OPEN SOURCE INTELLIGENCE (OSINT)

DATA ANALYSIS VERSUS INTELLIGENCE ANALYSIS: WHAT'S NEW IN THE GAME?

Andrada Nicoleta HALGAŞ - BLAGA[*]

Abstract

Nowadays, when information is abundant, ever-present and readily available, how does an intelligence organization adapt, in order to collect and process what is truly relevant, as well as provide "value-added" to decision-makers?

The information revolution has greatly impacted national intelligence. Collection has become increasingly focused on large sets of data (big data). As more of that processed data is available outside intelligence organizations, the role of intelligence professionals has shifted from analysing collected data to provide assessments of current events, to that of generating knowledge, predictions and warning that help decision-makers avoid strategic surprises.

Keywords*: data analysis, intelligence analysis, big data, analysis-driven collection, data analysis tools, tech skills, research and development, artificial intelligence*

Data and Intelligence analysis - Elements of difference and continuity

There is little consensus so far regarding the definition of data analysis. This is probably due to the fact that data analysis is traditionally closer to business intelligence, which means it hasn't until recently been the focus of IC research efforts in order to properly delineate it from similar concepts such as *data science* or *machine learning*.

In order to better understand our topic of discussion, let's take a look at the following definitions:

Data: 1. Facts and statistics collected together for reference or analysis (Oxford Dictionaries).

1.1 The quantities, characters or symbols on which operations are performed by a computer, which may be stored or transmitted in the form of

[*] Analyst, intelligence research and development specialist.

electric signals and recorded on magnetic, optical, or mechanical recording media (Oxford Dictionaries).

Intelligence: As in the case of data or data analysis, there have been several attempts to define intelligence, in the context which is of interest here. Seven years later, Michael Warner's words on the issue still hold and a definition of intelligence remains on the "Wanted" list of intelligence professionals worldwide (Warner, 2007).

Several of the most meaningful definitions, for the points further made in this paper, are as follows:

1. [Intelligence is] knowledge or foreknowledge of the world around us - the prelude to decision and action by US policymakers (Central Intelligence Agency, 1999).

2. On the one hand, it [intelligence] refers to organization collecting information and on the other to the information that has been gathered (Lauquer, 1985).

These two definitions lead to the conclusion that the common understanding among intelligence professionals is that intelligence is both a process and the product resulting from it, and that its purpose is to assist decision-makers in adopting policies.

Analysis:

1. Breaking down a problem into its component parts, assessing each part separately, then putting the parts back together to make a decision (Heuer).

2. The conversion of processed information into intelligence through the integration, evaluation, analysis, and interpretation of all source data (US Department of Defence).

If we take into consideration only the first acceptation of data provided by the Oxford Dictionary, then we have nothing new in the game of data analysis, compared to intelligence analysis: they both mean collecting and processing pieces, providing them with context and meaning and delivering the full puzzle to decision-makers.

However, if we look at the second definition of data, we may observe that there is a third party involved: the **computer**. In this case, data analysis implies a different sort of process, technology being a key element of it. Data is defined as "quantities, characters or symbols", which hints at the fact that before processing them by using specialized software, they might not be intelligible to the intelligence analyst. If in the first case, "facts" and "statistics" can be analysed even without the help of technology, for "quantities, characters or symbols", software processing is a prerequisite.

Therefore, we may conclude that data analysis is, in fact, based on the same principles as classical intelligence analysis, the main difference being that it has a greater quantitative dimension and it might (and almost always does) need specialized processing technological tools. Historically speaking, it is a product of the information age. In order to better follow this line of reasoning, let's take a look at Edward Waltz's timeline on the evolution of intelligence:

Age	**Agricultural**	**Industrial**	**Information**	
Approx. Period	**Until–1700**	**1700–2000**	**2000–Future**	
Wealth Creation: Power and Business	*Method: peasant-based crop production Resource: land*	*Method: mass production of goods Central resource: raw materials*	*Method: customized production of knowledge services Central resource: knowledge*	
Nation-State Warfare, Conflict, and Competition	*Object of conflicts: land Infantry warfare: attrition of infantry (target human bodies)*	*Objects of conflict: regional economies, access to materials Mechanized warfare: mass destruction of weapons (target mechanized weapons)*	*Objects of conflict: global economies, ideologies Information warfare: attrition of will and capability, precision targeting, speed and agility, management of perception (target the human mind)*	
Focus of Intelligence	*Human collection centric (covert access)*	*Technical sensing centric (remote access)*	*Network centric (network access)*	*Knowledge - centric (perceptual access)*
Intelligence Examples	*Moses, Sun Tzu, General George Washington*	*World War II: radio, radar, cryptography; use of air platforms Cold War: space reconnaissance*	*Post–Gulf War: emphasis on network-centric warfare, battlefield digitization, rapid targeting and data dissemination*	*Future emphasis on human cognition, decision-making and influence*

Table: Edward Waltz's timeline on the evolution on intelligence (Waltz, 2003)

What we learn from the above timeline is that starting with the industrial revolution, the traditional meaning of *intelligence* (analysed information obtained through secret means, based on human collection and covert access) has been expanded to encompass information obtained through technical, remote access sources. The specific difference here is that the data obtained through technical means is, in most cases, not directly intelligible to the customer. It needs processing before it can be transformed into an intelligence product, which adds an extra layer of primary analysis, consisting of the "translation" of the data into something that can be read and further evaluated. Then came the information revolution, and with it, a further expansion of the meaning of *intelligence.* The main issue here is not necessarily adding an extra layer of processing (although that, too, is the case), but an extra layer of *meaning,* of analytic input and evaluation.

The element of continuity throughout this series of evolutions is the main purpose of intelligence, which remains providing the customer with valuable insight that could not be obtained through other means.

There is a lot of buzz these days around **big data**, so much so that skeptics argue it is in fact a concept devoid of meaning, since data has always been big, in comparison to the collection possibilities of its corresponding time. As data gets bigger, so does the capacity of collection and processing tools. Indeed, just like data analysis, big data eludes a commonly accepted definition. However, since it is beside the point to make an inventory of its different understandings, we'll go with Schutt and O'Neil's definition: data is considered "big" when you can't fit (and process) all of it on one machine (Schutt & O'Neil, 2014). Why is the growing size of data relevant to intelligence, however?

The view of the author of this paper is that when considered in the context of national intelligence, data analysis is a blend of quantitative and qualitative research. Data analysis as we understand it now has as a main object evaluating **large chunks of computer-processed quantitative data**, and providing interpretations of it based on qualitative indicators, which can stem from **substantive expertise** in the field of research, the only input powerful enough to provide crucial insights into interpreting quantitative data.

It is clear that human expertise is still very much needed in order to select what's relevant and what should be further analysed. Of course, this was also the case when data was not that big yet, but its growing size poses a

correspondingly growing problem for analysts. Software that selects data making use of key indicators is of course available on the market, but experience has proven that automatically selected data is still far lower in quality and relevance than data selected by humans.

Researchers hope this technical glitch might be solved with the development of **artificial intelligence** (AI), but experiments conducted so far have proven that what humans consider relevant is not regarded in the same way by bots. An AI experiment by Facebook, for instance, was allegedly shut down after the bots started developing their own language, unintelligible to humans (unless linguistically analysed). The researchers assumed that the bots did this because they found human language sequences irrelevant to the tasks they were required to perform. On the other hand, the data scientists that were conducting the experiment called the newly-developed language "functional gibberish" and considered it inefficient (Quora, 2017).

Whatever the real reason behind Facebook pulling the plug on its AI program, one thing remains clear: bots and humans don't speak the same language yet. So if we can't put our faith in AI, what's the solution to our collection-selection-analysis issues?

The impact of big data on intelligence processes, products and human resources

- ***Analysis-driven collection: the solution to smart, efficient collection & processing in the age of big data***

The traditional intelligence cycle begins with planning, and up to a certain point, this still holds. When we think of planning what to collect, what were traditionally taken into account were the issues of national interest which stemmed from previous collection efforts. This is commonly known as **issue-driven collection**, an efficient way of organizing intelligence collection when data was a bit smaller, and definitely not as publicly available. However, issue-driven collection in the information age has led to a set of problems, including collecting more than can be processed, difficulties in ensuring the timely and correct selection of relevant data out of large data sets and the ethical issue of possibly collecting more than can be objectively justified.

The Joint Inquiry into the performance of the US Intelligence Community after the events of 9/11 concluded, among other things, that

the nation's greatest collector, the NSA, had very little time at its disposal to actually process and further select and analyse all the input it got daily. This was found to be the reason why a relevant lead to the future attacks was only processed a few days after the event (House Permanent Select Committee on Intelligence and the Senate Select Committee on Intelligence, 2002).

One of the reforms that took place afterwards was shifting towards **analysis-driven collection**, which is, in fact, a mere recognition of the fact that "collection priorities should reflect the needs of those crafting the analysis, who depend upon the collected intelligence" (Lowenthal, 2016). According to the former DDNI-A, Thomas Fingar, analysis-driven collection actually worked for the American IC, leading, among other things, to more focus from collectors on providing information that would lead analysts to crucial insights, and thus, making selection at least a bit easier (Fingar & Graham, 2010).

- ***Packaging and "selling" intelligence products***

In the present age, information (processed and analysed information, even) is literally everywhere. The massive development of OSINT has had a great contribution to this, since it gave rise to many different entities collecting and analysing public information in order to aid the decision-making process at different levels.

Therefore, for the first time in history, national intelligence became a business that had to compete for its' customers attention and time. Of course, even traditional intelligence products went by tradecraft standards that required them to be attractive and attention-catching, but this usually still only meant that analysts had to pay attention to their writing, since this was their fundamental skill.

Intelligence professionals today are not only concerned with gaining substantive expertise and mastering analytic tradecraft, but also with learning and implementing marketing principles, adapting to different platforms of putting their information forth and learning new, complex skills that go way past the ability to write well, in order to make their intelligence products or services attractive to decision-makers.

Regarding this issue, Ruben Arcos and Randolph Pherson make an excellent point on the rising necessity for analysts to acquire multimedia competencies, in order to better fit the current intelligence market and keep up with today's customers (Arcos, 2015).

- ***The role of the data analyst in the intelligence organization***

Speaking of skills, let us further discuss an altogether new position in the intelligence organization: the data analyst. What sort of creature should she or he be, and what are the corresponding roles to fill? There are, of course, many possible answers to this question. First of all, a data analyst is the person whose job is first-level processing of "quantities, symbols and characters" (which are mostly collected through technical or open sources). She or he must be an expert in the corresponding source of collection, in order to be able to process the collected data by making use of the appropriate software. After this initial stage of "translation", the data analyst selects and analyses the previously processed single-source data by assigning it source and information credibility, in accordance with the Admiralty code or whatever else single-source analysis scale the intelligence organization uses.

The resulting data report, translated into an intelligible form and assigned credibility, is further passed on to the **intelligence analyst,** whose role is to put together such reports from all collection sources, provide an evaluation by interpreting quantitative data in qualitative ways, and transform them into deliverable intelligence products.

This, however, is not the only application of data analysis in an intelligence context. Some types of data can be further processed through statistical modelling, and in such cases, they can also provide support at an operational level:

- **User-level modelling** is the process behind the so-called "online bubble" we all live in. Our online behaviour is constantly being scrutinized, stored and analysed through cookies, which ads companies or social media further use in order to provide us with online content that suits our preferences and interests. In a similar manner, data analysts can use publicly available market research statistics or data obtained from the analysis of social media behaviour in order to profile potential HUMINT sources or to target and monitor individuals that may pose security threats.

- **Population-level modelling** - In business intelligence, one of the main functions of data analysis is to aid decision-makers in establishing and carrying through their marketing and lobbying strategies. The same process of subtly influencing the collective narratives people believe in is employed by secret services or governments, in order to manipulate public opinion. In a TED Talk regarding Russian propaganda, cyberspace analyst Laura Galante describes the process through which data can be modelled and used in order

to erode the very core of the values we believe in, such as democracy or human rights (Galante, 2017).

The reversed process, which can easily be modelled through data analysis, can be of great use in counterintelligence efforts that aim to discover and prevent the enemy from projecting influence through information operations.

Conclusions

The current unprecedented technological advancements have provided intelligence organizations with an important asset: the ability to collect and process large amounts of data. However, the same changes have also given rise to a set of issues that need to be addressed. Modern intelligence organizations can no longer function in a traditional logic; they **need to adapt their processes, products and people**.

1. Big data, little processing - Intelligence organizations collect larger amounts of data than they can process, or, even better said, more than is truly needed.

Solution*: A full shift from issue-driven towards analysis-driven collection is necessary, for the following reasons:*

- to better organize the collection-analysis process;

- to ensure a higher quality input from collectors;

- to allow for more focus on strategic rather than current analysis.

2. Better tools and people who can use them - The 21'st century intelligence analyst is a substantive expert, a master of tradecraft and a proficient multimedia user. Moreover, today's challenges create new positions within intelligence organizations for tech savvy people, needed in order to bridge a current expertise gap.

Solution: *Apart from expanding the recruitment pool for potential candidates, intelligence organizations ought to make sure IC data and intelligence analysts are constantly connected to the best and newest research and training in their field, which means establishing a better and more solid connection with outside experts.*

Intelligence R&D has a key role here, both through providing the sets of requests for the development of software and analytic tools, as well as through being in constant contact with the world outside and its experts.

References:

1. Arcos, R. (2015). Multimedia competencies and the intelligence analyst. In R. Arcos, & R. H. Pherson, Intelligence Communication in the Digital Era. Transforming Security, Defence and Business (pp. 15-16). Palgrave Macmillan.
2. Center For Security Studies. (2007). Emerging Threats in the 21st Century: Strategic Foresight and Warning Seminar Series Final Report. Zurich.
3. Central Intelligence Agency (Office of Public Affairs). (1999). A Consumer's Guide to Intelligence. Washington DC: Central Intelligence Agency.
4. Fingar, T., & Graham, M. M. (2010, April). Five years later, a stronger intelligence community.
5. Galante, L. (2017, April). How (and why) Russia hacked the US election. Retrieved September 16, 2017, from Ted Talks: https>//www.ted.com/talk/laura_ galante_how_to_exploit_democracy
6. Heuer, R. (1999). Psychology of Intelligence Analysis. Centre for the Study of Intelligence. Retrieved from www.odci.gov/csi/pubs.html
7. Heuer, R. (2013). Limits of Intelligence Analysis. Retrieved from www.iwp.edu/docLib/20131120_HeuerLimitsofIntelligenceAnalysis.pdf
8. House Permanent Select Committee on Intelligence and the Senate Select Committee on Intelligence. (2002). Report of the Joint Inquiry into the Terrorist Attacks of September 11, 2001.
9. Krizan, L. (1999). Intelligence essentials for everyone. National Intelligence Meets Business Intelligence. Washington DC. Retrieved mai 4, 2017, from www.strategyawareness.com
10. Lauquer, W. (1985). A World Of Secrets: The Uses and Limits of Intelligence. New York: NY: Basic Books.
11. Lowenthal, M. M. (2016). Intelligence. From Secrets to Policy. CQ Press.
12. National Security Agency USA. (n.d.). Operations Security - Intelligence Threat Handbook. USA. Retrieved august 19, 2017, from https://fas.org/irp/nsa/ ioss/threat96/part02.htm
13. Office of the Director of National Intelligence USA. (n.d.). What is Intelligence? Retrieved august 19, 2017, from https://www.dni.gov/index.php/what-we-do/what-is-intelligence
14. Oxford Dictionaries. (n.d.). Retrieved September 14, 2017, from en.oxforddictionaries.com/definition/data
15. Quora. (2017, August 16). Why Facebook Shut Down its AI Program That Went Rogue. Retrieved September 16, 2017, from Forbes: www.forbes.com/sites/ quora/2917/08/16/Why- Facebook- Shut- Down- its- AI- Program- That- Went-Rogue/#322a5ae1710

16. Schutt, R., & O'Neil, C. (2014). Doing Data Science. Straight Talk from the Frontline. O'Reilly Media, Inc.

17. Waltz, E. (2003). Knowledge Management in the Intelligence Enterprise. Artech House Information Warfare Library. Retrieved from https://phamtrung.wikispaces.com/file/view/KMinIntelligentEnterprise.pdf

18. Wark, W. K. (2009). Secret Intelligence: A Reader. In R. J. Aldridge, Secret Intelligence: A Reader (p. 528). Routledge.

19. Warner, M. (2007, April 14). Wanted: A Definition of Intelligence. Understanding our Craft. Retrieved from Central Intelligence Agency: www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/csi-studies/studies/vol46no3/article02.html#fn7