, and federal laws before collecting this information.
- Any LE intelligence database that is federally funded must comply with 28 Code of Federal Regulations Part 23 or 28 CFR 23 – measures that protect the privacy of individuals who might not be involved in criminal activities. Have systems in place to not retain information on people who have moved away from criminal involvement (juveniles)
- If you find yourself blocked from certain websites, work with your IT department to get access. Ideally the analyst should have administrative privileges on their computers so they can install necessary programs. Can also have computers that are complete separate from the network that are available to analysts and investigators that need this.

Social Networking

- Social media is any type of media used for social interactions
- Web 2.0 refers to websites where users create content on that site.
- CompuServe had a CB simulator in 1979 where users with a home computer and a modem could talk to each other
- www.makeoutclub.com – 1999 users could create profiles and upload photos
- Friendster 2002, MySpace 2003, Facebook 2004, YouTube 2005, Twitter 2006, etc.
- Over 85% of American adults use the internet, 95% with teens and 67% using social networking sites
- Know which sites are gaining in popularity and which ones are waning.
- Real time situational awareness, gather intelligence, undercover operations, engage with community, deliver emergency notifications, etc.

Gathering and Sharing Information

- Can gather government records on intranet or VPN connections
- Public records can include land ownership, court records, tax assessment, voter registration, business licenses, and criminal records (some states).
- Bookmark useful sites and check frequently to make sure they still work
- Know Freedom of Information Act (FOIA) or sunshine laws to help determine what you can share outside of your agency, though sharing with the public will likely fall under LE policy
- For criminal investigations, find a suitable law enforcement contact for a social media site (may fall to the investigator) so investigator can send a Preservation Letter to the company so that evidence is preserved immediately. www.search.org/resources has contact info for many sites.
- Digital photographs may contain metadata in the form of EXIF (Exchange Image File) data, which could provide geolocation/time/device information, though most sites scrub this now.
- When setting up fake accounts, never use another person's photo without permission.

Conclusions

- Read the books and take the classes to strengthen understanding.
- Try to apply the things learned to your every day work to “make them stick”.
- Use the study guides.
  - https://iaca.net/about-clea/ (links for program outline and study guides here)
  - https://iaca.net/about-leaf/ (links for program outline and study guides here)
- Next month: Applied Research Methods (Skill Set 7) and Qualitative Analysis (Skill Set 10)

Any questions?