LUISS

Dipartimento
di Impresa e Management

Laurea Magistrale in Analisi e Metriche di Marketing
Cattedra: Customer Intelligence e Logiche di Analisi dei Big Data

The impact of a Business Intelligence & Analytics Software on an Electronic warfare Company inside the Defence Sector:
The development of the QUIPO Platform to enhance the firm's supplier Scouting.

Prof. Luigi Laura
RELATORE

Prof. Simone Ciccarone
CORRELATORE

Marco Smacchia - Matr. 721671
CANDIDATO

Anno Accademico 2020/2021

Summary

CHAPTER I

Introduction

Since the last decades of the 20th century, we have witnessed a real technological revolution that has completely overturned our society, projecting us into a digital world where space and time take on a new dimension. Most people are connected to the 'web', which not only reduces distance and time to zero but also allows an exorbitant amount of data to be transmitted every day. In recent years, we have entered a new cycle of economic development called 'Industry 4.0' dominated by technologies related to the Internet of Things (IoT), Artificial Intelligence, Big Data analytics, neural and quantum chips and adaptive manufacturing: in other words, the digital revolution of the 21st century (T.E.H. Ambrosetti, 2017, 2018).

This process of information digitisation is total and involves a complete change in the assets and organisation of all enterprises in the global market. Many companies have to cope with "Digital Disruption" and react to the digitization process (T. Genzorova, 2019). In particular, if in the past we were inclined to focus our attention on the ability of machines to relieve man of particularly burdensome tasks, today we are focusing on the ability of computers and the applications that run on them to manage in real-time the mass of data available on the network. In practice, software is dominating over hardware, making the management of large amounts of data a critical success factor. Moreover, digital transformation affects not only the change of individuals, companies and their business models, and time but also the conception of the space. This implies the creation of a new digital dimension that is the pivot of this revolution. Within this new space, the world's most valuable resources are constantly travelling data. Creating new (digital) spaces and modifying existing ones increases the strategic importance of knowledge and control of resources and infrastructures that not so long ago were scarcely used or even unknown.

The most important resource to control and learn about is data. The data economy is currently worth EUR 443 billion in the European Union and the UK[1] alone with an impact on GDP of 2,8% (in the USA, it is 9%)[2], a figure that is set to grow to EUR 1.036 billion in 2025.

[1] Big Data Dossier 2020 available at: https://www.statista.com/study/14634/big-data-statista-dossier/.
[2] Digital economy as percentage of the total economy (GDP) in the United States from 2005 to 2018 https://www.statista.com/statistics/961982/digital-economy-gdp-share-usa/.

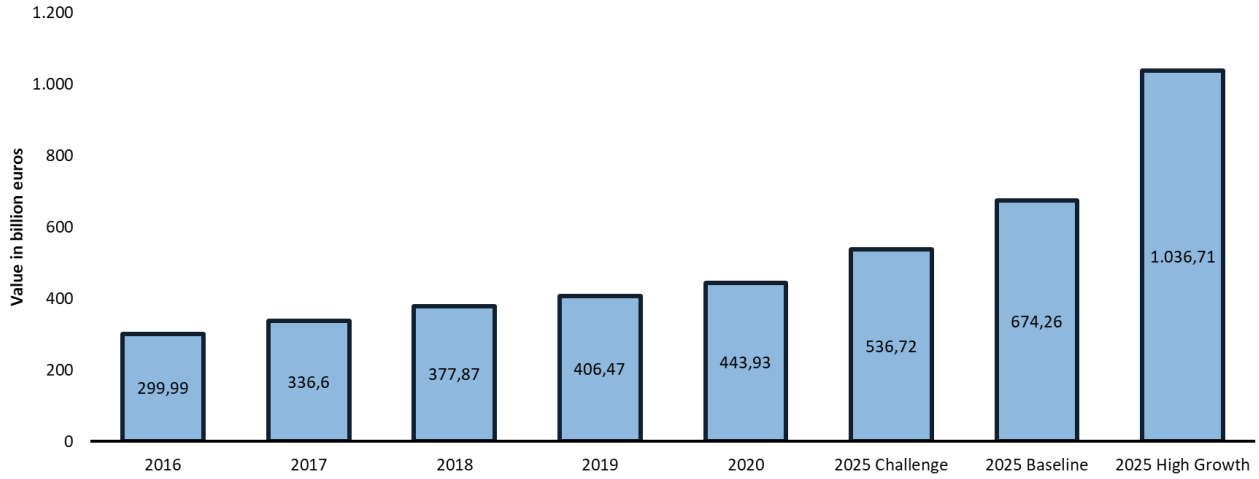


Figure 1: Value of the data economy in the European Union (EU) and the United Kingdom from 2016 to 2020 and in 2025* (in billion euros) (Source: https://www.statista.com/study/14634/big-data-statista-dossier/).

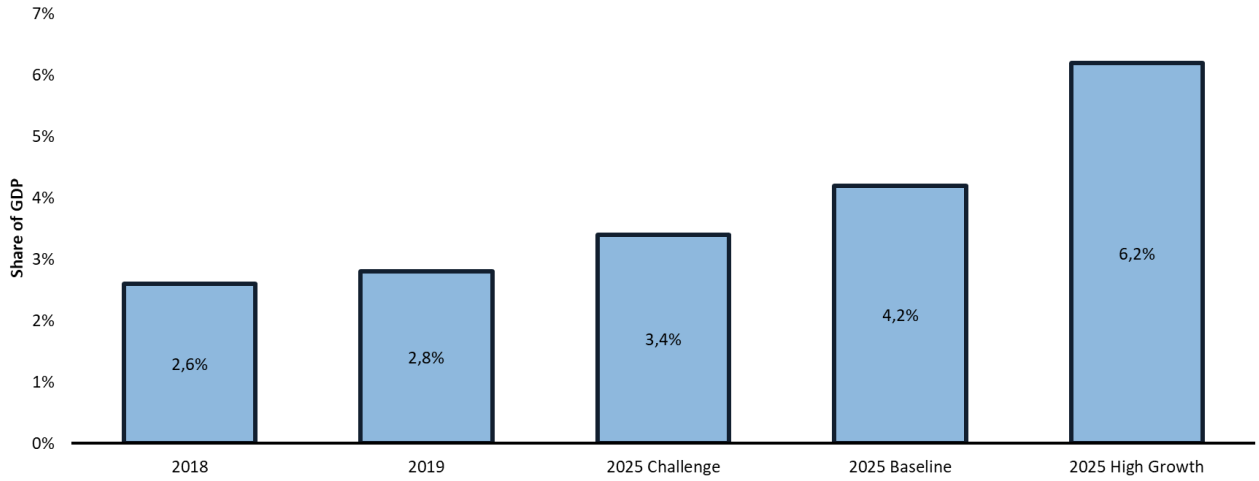


Figure: 2: Impact of data economy on overall GDP in the European Union (EU) and United Kingdom (UK) in 2018, 2019, and 2025 (Source: https://www.statista.com/study/14634/big-data-statista-dossier/).

The continuous growth of connected devices and the constant decrease in their price due to the advancement of technology means that today everyone has at least one or more devices capable of connecting to the internet, a fact supported by the number of active sims on the planet exceeding that of human beings. This phenomenon means that a huge amount of data is sent around the world every second.

It is estimated that in 2024 we could be exchanging 149 Zettabytes (149 trillion bytes) on the internet (now it is 74)[3].

[3] Big Data Dossier 2020 available at: https://www.statista.com/study/14634/big-data-statista-dossier/.

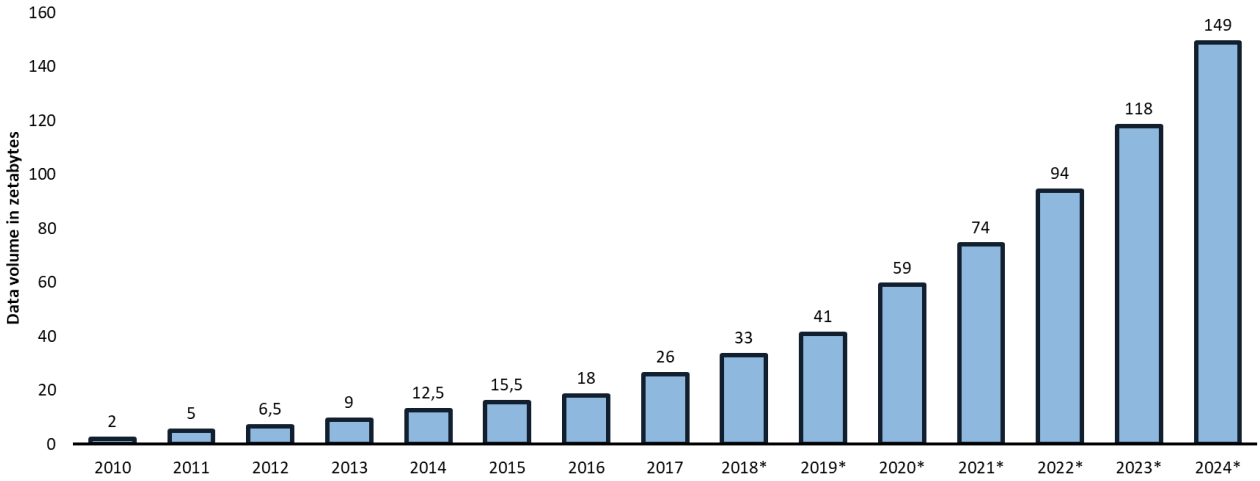


Figure 3: Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2024 (in zettabytes) (Source: https://www.statista.com/study/14634/big-data-statista-dossier/).

This introduces new privacy issues for business and non-business users. Institutions are constantly trying to keep up with technological innovation by passing regulations to protect personal and corporate data: the GDPR approved by the European Commission in 2018 is an example of this[4]. Another key aspect of this revolution is the resources that can manage the data present and available on the network. Companies are constantly looking for people with special skills that are currently very difficult to find due to the delay in adapting university courses to the change taking place. If, on the one hand, robotics will replace humans in many tasks, it is also true that most of the jobs required in the future do not yet exist or are very few[5].

An in-depth analysis of the 'Digital Transformation' cannot be limited to the SW algorithms needed to manage the data present on the network, and therefore cannot ignore the changes that are taking place in the devices on which these algorithms are based. In fact, despite the fundamental importance of software in the digital world, it would not be possible to process Zettabytes of data without the support of suitable hardware. This is why an analysis of the 'cultural' change in the world in which we live should not overlook the sector of components that form the basis of all devices: nanochips. The current evolution of these components involves the use not of silicon (which has now reached its technological limit) but of substrates based on "rare-earth metals". It is through these new components that companies in the technology and weapons sectors develop their products. Therefore, it is strategic for any country that wants to maintain technological supremacy to control the Rare Earth market[6].

[4] Digital Transformation opens up the High Road to GDPR: https://towardsdatascience.com/digital-transformation-opens-up-the-high-road-to-gdpr-8a1b11788e19.
[5] Nei prossimi 5 anni cambieranno 6 lavori su 10. Ecco come: https://www.ilsole24ore.com/art/ecco-come-prossimi-cinque-anni-cambieranno-sei-lavori-dieci--AB67zUaB (Sole 24 Ore).
[6] Le terre rare, ovvero la lotta per l'egemonia e l'Europa evanescente: https://www.econopoly.ilsole24ore.com/2019/07/11/terre-rare-egemonia/?refresh_ce=1

Currently, China is the market leader in the extraction and distribution of rare-earth metals[7], with a monopoly market share of 80%, and is able to influence supply and demand, as we can see during these years within the commercial war against the United States. The ongoing trade war between the two superpowers also involves this fundamental aspect of the digital revolution.

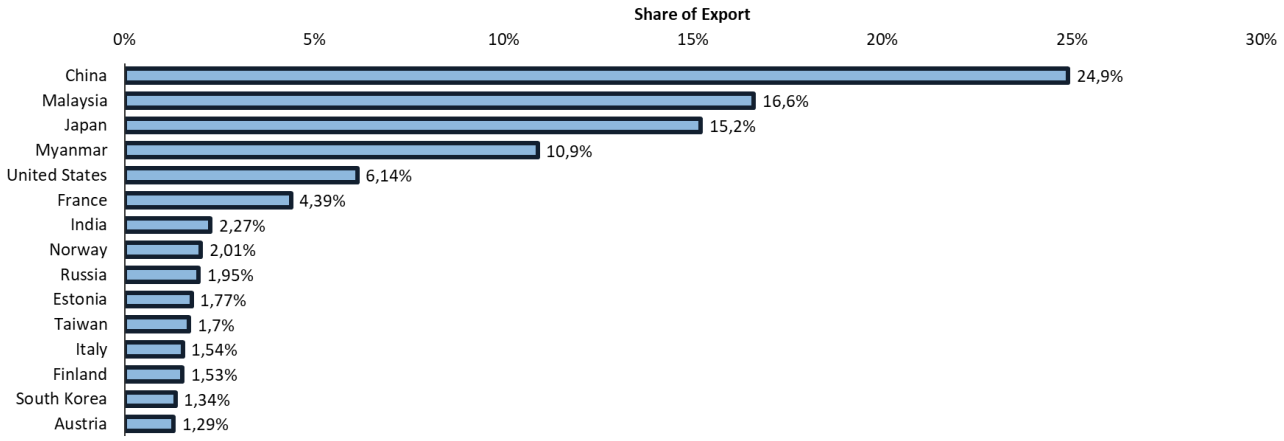


Figure 4: Distribution of rare earth metal exporting countries worldwide in 2019 (Source: https://www.statista.com/statistics/702689/global-rare-earth-metal-export-share-by-country/).

As a last aspect of the brief analysis carried out so far, it is important to mention also the infrastructures necessary for the extraction, exchange and storage of the "network" data. First of all, it is appropriate to focus on the communication networks, be they wired or wireless, which constitute their physical part. The web is made up of first and second level networks that combine to form the Internet as we know it today[8]. Another key infrastructure for data collection is data centres, which enable the implementation of two fundamental pillars of digital transformation: big data analysis and cloud computing that allows the management of large quantities of data and their storage in platforms that can be reached from anywhere in the world. This sector has a CAGR (Compound Annual Growth Rate) of 10,86%[9].

[7] Distribution of rare earth metal exporting countries worldwide in 2019 https://www.statista.com/statistics/702689/global-rare-earth-metal-export-share-by-country/.
[8] Structure of the Internet: https://teachcomputerscience.com/structure-of-the-internet/
[9] Big Data Dossier 2020 available at: https://www.statista.com/study/14634/big-data-statista-dossier/.

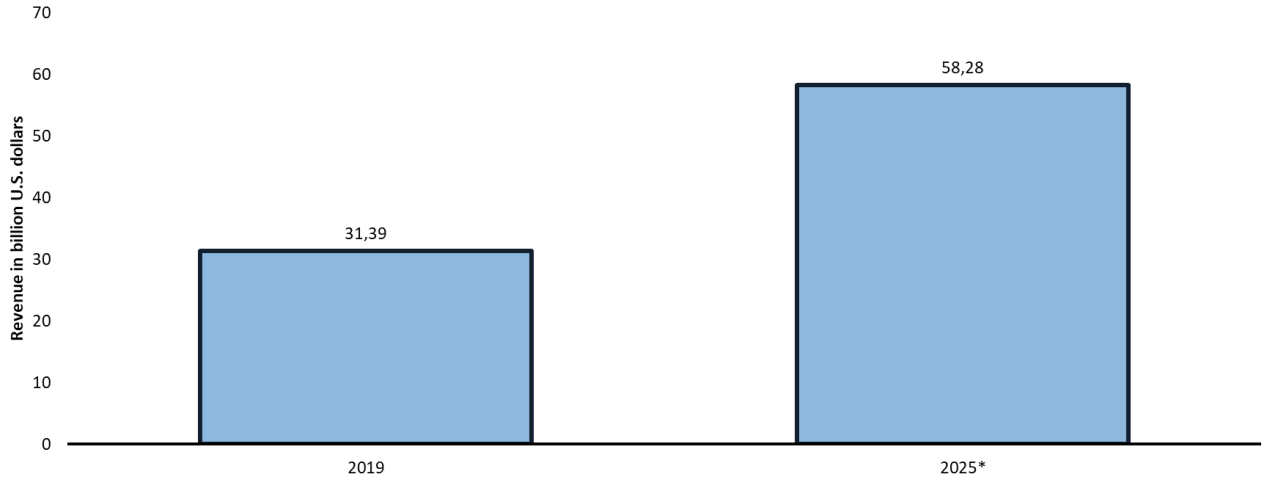


Figure 5: Colocation data centre market revenue worldwide in 2019 and 2025 (in billion U.S. dollars) (Source: https://www.statista.com/study/14634/big-data-statista-dossier/).

The first level networks, in particular, are made up of large cables, most of them submerged, which are able to carry thousands of Gigabytes per second from a continent to another. Today, there are about 450 undersea cables in the world, most of them privately controlled, often by large digital service providers such as Microsoft, Amazon and Google.

The implications of Industry 4.0 are of strategic importance. On the one hand, it represents a great opportunity for companies operating in this high-tech sector that are stimulated to innovate both in products & services and in their organisation and business model. At the same time, however, it represents the main risk factor due to the continuous need for technological innovation and new security standards for their Intellectual Property (IP). The speed with which technology becomes obsolete forces companies, much more than in the past, to invest in research & development in order to secure or maintain the technological advantage that allows them to compete successfully in the global market (Geissbauer, 2016). In particular, the evolution of the threats faced by the various states has led to a radical operational, tactical and organisational change which has forced the companies involved to review their entire product catalogue in order to cope with the new changing needs of their customers.

Enterprises are under substantial competitive pressures (Knight and Liesch, 2016), markets are changing faster and faster, and companies are subject to Digital Transformation challenges. According to McAfee (2012), there is a need to adopt a new source of data called Big Data which is defined as extremely large data sets in volume, velocity, variety, and veracity (Liu 2018). Big Data could help enterprises collect data and convert it into competitive advantages in the global market (McAfee 2012). La Valle (2011) states that the enterprises using data analytics to transform data into useful information outperform their competitors by helping them understanding their business more deeply and improving the decision-making process (Ahi et al. 2017). Consequently, data-driven

enterprises such as Amazon, Microsoft, and Apple in less than ten years have replaced oil companies becoming the biggest firms for capitalisation and profits worldwide (Sole 24 Ore)[10]. Big Data is also getting bigger as information is coming from a larger number of devices and sources. In the last 2/3 years, humanity has generated more data than ever in its history (Raconteur)[11]. It is no wonder that organisation are forced to use only a selection of the vast amount of data available to be effective (Redman 2017).

In this scenario, more and more companies are starting to use Business Intelligence software, applications that can identify and monitor market trends or significant events both inside and outside the organisation faster than human alone can[12]. They are able to transform raw data into intelligence, improving the decision-making process at all levels. In fact, Business Intelligence is a very large market, populated by very different applications, some specialised in specific tasks, others with a wider range of uses. The market forecasts are very promising, expecting stable growth over time. The push towards digitisation given by the pandemic in recent years could also increase these projections.

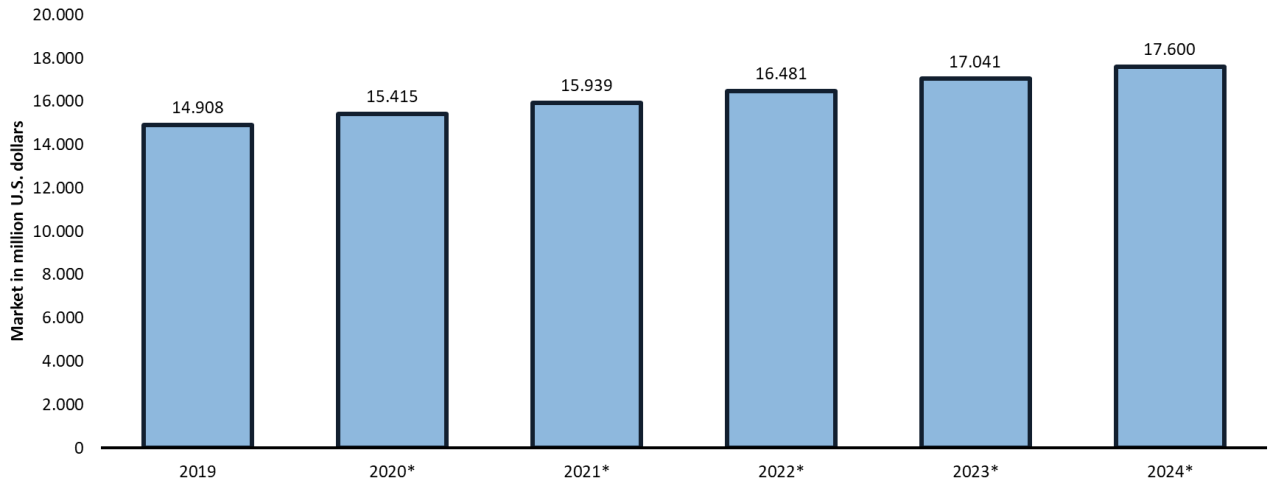


Figure 6: Size of the business intelligence and analytics software application market worldwide, from 2019 to 2024 (Source: https://www.statista.com/statistics/590054/worldwide-business-analytics-software-vendor-market/).

Business Intelligence Platforms

It is no wonder that organization are forced to use only a selection of the vast amount of data available to be effective (Quick, 2018). Considering such relevance of Big Data analysis, many companies must face constant challenges to keep up with digital transformation and become a data-driven company is one of them.

[10] *"Le prime 10 aziende al mondo negli ultimi 25 anni, Il sole 24 Ore 2018".* https://lab24.ilsole24ore.com/aziende-top/
[11] A Day in Data, Raconteur. https://www.raconteur.net/infographics/a-day-in-data/
[12] "Why is Business Intelligence (BI) important to your company?": https://www.linkedin.com/pulse/why-business-intelligence-bi-important-your-company-yass.

During this Digital Revolution, many companies rely on Business Intelligence (BI) software that can help analysts and management to more easily analyse and understand the primary and secondary data collected, facilitating the extraction of meaningful information for decision making.

Regarding BI Software, the two most important factors for its development and integration within a company are the creation of a meaningful data lake and correct taxonomic rules. According to Gibson (Babak, 2016 Gibson et Al. 2016), without data, there is nothing to analyse, and collecting data in a wrong way could lead to inadmissible failures, so it is imperative to collect data for the right reason and in the right amount. The constitution of a data lake following precise data gathering rules is essential to increase the performance of the software, to considerably speed up the manual analysis of the collected or inserted documents, to limit the semantic engine errors when analysing unstructured data and, especially in OSINT research, to facilitate the process of assigning the items reliability. Due to the exponential increase in data, it is becoming increasingly difficult to monitor all the information around us. In this scenario, semantic analysis can help process and visualise online information and extract useful knowledge that helps people make informed decisions. For this to happen, it is necessary to build new taxonomies each time the domain of interest changes (Carrion, 2019). To develop a new taxonomy, it is also essential to have a strong knowledge base on the specific domain (C. Wang, 2017). The creation of structured taxonomic rules is fundamental to increase the software's analysis accuracy and speed up the examination of documents. Analysing the literature, the study of the impact of Business Intelligence is of paramount importance. More research is needed to verify the value of Business Intelligence from both strategic and managerial aspects (Liang, 2018). In this frame, it will be fundamental to investigate the growing application of AI to BI platforms development (J.R.G. Evangelista, 2020). Business intelligence helps extract crucial facts from a vast amount of structured and unstructured data and transform them into useful information that makes organisation strategic decisions, improved operational efficiency and business productivity (Jackar et al. 2020). BI also helps to understand in effective, efficient and automatic way micro and macro/environmental factors. B.O. Hostmann (2007) stated in his research that business intelligence facilitates the functional organisation such as processing offline analysis, data mining, business analysis, business research network, and knowledge management and other related activities. His purpose is to help managers and other company functions to make better and informed business decisions. Organisation uses business intelligence to cut cost, identify new business opportunities and spot inefficient process (Jaiswal et al. 2003). Z. Sheng et al. (2005) stated in their research that business intelligence refers to the tools, technology, application, practices use to collect, integrate, analyse and present organisation raw data into actionable business information. Although the literature has tried to explain the Business intelligence & Analytics software benefits and despite this

is a sector with a positive trend of growth within next five years (Statista 2019), these tools are used only within few business areas and primarily for noncritical issues (Statista 2020). There are different definitions of Business Intelligence; one of the oldest and most exhaustive is Hans Peter Luhn's a researcher at IBM: "an automatic system... developed to disseminate information to the various sections of any industrial, scientific, or government organisation". Business intelligence is the process of transforming data into information, information into knowledge and knowledge into intelligence; with Artificial intelligence software, all this process is cyclical and automated. P. Karbhari (2006) stated in his research that business intelligence is a broad category of application and technologies for collecting, storing, analysing and providing access to data to help the organisation to make a better decision. There are five stages of Business Intelligence:

- Planning all the task to complete the analysis
- Gathering the structured and unstructured data from multiple sources
- Analyse the data and extract information and intelligence
- Prepare a Report to synthesise all the information gathered
- Dissemination of data through dashboard and graph construction

Chen et Al. (2012) and Davenport (2013) divide the evolution of BI software into three periods. The 1.0 phase begun in the 70s and concerns primarily the analysis of structured data with a focus on extraction, transformation and loading (ETL) processes to select decision-relevant data from transactional systems and bring them into the proper format for analyses (Chaudhuri et al. 2011). To store these data, companies mostly used relational database management systems and data warehouses. Data were analysed mainly with the statistical method. With the spread of the Internet in the 2000s, user-generated content and web analytics, collected through Web applications (O'Reilly 2020), are drivers for the second phase of BI. Business Intelligence 2.0 opened new frontiers for data analysis with more complex analysis techniques such as web and text mining and social network analysis (Chen et Al. 2012). With BI 3.0, we get into digital transformation, with data coming from mobile and IoT devices and the analysis of sensor-generated data (Chen et al. 2012). With A Business intelligence software with OSINT (OpenSource Intelligence), SOCMINT (Social Media Intelligence), AI tools, and limitless possibilities given by the Internet, we can access a large set of open information that provides us with a constant updating intelligence to improve decision making at all levels. Literature during these years has proven the benefits of BI, Torres et Al. (2018) explain that Business intelligence helps companies in their Digital Transformation and improve organisational outcomes. BI also leads to efficiency improvement, process optimisation, time and

cost reduction; taking in exam prior studies, we can also see an improvement in profitability, market share, and customer satisfaction (Fink 2017). According to Fink (2017), BI assets generate value via two parallel mechanisms, operational and strategic, and these two dimensions not always are aligned but still lead to value creation. According to Sumera et Al. (2020), Business intelligence could give insights not only in terms of design, manufacturing, production, retailing and marketing strategies but also to solve financial, social and environmental issues by saving money, time, energy consumption and material waste. Results of Business Intelligence studies also help organisation on two fronts: On one side, guide vendors to identify target markets and consumer, and make technology solutions with attractive offerings available, especially in developing countries which are facing more sustainability challenges because of mass shifting of manufacturing units from developed countries to less developed countries due to easy availability of labour and material at low costs (Sumera et Al. 2020). On the other side, help the stakeholders identify the best solution for their firm and develop custom KPIs that best suits them (Tripathi 2020).

OSINT and the Intelligence Cycle

In a 2011 document published by the Office of the Director of National Intelligence, OSINT was defined as: "intelligence produced from publicly available information that is collected, exploited, and disseminated in a timely manner to an appropriate audience for the purpose of addressing a specific intelligence requirement". OSINT is distinguished into two data types:

- Open-source data are published by individuals or groups without privacy restrictions. They have little value when taken individually but, when put together, can give a detailed overview of the situation being analysed.
- Open-source information (OSIF), on the other hand, includes all documents that can be obtained legally through request or purchase by citizens. They are usually more in-depth documents.

Through the use of OSINT tools, one can locate, extract and analyse documents from various sources such as blogs, social media, geolocation data, IP addresses, government documents etc. in order to produce intelligence (Quick, 2018). From OSINT analysis come other types of intelligence disciplines used for a variety of purposes. Among the various examples, we can mention SIGINT (signal intelligence), primarily used in the military, and the related disciplines ELINT (electronic intelligence) and COMINT (communication intelligence). More related to OSINT are instead the analysis of social media (SOCMINT), the analysis of images (IMINT) and Geospatial analysis

(GEOINT) that requires open-source data to be applied efficiently, such as some digital mapping services.

OSINT analysis has had a lot of luck over the years because it is a quick and efficient way to carry out in-depth analysis (Evangelista et al. 2020). It is widely used by many government agencies and in recent years also at corporate level. In fact, many companies that develop open-source intelligence tools are creating customised software for companies. OSINT is used to analyse links between various people, organisations, devices and many other entities (Maciołek, 2013). In the last two decades, there has been a shift from analysing to find hidden information to analysing to find relevant information (Nicart et al., 2016). The growth of data has created an environment where we are constantly flooded with information. It has become very difficult, if not impossible, for analysts to conduct manual searches without the help of software. In this scenario, artificial intelligence and machine learning algorithms are increasingly being used in Business Intelligence projects to partially manage various stages of the intelligence cycle.

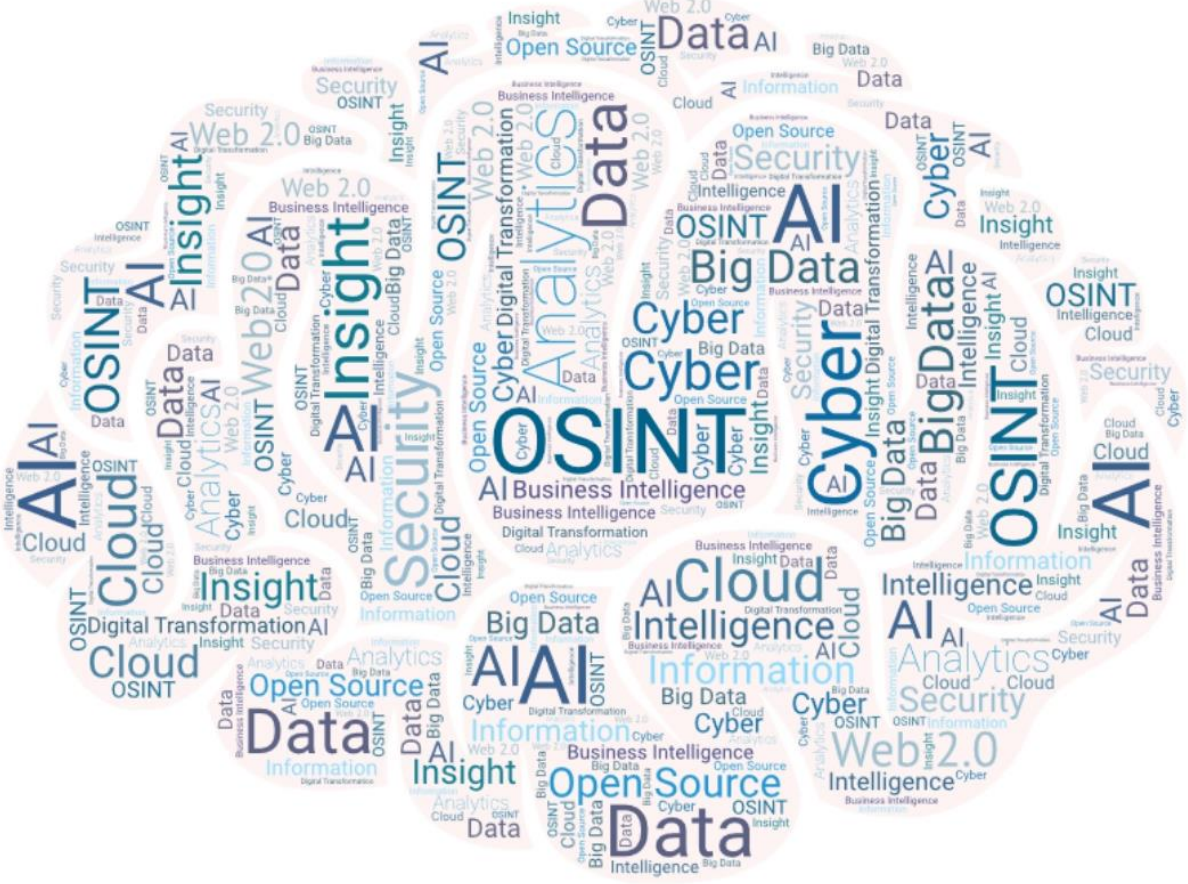


Figure 7: OSINT word cloud.

Like many technologies we use today, OSINT was initially developed in the military industry. It was used by the United States, more specifically by the Foreign Broadcast Monitoring Service (FBMS), to monitor and analyse the propaganda programs of the Rome-Berlin axis during the Second World

War (J. Roop, 1969). After the end of the war, the FBMS was absorbed by the CIA, becoming the unit responsible for collecting, translating, and analysing public domain documents. During the Cold War, the agency's work was fundamental in monitoring the actions and strategy of the Soviet Union, with about 80% of the information used to monitor the collapse of the USSR is attributed to open-source documents (Adm. Studeman, 1992). During the 1990s, with the advent of the Internet, it started to become evident how much the concept and function of OSINT were about to change due to digital technology and how important it would be, so in 2005 the FBMS was replaced within the CIA with the creation of the Open-Source Center (OSC): (Riddel, 1992). The new office was created to go beyond just translating documents, dealing with the entire intelligence cycle, as well as the development of new tools and technologies. In 2015, the OSC was renamed the Open-Source Enterprise (OSE). It was incorporated within the Directorate for Digital Innovation, focusing on cyber threats and the development of new analytical tools (Williams et Al, 2018).

Although there are various interpretations of the intelligence cycle, it is commonly defined in 5 steps[13]: planning, collection, processing, analysis and dissemination. These stages include the acquisition and validation of information, the identification of the value of the information and the distribution of the information to the client.

- In the planning stage, the purpose of the analysis is planned, along with the activities, timing, case manager, and all the people who will be working on the case and possibly the sources where the data will be obtained.
- During the data collection phase, the sources of potentially helpful information that will need to be collected are identified. In this phase, the analyst's ability to recognise and prioritise the suitable sources is essential to facilitate the next steps. Whether it is manual or automated OSINT analysis, not all data have the same ease of collection and retrieval. While newspaper or blog articles are very easy to find, grey literature often requires extensive research and access to particular databases. Despite appearing to be easy to find, social data presents many problems, firstly because of the vast amount of data available (think of the tweets published every day) and secondly because of many problems related to privacy. Much of what is published on some social networks such as Facebook and Instagram is not open source, so care must be taken with the data that is collected.
- In the processing step, the data is validated and normalised, making it usable for analysis. While before this step consisted only of document translation, with the advent of web 2.0 normalisation has become much more complex. In fact, there has been an exponential growth

[13] The intelligence process/cycle (Source: “Joint Publication 2-0, Joint Intelligence”. Defense Technical Information Center (DTIC). Department of Defense. February 2013)

of data and data formats, so the translation is accompanied by conversion, transcription and many other actions that make a document ready for analysis. At this stage, the support of digital tools is becoming essential to gaining intelligence.

- In the analysis phase, the relevance of information is determined. In fact, it is established which information is reliable and meaningful with which to compile the report. At this stage, the ability to adapt to the organisation's needs and distinguish meaningful data is crucial. The analyst must authenticate the information by understanding if what it says is accurate and also assess its credibility and contextualise it. This is relatively easy for news from official websites and scientific articles. It becomes much more complex in social media analysis.
- In the production phase, the information is made available to the end-user. In this step, data is disseminated through the compilation of a report and the creation of dashboards. This last aspect has become very important in recent years as more and more tools specialise in data visualisation.

According to Evangelista's study (2020), the main areas of application for OSINT software are Information Security, Digital Marketing and Artificial Intelligence. This last technology is used in software development through the adoption of Machine Learning and Natural Language Processing that allows to make a first analysis in a fully automatic way. The study concludes by saying that presently the most studied area in literature is Cybersecurity, followed by social media and military purposes (Evangelista et Al, 2020).

Research Question

In light of this analysis, the study of the impact of Business Intelligence is of paramount importance because:

- Most previous research focuses on the positive side of Business Intelligence for promoting the technology, but it has not yet proved the value creation or the avoidance of negative impact (Liang 2018).
- Most existing case reports are based on anecdotal evidence (Liang 2018).
- More research is needed to verify the value of Business Intelligence, both strategic and managerial (Liang 2018).
- Data exploration techniques such as data mining could be helpful to have more insights into Business Intelligence utilisation and success (Ain, 2019).
- It is recommended to investigate the application of OSINT with AI (Evangelista et Al, 2020).

This research is a starting point to study the validation process of a specific “use case” resulting from BI platform implementation by answering, through the formulation of a conjecture, to the following research question:

"In the frame of the implementation of an OSINT Business Intelligence Platform Analysis in the defence market, does the creation of a meaningful Data Lake together with an appropriate Taxonomy structure make the experimental analysis results comparable and coherent to equivalent analysis validated by internal experience and know-how of the firm ?".

In this frame, a segmentation of the BI Platform Market is performed in order to give an overview of the positioning of the QUIPO, a BI platform developed by Cy4Gate S.p.A.

Research Aim

This work aims to study the impact of a BI software platform on the field by evaluating both the positive and negative sides of its use and validating with empirical data its functioning to help organisations with the decision-making process. More specifically, the purpose was to inspect the application of a Business Intelligence approach that applies OSINT (Open-source intelligence) and SOCMINT (Social Media Intelligence) techniques to the Electronic Defence Market to investigate how much this technology could facilitate Companies’ decision-making process by providing them with a distinct competitive advantage. In this setting, the QUIPO intelligence platform (developed by Cy4Gate S.p.A.) will be used for an industrial analysis purpose within the Electronic Defence Sector. The potential growth, scalability and adaptability of the QUIPO application will be assessed to extend its use to other business areas. The research intends to find an empirical explanation according to what literature reports. It is indubitably important that companies understand how and why the adoption of Business intelligence software is essential in the Digital Transformation era. Many companies have difficulties coping with technological progress.

According to Statista (Statista, 2018), in many sectors, the penetration of Business Intelligence solutions is under 10%. Only the Insurance and Business Services sector are most likely to adopt this technology. Although the adoption trend is increasing, many organisations don’t use (or use only for noncritical operation) BI software. Nonetheless, data gathering and data analysis are central to every organisation that operate in all sectors. Through data, every business can unlock new paths leading to a significant competitive advantage (Forbes 2021).

Nevertheless, going digital is not always straightforward; organisations must understand what suits their needs best and how well these technologies work. With more and more companies that started

to offer BI services, there is also a choice overload with whom is difficult to cope with. According to that, more empirical examples must be studied that should give support to those who are reluctant or scared to change their business. The research must also consider that this type of studies could help companies that sell these products. In fact, precious insights could be gained, such as target market and consumer, differentiating factors and new ways to sell their products (Sumera et Al. 2020). Finally, enterprises must measure everything inside the organisation, and an example of BI software implementation could give an idea of the metrics involved in these processes. This research is a starting point to study the validation process of a specific "use case" resulting from BI platform implementation compared with OSINT and internal company knowledge within a Defence company. A second step will be to understand how such software can be extended to all business area. From a research standpoint, the study will also contribute to explain the evolution of Business Intelligence through Digital Transformation in Defence Industry.

Research Methods

In order to study what mentioned above, is employed data from internal company databases and companies' websites to identify competitors to the QUIPO platform and to segment the market. In relation to the research objectives, will be made a market segmentation through the construction of a two-variable positioning map. The aim is to examine a sample of 14 companies operating in the business intelligence market and position them on the map according to their focus on the use of OSINT data and the possibility of customising taxonomies.

In the second part of the study will be used an applied research method to empirically validate the use of the BI platform when applied to a specific "use case". According to Kothari (2008), Applied research "*aims at finding a solution for an immediate problem facing a society, or an industrial/business organisation, whereas fundamental research is mainly concerned with generalisations and with the formulation of a theory*" (Kothari, 2008).

More specifically, the study will go through five stages to develop the experimental research on the QUIPO platform:

- Planning all tasks to implement the job
- Starting the Data Gathering process from multiple OSINT sources
- Data Analysis implementation through the creation of a specific taxonomy for the case study
- Creation of a report
- Data Dissemination

Each phase will be procedurally and analytically analysed and issues will be reported along with suggestions for desired changes.

Then, to validate the platform's performance, it will be studied the correlation between a manual OSINT study based on Market and Company Reports together with data available on the web and an experimental analysis performed entirely with QUIPO. Variables will be created based on the information contained in the QUIPO dashboards and compared with those obtained from the manually compiled Excel file using Pearson's correlation index. The research will be conducted with a dedicated team within the Company.

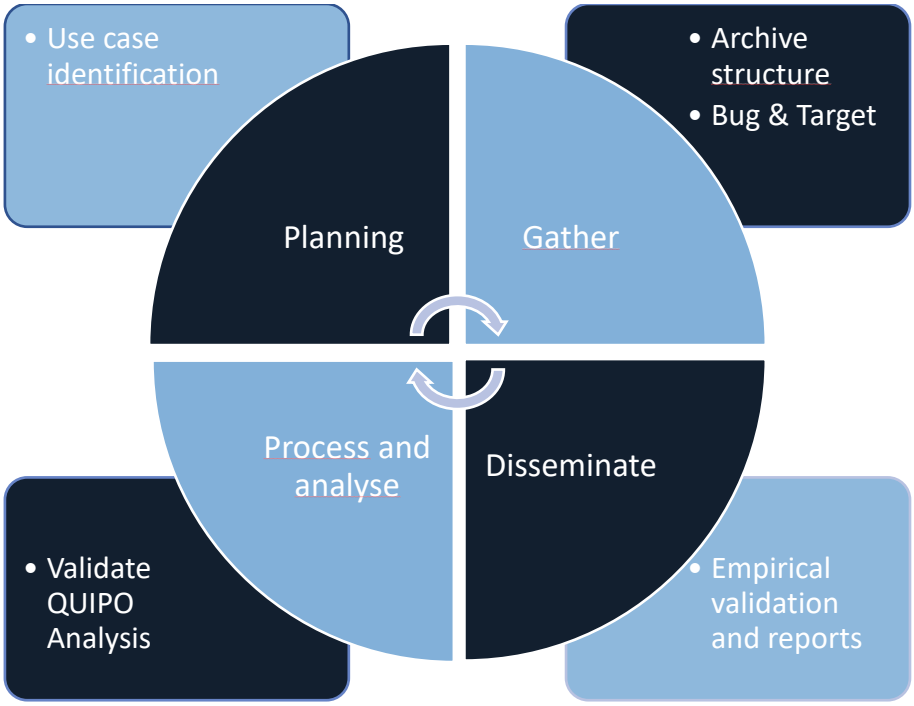


Figure 8: Steps needed to finalise the research.

CHAPTER 2

Market Summary

In recent years, due to the exponential growth of data available online[14], there has been a shakeup in the OSINT software market, especially during the pandemic period related to the spread of the Covid-19 virus, which was followed by the increase in information shared on the Internet, with data traffic growing by more than 100% globally[15]. Many companies, as a result, are investing in these platforms, either through the acquisition of companies or internal development. The market is currently worth about $6 billion and is growing steadily with an expected CAGR of more than 20% by 2027. North America currently dominates the market, also aided by the many government agencies using Business Intelligence platforms such as Palantir. However, it is expected to grow globally in the coming years, especially in the Asian market. Open-source Intelligence software is used for numerous purposes in numerous industries. They can prove useful at the corporate level to conduct market research and formulate strategies by analysing social media and breaking news. At the same time, they are used by government agencies for national security and overseas missions through continuous collection and analysis of breaking news, combined with the data in their possession. Lately, these platforms are also widely used in Cybersecurity to prevent cyber attacks and discover systems flaws. The most significant players often coincide with companies operating in the field of defence, including Palantir and Hensoldt, after purchasing Sail Labs. Among the Italian players, Expert Systems stands out as a partner of Cy4Gate in the development of the Quipo Business Intelligence platform.

Palantir

Overview

Palantir Technologies is an American company specialised in the development of Big Data analytics software. The Company was founded by the entrepreneur Peter Thiel in 2003 and entered the stock market in 2020. It is currently valued at $41.147 billion.

[14] «Il traffico internet è cresciuto con la pandemia, la rete ha tenuto grazie al cloud» - Il Sole 24 ORE https://www.ilsole24ore.com/art/il-traffico-internet-e-cresciuto-la-pandemia-rete-ha-tenuto-grazie-cloud-ADqZk1HB.
[15] Come è cambiato il traffico su Internet durante l'emergenza coronavirus (agi.it) https://www.agi.it/innovazione/news/2020-04-29/aumento-traffico-dati-internet-coronavirus-lockdown-8471456/.

Palantir has 139 customers and is used in more than 150 countries, including corporate clients such as the investment company Morgan Stanley, Scuderia Ferrari, defence giant Airbus and FCA Group[16]. However, the Company's real strength lies in its collaboration with US government agencies, both in the civil (National Institutes of Health) and military and intelligence fields (FBI, CIA, U.S. Air Force, LAPD). The US market is key for Palantir, as it accounts for 52% of its revenues[17]. Since 2018, it has been part of Project Maven[18], the US government project to analyse surveillance videos collected by unmanned aerial vehicles, after Google stepped away. The Company repeatedly emphasises that technology alone is not the answer to all problems but must support human beings in analysing events and making decisions.

Ecofin Data

In 2020, Palantir had revenues of 1.1 billion, a 47% increase over 2019 and a contribution margin of 594 million. It has 139 clients, with a $7.9 million revenue per customer (41% higher than 2019). Of the 139 customers, the top 20 make up 61% of its revenue. This is one of the company's highest risk factors since even losing just one of these customers could have severe repercussions on its business. Palantir focuses on both the government sector with Gotham and the commercial sector with Foundry. Reports show that its relationships with government agencies are worth 56% of its revenues, with the remaining 44% dependent on its commercial clients. As mentioned earlier, its business base of 52% of its revenues is mostly located in the United States, with the rest spread across more than 150 countries. In the last period, however, there has been significant growth in exports, one example being its partnership with Airbus[19] to manufacture and operate more than 9,000 Airbus A350s worldwide. Although the economic consequences due to the Covid-19 pandemic are still uncertain, reports indicate that there does not appear to have been a slowdown in business as most tasks are managed remotely. In addition, increased efficiency in sales resulted in more customers in 2020. Palantir has invested $560 million in research and development in the last year and $2 billion since 2008[20].

[16] Ipo in vista per Palantir, la società tech di dati sensibili finanziata dalla Cia: https://24plus.ilsole24ore.com/art/ipo-vista-palantir-societa-tech-dati-sensibili-finanziata-cia-ADiKDed.
[17] What you need to know about Palantir's direct listing: https://www.protocol.com/palantir-ipo-what-you-need-to-know?rebelltitem=4#rebelltitem4
[18] Leaked Palantir Doc Reveals Uses, Specific Functions And Key Clients: https://techcrunch.com/2015/01/11/leaked-palantir-doc-reveals-uses-specific-functions-and-key-clients/.
[19] Airbus believes its AI can eliminate flight delays: https://venturebeat.com/2018/10/22/airbus-believes-its-ai-can-eliminate-flight-delays/
[20] Palantir Technologies Inc. ANNUAL REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1 0001193125-21-060650 (d18rn0p25nwr6d.cloudfront.net).

Business Model

Palantir's targets are large corporations or government agencies that have tough challenges to achieve. The company tries to fit in where there are risk-averse companies that hadn't a positive outcome from their previous projects. In this scenario, the barriers to entry for competitors are very high due to the high risk of failure, the very long sales cycle, and the situation's complexity. One of Palantir's strengths is focusing on partnerships within its sales channel, signing agreements with public and private cloud service providers, and joint ventures to expand opportunities internationally. Currently, Palantir expects revenues from existing contracts for 3.6 years equal to the dollar-weighted average contract duration basis. The company's business strategy is based on three steps: Acquire, Expand and Scale. In the first phase, the company generally works at a loss through trial periods that have no cost to the customer to make the platform's added value. In the “expand” phase, you invest in the customer to understand the main challenges they face and to make sure the software delivers real results. In the last step, however, the contract has reached its maturity stage and the Company's investment normally decreases, while the customer's added value from using the software grows. In this phase, the customer is almost self-sufficient in the use of the platforms. Only operation and maintenance support is required by Palantir.

Risks

Despite being a very innovative company that has had record profits over the years, Palantir has to deal with many issues that could slow its growth, if not stop it altogether. First of all, one of its big problems is that it depends on a few customers since 61% of its revenues depend on its top 20 customers. The same could be said about its high volume of business in the United States, especially with government agencies. In these times of economic uncertainties, with the entry of several competitors and the continuous evolution of the US and international political scenario, the probability of losing vital contracts is very high and even the loss of one key customer could bring serious problems to the company. This is reinforced by the fact that the sales effort for new account acquisition is very high in terms of time and money, with no guarantee of taking the customer from the acquire to the expand phase.

A second problem is an increase in the number of cyberattacks on companies in recent years. Palantir is exposed both in first and third person since it needs third party devices and software to provide its services. A potential loss of its customers' data would be massive damage to the Company's image from which it may not recover, which is compounded by its heavy reliance on contracts with U.S. government agencies.

Finally, the risk arising from the pandemic should not be underestimated, not only for Palantir but especially for its customers; in fact, many businesses are having great difficulties and could reduce or interrupt their relationship with Palantir.

Products

Palantir offers three products: Apollo, Gotham and Foundry[21]. Apollo[22] lies at the heart of the other two platforms, is a system that continuously delivers updates and new features in Gotham and Foundry. Each platform consists of hundreds of services, each managed by a development team. Thanks to its modular approach, Apollo can upgrade the other two platforms for individual services independently, which ensures speed of development and deployment. This is done irrespective of the software application environment, so it applies whether the platforms are delivered via the cloud or installed on the customer's servers. This has enabled Palantir to manage updates to its platforms independently of environmental restrictions and without requiring any intervention from the customer.

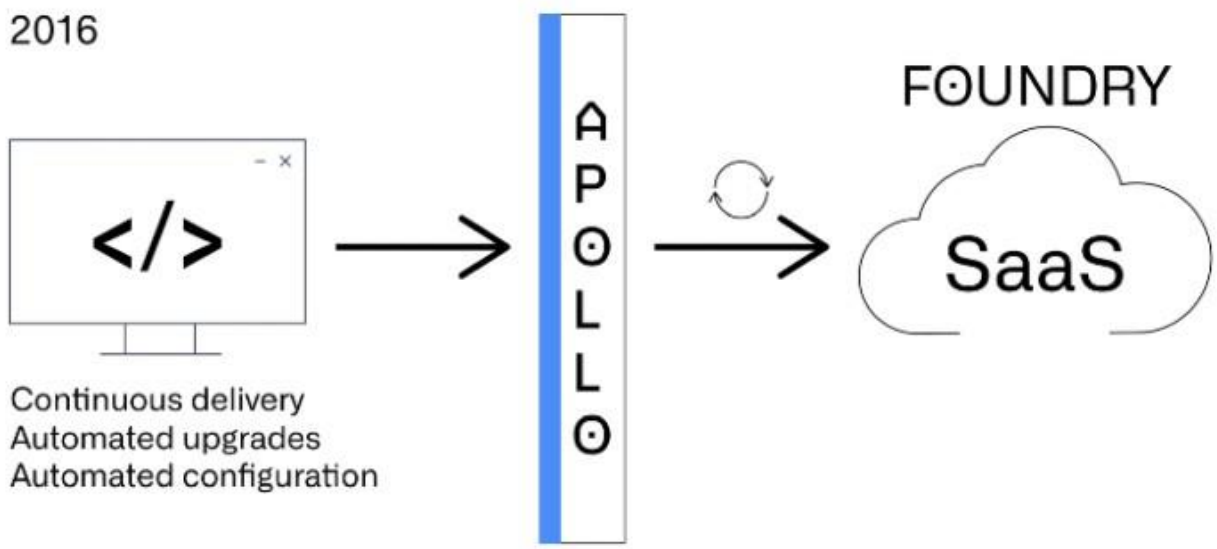


Figure 9: Palantir Apollo Operations (Palantir Apollo: Powering SaaS where no SaaS has gone before, 2020).

Palantir Gotham is an intelligence software explicitly designed for government and intelligence agencies. It is mainly used as decision support in missions, investigations and Cybersecurity. Starting from a database, primarily internal, Gotham can collect, manage, and analyse data in real-time, thanks to compatibility with various types of sensors, giving insights and allowing analysts to manage the mission and communicate in real-time with all the people involved. Thanks to machine learning

[21] Palantir Homepage - https://www.palantir.com.
[22] Palantir Apollo: Powering SaaS where no SaaS has gone before: https://blog.palantir.com/palantir-apollo-powering-saas-where-no-saas-has-gone-before-7be3e565c379.

algorithms, it is able to transform data into useful information, drawing plots, maps and graphs, always indicating the sources used. In addition, Gotham can be enriched with different modules that allow you to automatically analyse information and set alerts as soon as there is a connection with the object of the study, to write dossiers that multiple users can edit, to organise troops in real-time for field operations, making communication much easier thanks to the integration with sensors and mobile devices and finally to synchronise your data in real-time around the world.

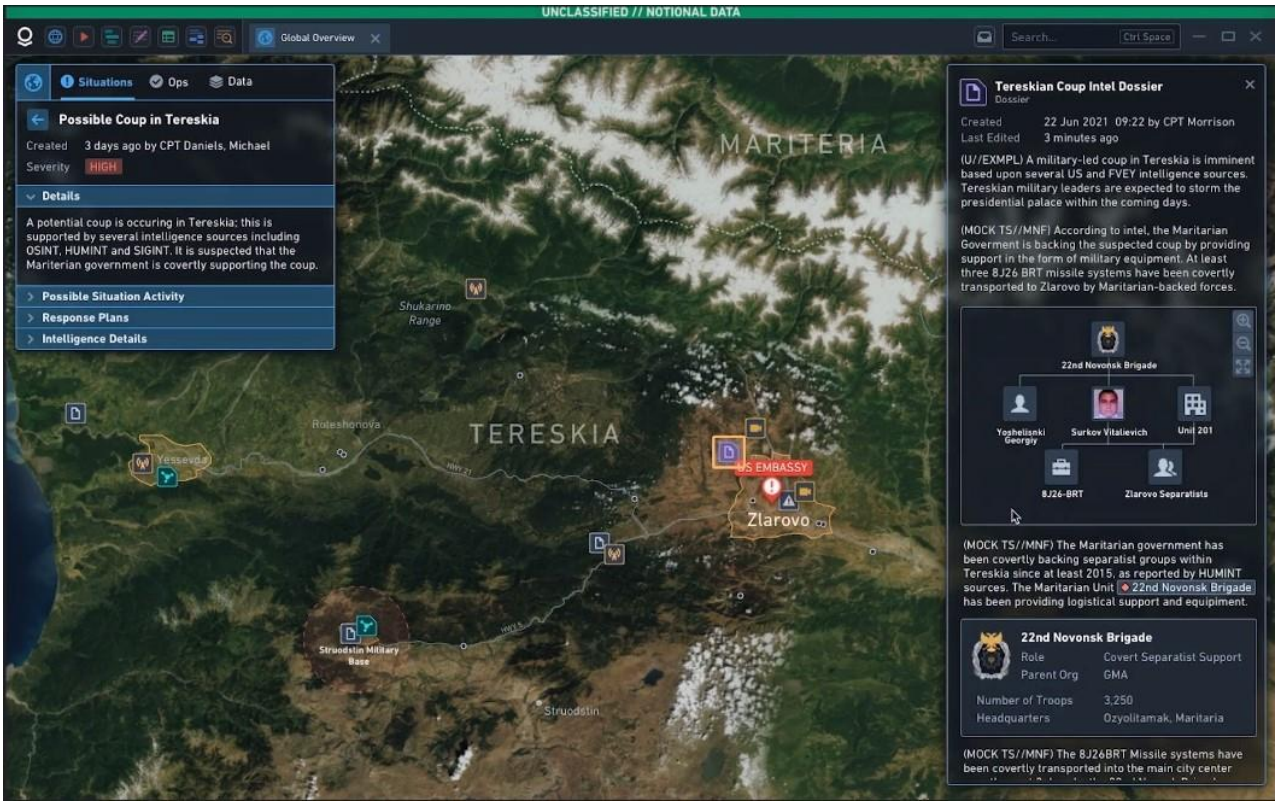


Figure 10: Palantir Gotham Demo (www.palantir.com, 2021).

Palantir Foundry is a Business Intelligence software that allows companies to collect their data and analyse it quickly to get useful intelligence in the decision-making process. The platform is currently used in various fields such as finance, health and social sciences, insurance, legal and manufacturing, to name a few. The system can automatically recognise multiple data types from various sources and manipulate them to make them clear and understandable. The goal of Foundry is to simplify back-end data management and facilitate front-end analysis so that the application is usable by most users within the enterprise. The platform's strengths consist in the speed of data representation, in the uniformity of concepts given by the construction of a common ontology, so that everyone in the Company speaks the same language and in the ease of use.

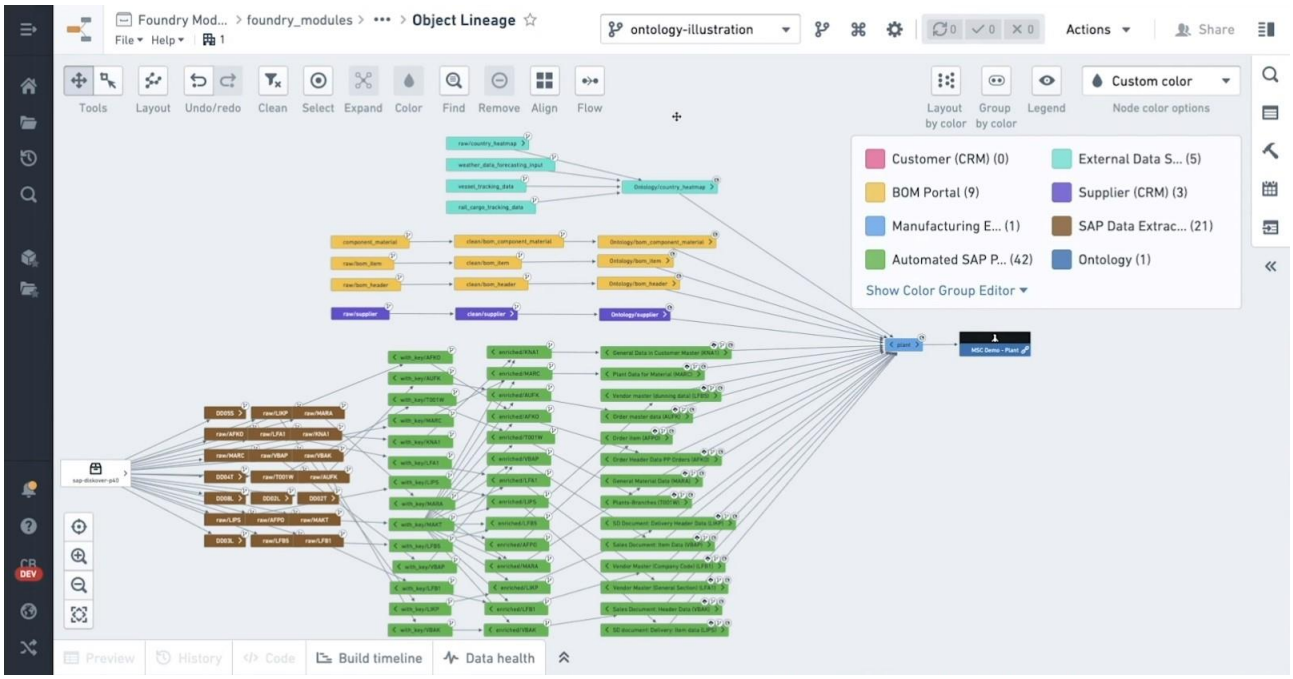


Figure 11: Palantir Foundry Demo (www.palantir.com, 2021).

Business Cases

As mentioned before, Palantir's platforms allow to store, query and visualise very large data sets, helping analysts to solve problems, analyse an event or formulate a strategy by finding essential information in a sea of data (Munn, 2017).

However, in order to conduct an analysis that reflects reality, two essential steps are required. First of all, platforms must have a considerable amount, variety and speed of data available. In this part, confusion, ambiguity and repetitiveness of data are an indication of authenticity. All of this is manageable through Palantir's back-end technology that can capture large amounts of data and process it quickly. Initially, an extensive database must be achieved so that it is meaningful and can represent the big picture. The limitation of having so much data is in its inhomogeneity; it becomes impossible to rely on the relational schema of traditional databases, making it mandatory to switch to a non-relational schema that provides more flexibility in data fusion. Palantir specifically uses Cassandra, a NoSQL database capable of managing data from different sources more effectively.

Secondly, the data must reflect what is happening in the present. Therefore, it is necessary to constantly update the data on the platform. However, it is not only sufficient that new data is continuously available. It must also be available for analysis in a very short time. To solve this problem, Palantir uses Map Reduce, a fundamental component of the Apache Hadoop system. Map Reduce allows operations to be divided into different groups and mapped to different nodes. Once the operations of each node have been completed, the partial results are reduced to a single final

result. This process allows you to harness the power and scalability of cloud computing, making it faster and more efficient to process large amounts of data.

With the Back-End processes, where data are grouped and processed, a thorough analysis of the Front-End processes is essential. It is, therefore, necessary to make sense of the collected data. This can be done in the platforms through automatic and manual tools that allow to analyse key signals within the data lake, examine paths and find relationships.

- Search Around is a tool that allows you to search for all relations in an item. It is represented by a graph where the sought-after item is at the centre, linked to all the other nodes according to the interconnected relations. It is possible to search for different types of objects, from people to IP addresses and even telephone calls from a given number. The relationship between items is established through a metonymic (and non-taxonomic) logic. Associations between two people, for example, are made by examining specific information instead of analysing all the related concepts.

- This tool produces an array of effects, each tied closely to its formal properties. Flows crystallise, solidifying connections between entities. Flows, on the other hand, is a plugin that allows material flows to be displayed. Flows make it possible to visualise a network of relations between various objects collected in several clusters and to prioritise the most critical relations in each group, giving the analyst a view of the essential occurrences within the network.

- Timeline allows you to specify a time window in your analysis regarding, for example, a specific action. This allows the analyst to observe a particular phenomenon at a specific time, allowing him or her to discover a significant path or monitor a specific activity. An example would be to use a timeline in an investigative inquiry to see if there is a relationship between certain events that repeat over time.

Both of Palantir's platforms are widely used globally for various purposes. Airbus uses Foundry to monitor multiple components' operations, both for their development and to predict maintenance actions. J.P. Morgan, on the other hand, uses it to identify employees who commit fraud. The LAPD uses Gotham for Operation Laser, a program to identify and deter people who might commit a crime. Intelligence agencies in the United States use it internally to prevent terrorist attacks and abroad during battlefield missions.

Among these cases, two of the most interesting and controversial are the use of Gotham by the Los Angeles Police Department (LAPD) and the use of Metropolis by J.P Morgan. The LAPD was one of the first U.S. government agencies to adopt Palantir for predictive policing in his Operation Laser. The main target of the platform was to get as much information about the individuals in the territory

as possible to have a system that could recognise and trace anyone. One of the tasks of Palantir Gotham was to collect data from surveillance cameras that had the function of reading car license plates, analyse them and compare the license plate number with federal databases to check for irregularities. In this way, it was easy to identify if the car was stolen, if it belonged to a criminal who had violated the house arrest or if he had simply not paid some fines. Once reported, the agents could proceed with the necessary operations. In addition, Gotham collected data on purchases, word reports, toll road information, social media, hospital records, interrogations and many other sources. In this way, it was possible to arrive at an almost perfect identikit of the person and even have a network of his relationships. However, platform errors have often caused officers to make serious mistakes; there have been several instances of incorrect readings of a license plate leading to wrongful arrest. Or the association of a person within a gang just because he appeared in some photos with the exponents with the consequence of being stopped decide times for crimes that you have not committed.

Palantir has also been deployed within J.P. Morgan, one of the world's largest investment banks. Thanks to Palantir's platforms, an internal task force within the company was able to store and analyse all data from corporate cell phones, computer printers and badges. If an anomaly was found, the employee was immediately searched. During its use within the company, it proved useful in identifying those responsible for leaks. However, its deployment has also caused discontent within the workplace, with many employees complaining of excessive monitoring.

Hensoldt Analytics (formerly Sail Labs)

Outlook

Founded in Vienna in 1999, Sail Labs is considered one of the leading players in providing artificial intelligence-based OSINT services. Its platform can collect, process, analyse and represent both unstructured and structured data. Its strength lies in automatic speech recognition in more than 30 languages. In early 2021, it was acquired by German defence giant Hensoldt[23] in order to integrate this technology with its systems. After the acquisition, Sail Labs was incorporated within Hensoldt Analytics, which inherited and expanded its functions. Hensoldt Analytics is also one of the founding

[23] HENSOLDT Acquires SAIL LABS - European Security & Defence (euro-sd.com) https://euro-sd.com/2021/02/news/cyber-coms/20935/hensoldt-sail-labs/.

members of EUROSINT, the European forum on open source intelligence present in many projects worldwide[24].

Ecofin Data

Hensoldt Analytics has total assets of approximately $6 million, most of which is debt. Until 2019 the Company was considered a high financial risk and was still orphaned by Hensoldt, whose purchase was central to its business; it will be interesting to see future financial data[25].

Products

HENSOLDT Analytics System is an integrated platform for analysts and decision-makers that extracts metadata and key information from multiple sources in various languages and in real-time. The platform allows you to collect data from various external sources, such as major social media, search engines, news websites and traditional media in different formats (video, text, images), process the collected media and analyse them to produce relevant information. Its main strength is mainly due to the automatic ASR speech recognition that transcribes audio and video files and translates them into about 26 languages. Among its other functions, Hensoldt analytics enables the creation of graphs and dashboards for the representation and comparison of large amounts of data. It allows analysing text extracting concepts, topics, and logos. It can process videos conducting sentiment analysis on the speech, identifying keywords and the speaker. Another very interesting feature of the platform is the presence of the Crisis Room, which allows you to monitor live media streams from various communication channels in a single dashboard, useful in the most critical situations.

[24] HENSOLDT Analytics System - HENSOLDT Analytics (hensoldt-analytics.com) https://www.hensoldt-analytics.com/hensoldt-analytics-system/
[25] ORBIS, Data Update 209,001 (02/04/2021) © Bureau van Dijk 2021 Hensoldt Analytics Complete book.

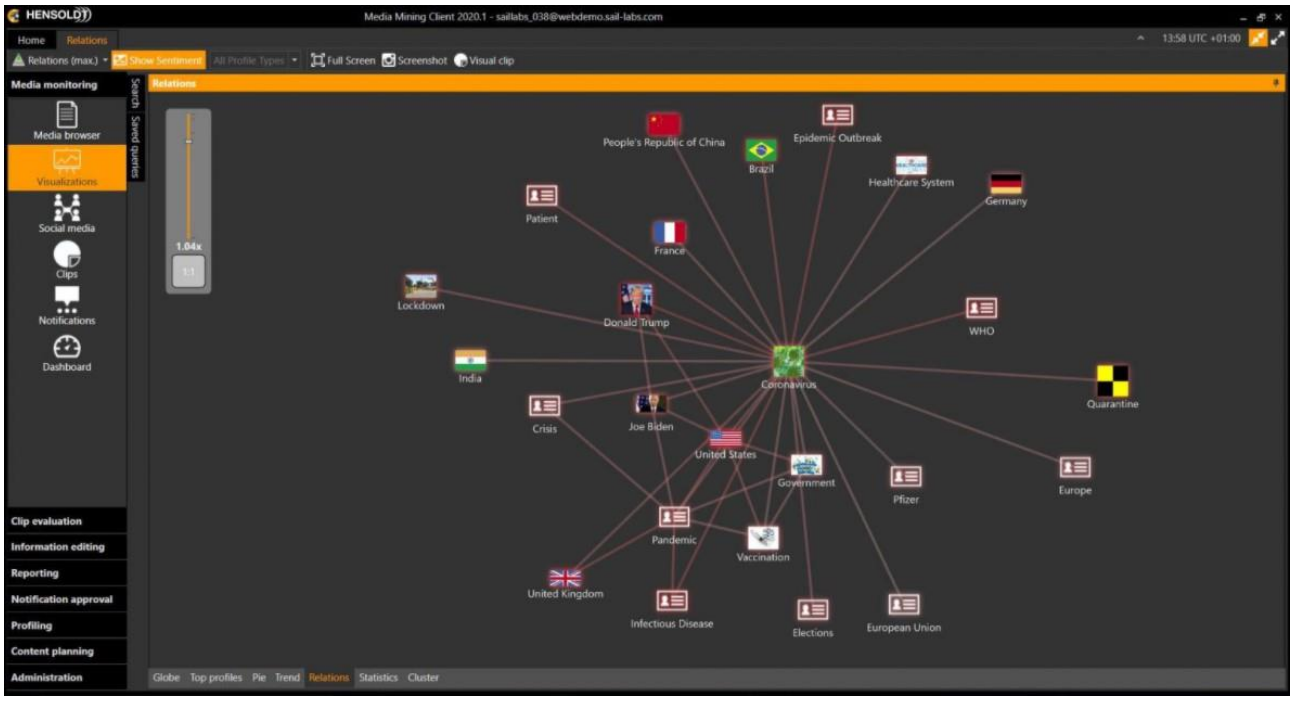


Figure 12: Hensold Analytics Link Analysis (https://www.hensoldt-analytics.com/, 2021).

Hensoldt Analytics was born with a focus on the defence and security sector. However, it is employed in various industries belonging to different industries. On the B2G side, the platform is used in national security and public health surveillance to protect the health of citizens; it is also used in the management of natural disasters and the protection of infrastructure and the formulation of communication strategies in foreign affairs[26]. In the commercial sector, the solutions are various: it is used in the travel industry to detect in advance security threats or adverse conditions, in political strategy to help in the monitoring of social and classic media, in the financial and pharmaceutical industry to keep up with the latest policies and regulations.

[26] Application Areas - HENSOLDT Analytics (hensoldt-analytics.com) https://www.hensoldt-analytics.com/application-areas/.

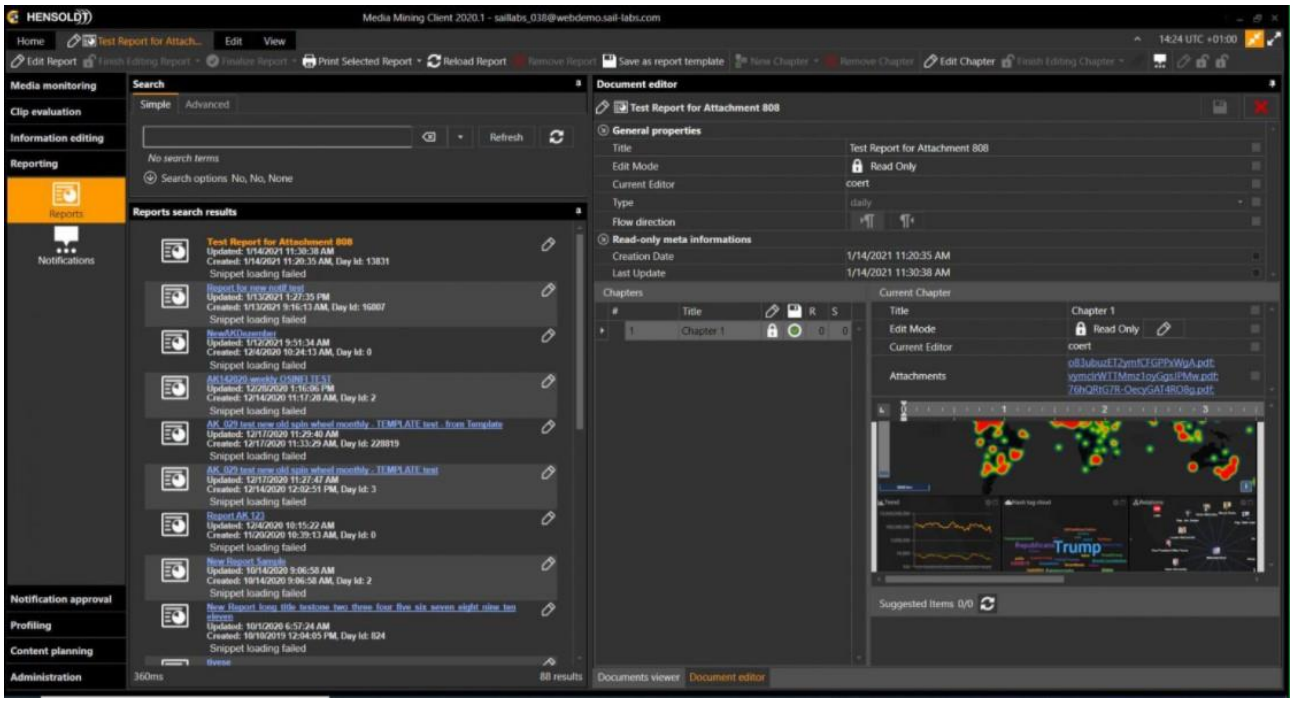


Figure 13: Hensold Analytics Link analysis dashboard (https://www.hensoldt-analytics.com/, 2021)

Risks

The risks for Hensoldt analytics are very similar to any other company in the business intelligence market (Marketline, 2021). First, one of the most significant risks comes from the high degree of innovation in the industry. In recent years, the global pandemic of the Covid-19 virus has prompted companies to invest heavily in technology infrastructure and new software development. Since there are many new developments, especially in the field of Big Data management and analysis, companies operating in this sector need to continuously innovate in order to stay one step ahead of their competitors. In this scenario, Hensold Analytics can't afford any missteps, or else they will lose many customers.

A second risk, similarly to Palantir, is due to cybersecurity risks. Since cyber-attacks successfully hit so many big names in the HiTech field, such as Microsoft and Facebook, no company can consider itself safe. Hensoldt Analytics system works very often with sensitive data and any data leakage could put its reputation at serious risk. In addition, a risk factor is the defection of its service providers, which, in some cases, could put its operation in serious jeopardy.

Finally, another issue is the increase of competitors within the market. In fact, more and more companies realise how much digital transformation is shaping the market and how important it is to enter it either by buying a company that operates in this area or by developing the system from scratch. In this scenario, the main risk is the possible price squeeze that could cause the company to lose market share.

Business Cases

As mentioned above, Hensoldt Analytics system, unlike older OSINT platforms that focused primarily on data in text form, can also integrate various types of information in different formats, including images, audio and video. All this is possible thanks to the Media Mining System (MM-System), a modular system that is able to cover the entire OSINT Intelligence cycle, from data collection to dissemination. It is composed of different tools that can also be installed separately according to the customer's needs. It is composed of four main components (G. Backfried, 2015):

Media Mining Feeders: represents the MM-System's input interface from the outside world. It is the component responsible for the ingestion of the various data and has the task of indexing the documents assigning them characteristics depending on their origin, the date on which it was inserted. Its characteristics also have the function of normalising the text and identifying the files.

Media Mining Indexer: this is the core component for audio and text analysis within the MM-System and is composed of different technologies that allow the analysis of various document formats. **Audio Processing** is the component that allows to decode and analyse audio files, the major strength of the Hensoldt Analytics platform. It consists of five main phases (E. Dikici, 2019): the first stage is called Audio pre-processing, where the audio/video input is extracted, converted and normalised into the native audio format. The second phase is Audio segmentation, where it is able to recognise and segment words, music, rhythms and non-language sounds. Then, in the Speaker identification phase (SID), the system can identify the speaker from a set of predefined targets; if it is not recognised, it can identify age, gender and similar patterns. In the Automatic speech recognition (ASR) step, the platform can transcribe the speech in 25 languages. Finally, in the Post Processing phase, the text is enriched and analysed by the system thanks to the Named Entity Detection (NED) to identify different morphological concepts.

Regarding this last point, it is worth mentioning the Continuous Vocabulary and Language Model Adaptation (CAVA) that use algorithms based on frequency, recency and variety of the various terms, can identify new trendy words and eliminate terms that are no longer used. **Text Processing**: as we have seen for the transcription of audio and video files, also for text files, the Named entity detection (NED) component is able to enrich documents with analysis of concepts and terms inserted in an entity category. The system is also able to perform sentiment analysis. La parte di **Visual Processing** consente invece di riconoscere volti e testo in un'immagine.

Media Mining Server: includes servers used for file storage and a set of tools and interfaces used to update and query database contents. MMS is based on Oracle 11g and Apache Lucene for search and

retrieval functions. The MMS also comprises semantic technologies forming the basis for all ontology-related operations, a geo-server and translation functionalities.

Media Mining Client (MMC): allows user interaction with the platform; this level allows to search, filter results, display analysis and build dashboards, as well as represent data in the form of a graph or map. Here are shown all the collected documents ready for analysis.

The Hensoldt Analytics system is applicable in various fields. A very interesting study concerns the use of sentiment analysis to analyse German social media during the 2013 German-Austrian flood (G. Shalunts, 2014). The model was developed by examining and comparing three sentiment analysis methods. The first one is a lexicon-based method with already defined vocabularies and patterns. The second one is an evolution of the first one, adding the specific domain of general news and natural disasters, improving the multilingual support, adding the possibility to manage data from traditional and social media, preparing the reading of long articles and adding a classification scheme in four classes. The third method is based on human analysis. The results establish a substantial equality between the second and third methods, with an agreement rate almost equal to more than 90%, more than double that of the first method. Finally, the sentiment was represented in a dashboard by making a temporal and quantitative analysis of the tweets examined.

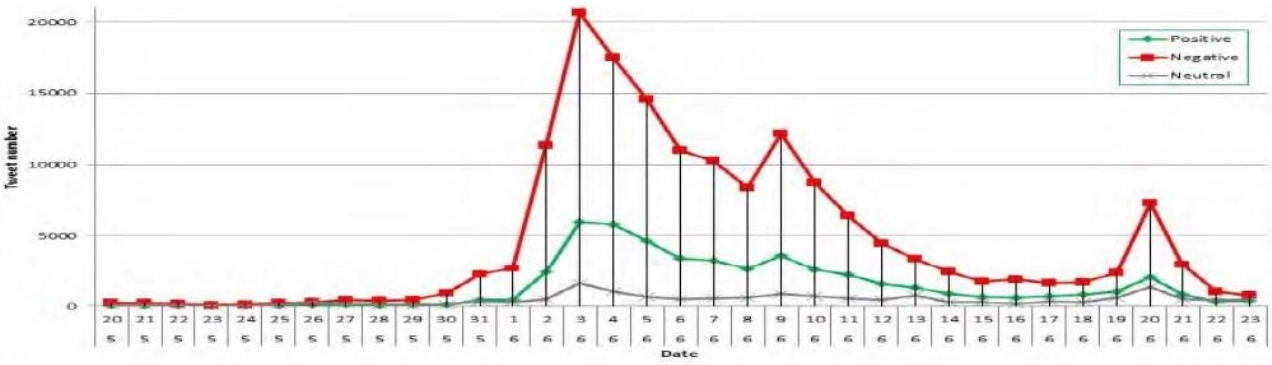


Figure 14: Results of Temporal Sentiment analysis of tweets in the period May 20th - June 23rd (Shalunts et al. 2014).

The use case of the Crimea crisis in 2014 shows that by analysing different sources of information ranging from social media to radio, it is possible to create several different reports that have different applications. For example, a report can be created for national security, analysing the crisis and understanding how the situation can be managed. A report can be created solely to understand how best to distribute humanitarian aid or formulate foreign relations strategies.

Maltego Technologies

Overview

Paterva is a software development company founded in Pretoria, South Africa, in 2007, famous for its open-source OSINT platform Maltego. The company's characteristic has always been to operate through a small team with the support of the community of more than half a million people worldwide. In the last few years, due to the industrial revolution 4.0, many companies have started to invest in business intelligence software; as a result, Paterva has seen a significant increase in enterprise customers and it has become challenging for a small team to take care of so many customers around the world. In addition, some founders left the company in search of new challenges. For these reasons the Company in 2018 moved customer operations and new solution development to Germany, under the leadership of the Maltego Technologies team.

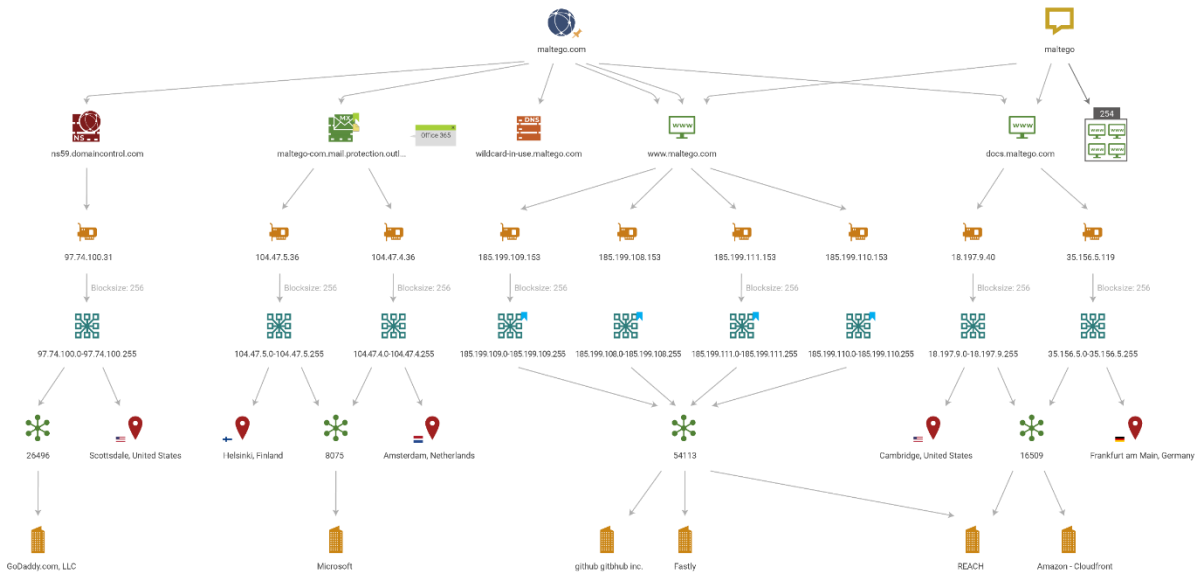


Figure 15: An example of Maltego Link Analysis (https://commons.wikimedia.org/wiki/File:Maltego-example-Graph-L_(1).png, 2021)

Ecofin Data

Maltego Technology has total assets of approximately $2 million. Through 2018, the company was rated sufficient (C) in the modeFinance Qualitative score[27].

Risks

The risks for the company are primarily competition. Although Maltego is one of the few affordable options to small and medium-sized businesses because of its competitive pricing, it suffers much competition at the enterprise level, where it competes with platforms with more features and flexibility. Should there be a general price drop in the market, Maltego Technologies could suffer. As for the other companies, the problems due to the rapid evolution of the sector and the IT security also arise, although being able to count on such a large community can help the Company keep up with the times.

Product

Like other OSINT platforms, Maltego can be used to collect various types of publicly available data. Used mainly in investigations, it can find links between many entities such as people, religious groups, companies, associations, network infrastructures (IP addresses and internet domains for example, thanks to compatibility with Shodan), affiliations, documents and many others. Thanks to Maltego, it is possible to represent in graph form a lot of data available online from different sources (some of which must be paid) and obtain the information you need in much less time than a classic search. Very important to understand the functioning and description of the platform is. The concept of Transform: Transforms are pieces of code that take a bit of information (in the form of an Entity) as input and then return related information in the form of more Entities as an output. A Transform runs on a Transform server hosted by either Maltego (for Maltego Standard Transforms) or by one of our Data Partners. Transforms are the mechanism that enables exploratory link analysis from within the Maltego Desktop Client application. They can also be written by Platform's users, providing the flexibility to connect to their own data.

[27] ORBIS, Data Update 209,001 (02/04/2021) © Bureau van Dijk 2021 Maltego Technologies GMBH Complete book.

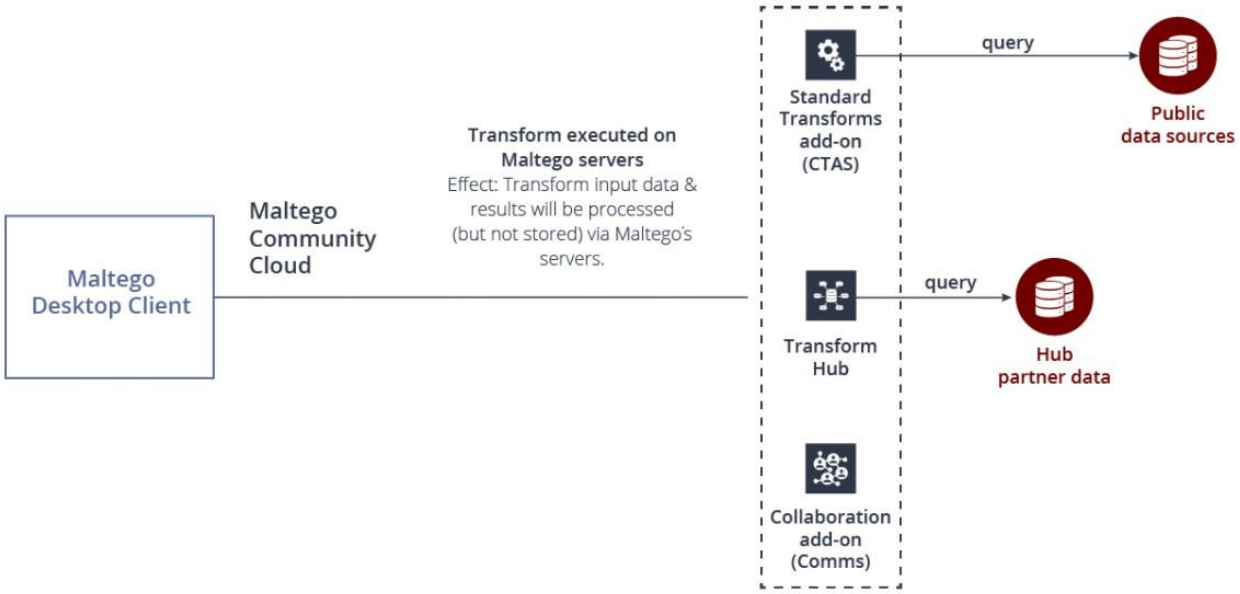


Figure 16: How Transforms works (www.maltego.com, 2021).

Maltego has two types of licenses: client and server. Maltego servers can be deployed within an organisation enabling it to host its Transform servers on infrastructure controlled by the organisation rather than having its Transforms run over Maltego's infrastructure. An internal server gives a firm the ability to integrate with its internal data and leverage internal processes and distribute these Transforms across the enterprise. Maltego client is the version that allows the user to interact with the GUI to perform data analysis. There are different versions designed for various types of clients; they differ mainly in the number of entities that a search can show in a single graph. Maltego One, for example, is the new professional license that incorporates the XL version for enterprises. Maltego Community Edition is the free version of Maltego that allows anyone to try out the program's potential. Finally, Maltego CaseFile is the version that allows the user to work with offline data. Maltego offers the possibility to integrate information sources with databases managed by third parties, which makes it very versatile[28].

Business Cases

Very interesting research concerns the use of Maltego in social cyber forensic by analysing open-source data and social media to advance cybersecurity informatics (S. Al khateeb, 2019). The study focuses on the 2015 Operation Trident Juncture, where the U.S. Army, along with more than 30 NATO countries, conducted exercises moving more than 36,000 soldiers to different parts of Europe to demonstrate adequate capabilities to deal with security issues. At that time, some opponents started

[28] Homepage - Maltego https://www.maltego.com/.

a disinformation campaign on blogs and major social media, inciting a revolt against NATO and Operation Trident Juncture. Starting from a set of 12 blogs, the researchers extracted, through Maltego, various types of metadata such as Web Tracker Code like Google Analytics ID that allows the owner of a site to view the statistics of visitors when they visit his webpage. The same ID can manage multiple sites. Through the analysis of this metadata, it was possible to discover a connection between these blogs and others that published the same news, sometimes in different languages. By analysing these blogs, the researchers collected a large amount of information such as IP addresses, name of the site owner, phone numbers and locations that allowed to obtain three clusters based on IP address geo-localisation, useful to know the origin of the blogs. A social media analysis was also conducted to find the most used hashtags, the most tweeted URLs and the most used domains to understand which other social media were most used. Finally, sentiment analysis was conducted on blog articles, as well as an analysis of keyword trends, publication frequency and bloggers. In this way, it was possible to have a precise identikit of the dissidents.

Maltego is also used to support User Behavior Analytics (UBA) tools (Maltego, 2021), software that can detect suspicious activity that could threaten a business. In particular, UBAs can tell when there is an inappropriate use of account privileges, when confidential data is stolen or when credentials are compromised. If any of these events occur, the system immediately notifies the possible threat, but the alerts need to be contextualised and this is where Maltego comes in. By integrating internal data with OSINT data retrieved from the platform, the UBA alert notification can be contextualised and a false positive avoided. Moreover, thanks to this analysis, it is possible to improve the algorithm of the UBA tool and improve the quality of the alerts.

Other Competitors

Cy4Gate's direct and indirect competitors include many other companies developing business intelligence platforms that can compete with Quipo. Among these, the Italian Company Almaviva, Voyager Labs and Tableau Software are worth mentioning, three platforms that are different from each other but whose goal is to make a company become data-driven.

AlmavivA Group is one of the largest Italian IT companies, a leader in driving Digital Transformation of public administration, with a turnover of about 890 million euros and 45000 employees worldwide. It collaborates with the largest tech companies in the world; among its customers there are many national and international public bodies such as the Italian Ministry of Defence and NATO, and it is also used by many companies in many countries. The company deals mainly with CRM and IT System Integration and operates in Finance, Defence, Cybersecurity,

Energy and many other markets. Almawave is part of AlmavivA Group and is a leader in the integration of customised IT systems according to customer needs, thanks to its many capabilities that include natural language processing, development of artificial intelligence algorithms, sales force management and the use of Distributed Ledger Technology (Blockchain) techniques.

Tableau Software, the homonymous business intelligence platform developer, is one of the key players in the market. His offer is distinguished by the numerous variants that adapt to the different needs of customers. There are, for example, a desktop version and a server version for larger companies and a version focused on CRM. The platform focuses mainly on the management and representation of structured data and makes user-friendliness its strong point. Another advantage is the vast community and the availability of several online tutorials that manage to give an overview of the platform in just a few lessons. Competition in Tableau's market niche is very high. In fact, numerous companies are betting heavily on the development of similar platforms. Among the key competitors are Microsoft with power BI and Elastic with Kibana.

Developed by UK-based Voyager Labs, Voyager Analytics is a business intelligence platform that can analyse OSINT documents collected from external sources. The platform uses artificial intelligence algorithms to process unstructured data such as text and images to find insights. As with other competing platforms, Voyager Labs' fields of application are multiple at both the government and corporate levels. The platform can be used for formulating strategic plans, cybersecurity operations, investigating terrorists and criminal organisations, and many other areas. Voyager labs is funded by technology companies such as Oracle, along with several UK edge funds.

Spiderfoot and Recorded future make Threat and security intelligence their core business. They are platforms focused on collecting and analysing OSINT data with the specific feature of automating the functions of Cyber Threat Intelligence, IT Asset Discovery, Security Assessments, and Attack Surface Monitoring. Shodan is one of the most famous search engines for finding information about IoT devices. The vast majority of the programs mentioned above use Shodan as a source of data for some of their functions.

Questel Orbit Intelligence, a software specialised in patent searches and the study of new technologies, allows you to consult dashboards with very important metrics for market analysis. The Crunchbase database is integrated into the platform and is useful for finding data on the companies analysed, particularly start-up. Zoominfo is one of the world's most popular CRM apps, used to optimise lead generation and manage the sales force within a company, providing information on sales and market performance. Marketing departments also use the app to facilitate go-to-market

strategy and by human resources to recruit and develop talent within the company. KPI 6[29] is an Italian start-up that aims to automate market research through a combination of data from its proprietary databases. The app provides a vast source of information that allows it to display a variety of information, from market trends to customer consumption habits. The platform consists of several tools, each with a specific analysis function. Khoros is a Business Intelligence platform specialising in customer engagement, marketing campaigns and customer care services. The app monitors all communication channels of the company and provides a clear understanding of customer behaviour. Due to the new customer-centric culture supported by the vast majority of companies, these apps are becoming increasingly popular.

Platform Comparison

The following is a comparison of competitor platforms with the platform developed by Cy4Gate, Quipo. The feature matrix and positioning map give a key insight into the capabilities of these applications and how they are perceived within the market.

Business Intelligence Platform Positioning Map

The two variables for the market segmentation have been selected by studying OSINT techniques and Business Intelligence literature. The dimensions relate to:

- **Focus on OSINT Analysis**, representing the degree to which the platform focuses on collecting and analysing external data available on the Internet.
- **Focus on TAXONOMY Customization** represents the degree to which the platform focuses on taxonomies customisation according to the application cases.

Analysing the results deriving from the positioning map (see fig. 17), four market niches were noticed corresponding to the quadrants outlined by the positioning map.

Specific In-House: includes platforms that have a rigid semantic structure and that refer to internal non open-source data. These software collect data provided by the developing company or customer databases. They are used for specific purposes within the organisation that cannot be extended to other business functions. Since these are programs developed to solve a specific problem, the taxonomy structure is very rigid, focusing only on their scope of application, such as CRM applications. The advantages of these systems come from their specialisation. As they are used for very specific tasks, they require less time to be implemented and benefit from more established

[29] Alberto Nasciuti ci parla di Kpi6: Dati e insights per la consumer market research (startup.info) https://startup.info/it/alberto-nasciuti-kpi6/

taxonomies in their field of application. In this section, there are well-known applications of CRM and patent analysis. Although they are very comprehensive software, they are highly specialised in their field of use, unlike Palantir, which is used for several purposes.

Multi-Purpose In-House: includes platforms whose primary purpose is to use in-house data, but unlike the previous group, they have the ability to modify taxonomies depending on their field of application. They are used for various purposes and are very adaptable to various markets and various business functions, which is precisely the reason why they need to adapt the taxonomy structure depending on the type of analysis required. The advantages of these software lie in their greater elasticity of use, making them attractive in many sectors and for the most various problems. While being able to analyse external documents, Palantir focuses more on the analysis of internal primary data, ensuring maximum flexibility of use. Almawave, instead, is a company that allows a highly customisable platform development and can adapt to many markets. Both platforms have a wide scope of use, which makes them easily customisable to the customer's needs.

Specific OSINT: includes platforms that focus on the collection and analysis of open-source data and have a limited ability to customise taxonomies for semantic analysis. Like the programs in the first quadrant, these software are also developed for very specific functions within organisations that require analysis of externally sourced data. The taxonomies are often fixed due to the high specificity of the task, even though platforms dedicated to OSINT research are very widely developed in their scope (e.g. this group includes cyber intelligence software). These platforms are able to manage the data they collect more quickly and accurately because of their knowledge of the primary sources from which they collect information and their highly developed semantic analysis around the primary function. We find platforms such as Maltego and Spiderfoot known for their specialisation in investigation and cyber threat intelligence in this quadrant. Over the years, these applications have been developing more and more, expanding their scope of use.

Multi-Purpose OSINT: In this niche, the Business Intelligence platforms make open-source data collection their core business and allow developers and designers to create new taxonomies for semantic analysis depending on the platform's field of application. The tools in this quadrant have various features that make them suitable for various tasks, and their flexibility makes them usable by multiple functions within the company. Although they focus on OSINT research, these software are, in most cases, capable of handling internal data to combine the benefits of external and internal research. Since the strength of these platforms is their adaptability to various environments, they require ad hoc taxonomic development to reflect the client's needs. While implementation can be slightly time-consuming, once embedded in business operations, it provides a high degree of

flexibility at all levels. In this section, Cy4Gate, Hensoldt and Voyager Labs are in very close competition with each other. Quipo, Hensoldt Analytics and Voyager Analytics focus on external analysis of unstructured documents using artificial intelligence and machine learning. While the three platforms have differences, they are on the same end of the market.

While aiming to help organisations collect, manage, and analyse data, BI platforms are very different from each other and can be used for different purposes. QUIPO, the BI Platform under study, is positioned within the Multi-Purpose OSINT industry market with its related key features to be considered for adoption.

Matrix

The construction of the matrix was designed to provide a benchmark of the salient technical features of the various platforms. As analysed in the positioning map, the software, while offering similar functions, sometimes differ in purpose and focus on different functions. Therefore, within the matrix, it was indicated with a green dot whether a platform has a certain function.

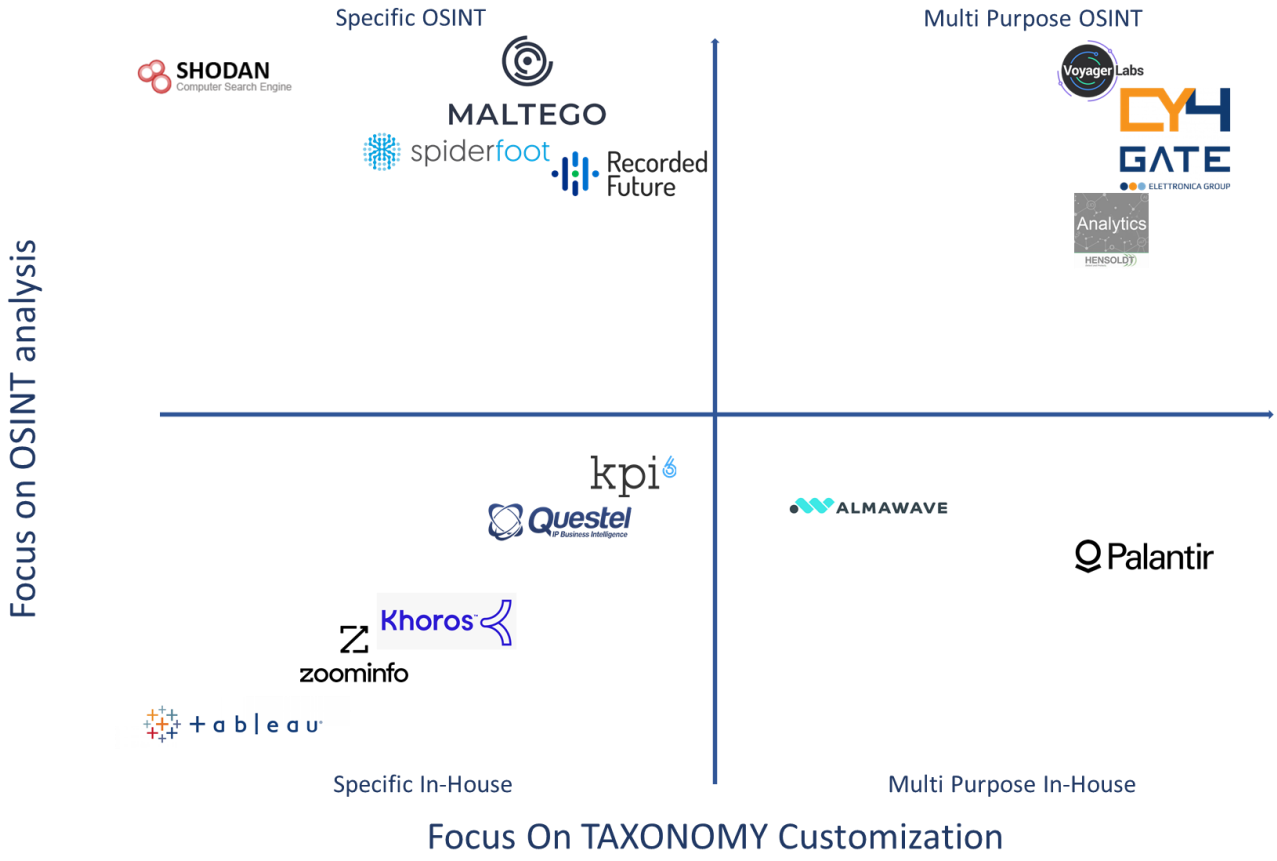


Figure 17: OSINT Competitors Positioning Map.

	Quipo	Palantir Gotham/ Foundry	Hensoldt Analytics	Maltego	Voyager Analytics	Almawave	Tableau	Spiderfoot	Shodan	Recorded Future	Questel Orbit	KPI 6	Zoominfo	Khoros
Ingestion and analysis of structured data	●	●	●	●	●	●	●	●	●		●	●	●	●
Ingestion and analysis of unstructured data	●	●	●	●	●	●		●		●	●	●	●	●
Audio and video processing	TBAS	●	●			●								
Image Analysis	●	●			●	●						●		
Social Analysis	●	●	●	●	●	●						●		●
Dashboard and report creation	●	●	●		●	●	●	●		●	●	●	●	●
Automatic and continuous ingestion	●	●	●		●			●		●				
Link and sentiment analysis	●	●	●	●						●	●			●
Natural Processing Language to analyse unstructured data	●	●	●		●	●					●	●		

Figure 18: Competitors Technical Benchmark. Note: TBAS is the acronym for To Be Added Soon.

CHAPTER III

Company profile: Cy4Gate

CY4GATE was established as a joint venture between Elettronica Group and Expert Systems in 2014 with the task of responding to a demand for unconventional cyber security. Listed on AIM (FTSE AIM Milano) since June 2020, CY4GATE was conceived to design, develop and manufacture technologies and products, systems and services, which are able to satisfy the most and modern requirements of "Cyber Intelligence & Cyber Security" expressed by the Armed Forces, Police Forces, Intelligence Agencies and Companies, both at home and abroad. A unique Italian industrial project, CY4GATE operates in the cyber market at 360°, with proprietary products that meet both information collection and analysis and security needs.

CY4GATE operates in two macro areas: 'Cyber Intelligence' and 'Cyber Security'.

Cyber Intelligence: the Company develops and licenses programmes for the collection and analysis of information present online or transmitted via the Internet (Open Source Intelligence or "OSINT") and the collection of information produced through the use of electronic and digital devices (Social Media Intelligence or "SOCMINT").

Cyber Security: the company develops cyber security products and services to protect customers' information systems, enabling anomaly detection and generating response actions. The Company offers cyber security products and services.

Figure 19: Cy4Gate History.

QUIPO: A Business Intelligence Platform

Quipo is Cy4Gate's proprietary Business Intelligence platform focused on supporting and optimising decision making. The software supports the analyst throughout the intelligence cycle by collecting, analysing, and representing OSINT data. Born in 2017 from the collaboration between Cy4Gate and Expert System, Quipo (formerly known as D-Sint) uses the integration of the Cogito semantic engine leading in semantic understanding and natural language processing by leveraging artificial intelligence and machine learning algorithms. The primary objective of Quipo is to support analysts in intelligence analysis by partially automating the various steps involved in developing a report on a given topic. In fact, with the exponential increase of data in recent years, it has become impossible to carry out intelligence analysis relying entirely on the human workforce. The advent of social networks such as Twitter or Facebook and the increase of newspapers and blogs means that more and more news and information are released every day, making it impossible, even for teams of many people, to acquire all the most important information. Business Intelligence platforms such as Quipo make this task easier by collecting and analysing information in real-time so that all the most important data can be filtered to gain the knowledge needed for analysis. Expert Systems' presence in the project, with its great experience in the field of artificial intelligence and information technology, has contributed to a substantial improvement of the platform, which now boasts one of the most reliable and famous semantic engines in the world. As mentioned earlier, Quipo supports the analyst throughout the intelligence cycle, allowing automation of certain functions throughout the process. In addition to these functions, it also allows the end-user to customise certain aspects of the software without having to resort to writing code and furthermore, thanks to the integration of the business intelligence and data visualisation software Kibana; it allows the data collected to be represented in dashboards that the user can fully customise according to his or her needs. The platform also allows users to take advantage of taxonomic constructs to create graphs that allow you to search for information both on individual nodes and on the branches that connect them.

QUIPO Structure

QUIPO's structure consists of three main modules linked together by tools that use artificial intelligence and machine learning algorithms. Their function is to collect, save, read, edit, aggregate and represent internal and external data from different sources. The collected documents are saved within the platform and uniquely identified. The data is then analysed by Cogito, which, depending on the taxonomies created, enriches the documents with aggregated and graphically represented metadata. Parallel to the text documents, the images are analysed by AI, which is able to recognise

faces or objects and tag them. Finally, all data can be consulted by the end-user for research purposes and disseminated through the AMICO platform.

QUIPO's core modules are:

Extract, Transform and Load (ETL): the first module consists of a non-relational database with the task of indexing documents from internal and external sources. An archive is then created containing all the documents, enriched with the first metadata that establishes their uniqueness and allows the internal search engine to retrieve them quickly. After being saved and indexed in the database, the data is analysed by the platform's semantic engine or, in the case of images, by image and face recognition tools. In this phase, the documents are assigned additional metadata, which is very important for the representation and analysis by the user.

Heterogenous Data Analytics (HDA): the second module deals with the representation of the documents collected in the database using the metadata produced in the previous phases. Through the aggregation of this data, the module allows the creation of dashboards and graphs which contain metadata derived from image analysis and semantic analysis. Through the taxonomic concepts, it is, in fact, possible to create various graphs that help to have an overview of what Cogito was able to extrapolate from the collected documents.

Operation Control Platform (OCP): the third module is the platform's GUI, which allows the end-user to manage all the tasks necessary to carry out the analysis. It can be considered as the point of contact between the backend tools and the analyst. Closely linked to this module is the data dissemination platform that allows the textual and graphical display of the work done. This part, entirely developed by Cy4Gate, is essential for interacting with the platform and indicates its degree of usability. For a business intelligence platform, it is indeed essential to be easy to use in order to have a high acceptance level by end-users.

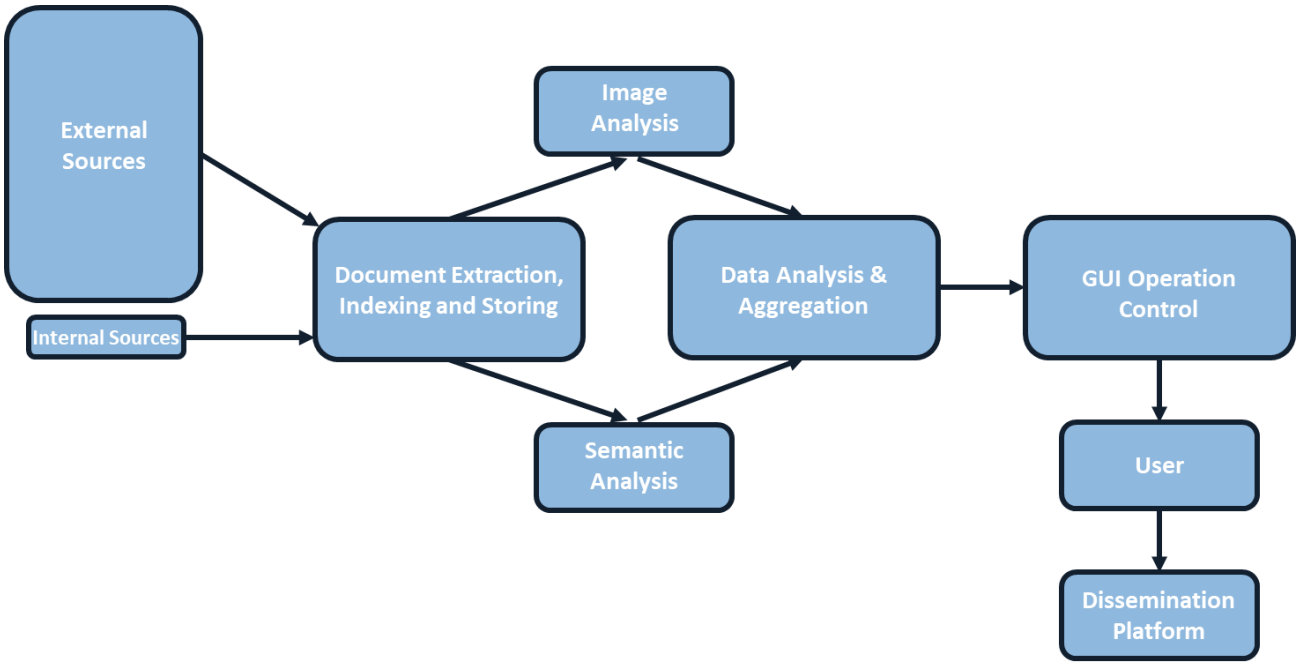


Figure 20: Visual representation of Quipo Structure.

Quipo Graphical User Interface (Operation Control Platform)

QUIPO is an advanced intelligence analytics platform based on a modular, scalable and open architecture to support the Customer to gather and analyse in real-time a large amount of data collected from multiple sources and turn them into intelligence information on targets. QUIPO is an intelligence platform conceived for law enforcement agencies, intelligence agencies and military organisations.

QUIPO is able to:

- Enact the Intelligence Cycle to produce knowledge;
- Manage and analyse information from any kind of source;
- Provide data enrichment on all information obtained from the sources, using the power of proprietary semantic technology;
- Integrate collaboration features to allow team definition and task management;
- Disseminate the information to the decision-makers.

The QUIPO solution design has been conceived to pursue the full intelligence cycle (Planning, Collection, Processing, Analysis, and Dissemination) to process structured and non-structured data (multi-media, multi-format), which are collected from:

- Open sources (web, social networks, emails, databases, etc.)
- Local databases (emails, warehouse documentation, and proprietary data)

Main Page

The landing page shows all the cases accessible by the user. From this page, it is also possible to monitor the platform's progress and create new users, cases, dashboards, and charts. Some functions are reserved for users with supervisor status.

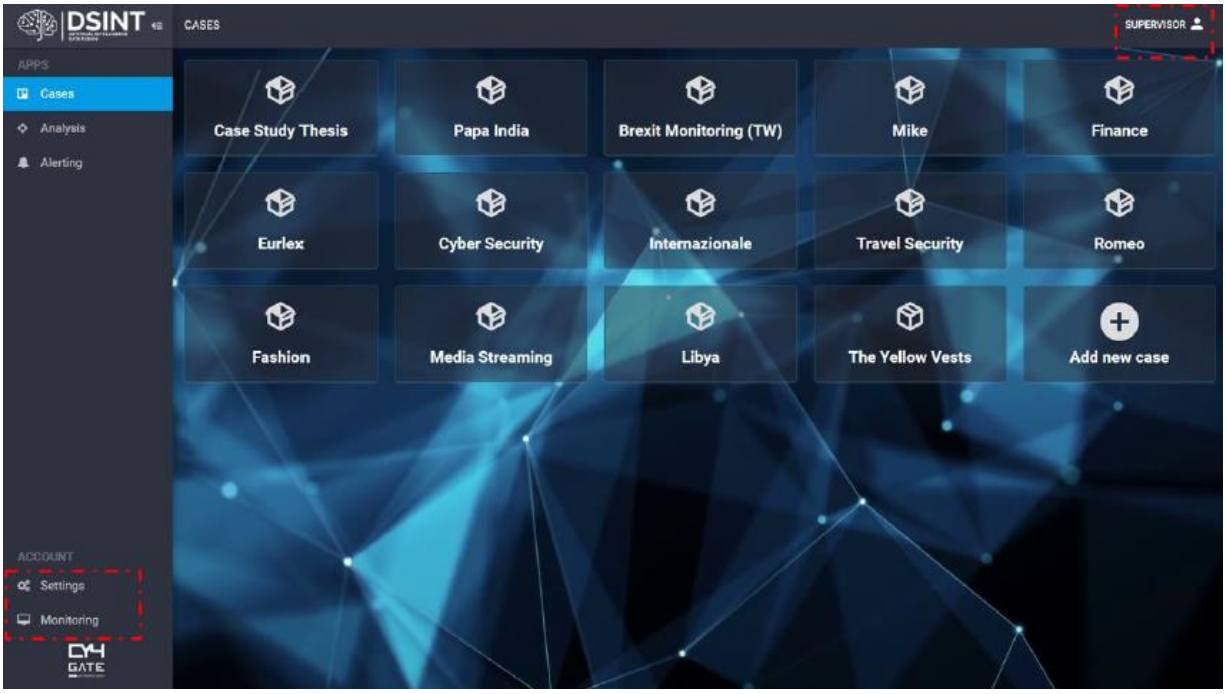


Figure 21: Quipo landing page.

The Intelligence Cycle

Using QUIPO Platform, the user can go through five different stages that match the Business Intelligence Cycle:

Figure 22: Quipo Intelligence Cycle.

1. **Planning** (top right): the Planning phase allows you to create tasks and targets to be assigned to one or more analysts and to start collecting data from OSINT (News) and Social Media (Social).
2. **Data Gathering** (middle right): you can select specific sources from which to retrieve data and information; the source types are the ones as follows:
    - RSS (Really Simple Syndication): RSS Sources allow you to update the data input from websites on the QUIPO automatically. RSS is often used to crawl data from online newspapers.
    - Google News: it allows the user to perform queries using the Google News methodology.
    - EMM allows the user to perform queries using the Europe Media Monitor methodology.
    - Twitter: it allows the user to collect tweets from Twitter accounts, pages and hashtag.
    - Web Crawler: allows the user to add a single specific URL address in the Data Gathering process. Furthermore, it implies a manual addition of sources into the platform.
    - Facebook allows the user to collect information from public Facebook pages.
    - Instagram allows the user to collect information from public Instagram accounts, pages and hashtag.
    - Search Engine allows the user to retrieve documents from the web through a keyword-based search on the search engines Duck Duck Go and/or Qwant.
    - Wikipedia allows the user to collect documents from the online encyclopedia Wikipedia.
    - Local Files allows the user to upload the files the user are interested in, such as images, html pages or PFD files.
    - Email module allows the user connecting to a specific mailbox, if the user know email address, password and collecting data from it.
3. **Analysis** (down right): the cockpit of analysis merges different modules in relation to the different sources of information you selected in the previous phase:
    - The module of Special Analysis allows you to implement dashboarding with the creation of many plots, thanks to Kibana integration and social media trend analysis.
    - The Integrated Module is also called Explore and it merges all the different modules to boost the analysis activity. The Explore module is the central cockpit of analysis: it opens with a specific interface which has its core in the semantic analysis

(unstructured data) and has the main duty to show you all the data collected together with the analysis modules. It also includes a dedicated section for image analysis (face recognition, image tagging, objects, or physical concepts recognition). In this section, you have the possibility of filtering documents thanks to taxonomies, a hierarchical classification of concepts that can be completely customised according to the client's needs and that allows you to speed up the search for important documents for analysis. Taxonomies are used by the Cogito semantic engine for document understanding and analysis. There are also secondary filters that allow the range of documents to be further narrowed down to facilitate the analyst in their tasks. You can also run a link analysis to show the relations between the entities inside the documents.

4. **Production** (down left): Once the analysis phase is concluded, all relevant documents have been collected, assigned to a task and saved in the notebook, you are ready to start typing your personal report.
5. **Dissemination** (middle left): you can disseminate the report to decision-makers, redirecting him to the AMICO dissemination platform.

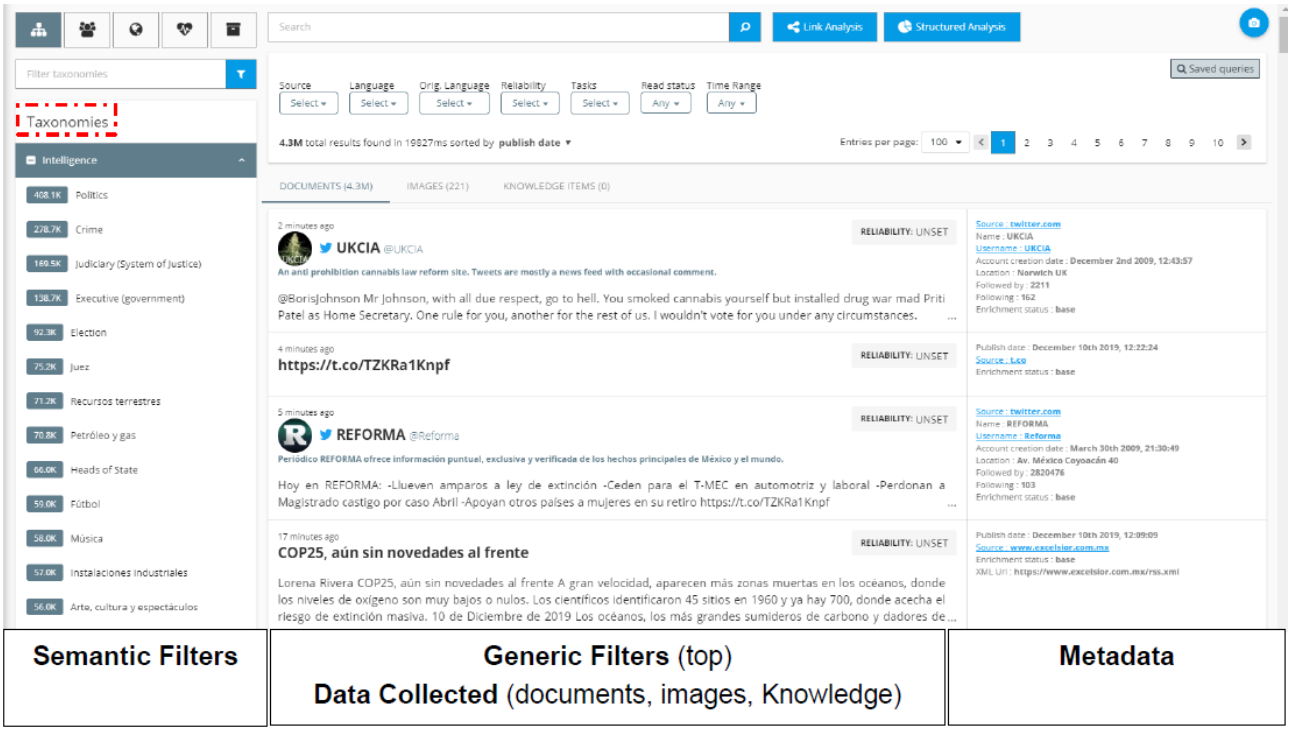


Figure 23: Quipo Explore Section.

In the inner circle there are five useful functions to customise the platform even more:

1. With the **automatic alerting** function, you can automatically insert articles, appropriately filtered with collection rules, inside a task and make them available for the production and dissemination part.
2. Through the **reliability** function, it is possible to automatically set the reliability of a data source to facilitate the identification of important documents.
3. With the **archives** function, you can create spaces containing the data lake of the collected articles.
4. With the **Knowledge Base** function, you can manually create a graph to highlight the most important relationships within the analysis.
5. With the **normalisation** function, it is possible to create normalisation rules in order to better manage the results expected from the taxonomies. In fact, it may happen, for example, that a company is reported in the articles with different names or abbreviations; through the normalisation function, all interpretations can be united under a single name.

AMICO (Advanced Multimedia Information Cockpit) is a data dissemination platform that provides, in an easy-to-use interface, the following features for operational centres and decision-makers:

- Management of analysis' results based on targets of interest (reports, knowledge base);
- Presentation of relevant documents and their semantic metadata;
- Presentation of structured or unstructured data merged together.

AMICO Platform is structured as follows:

- Live News: This section shows customisable live news under the form of Streaming channel (swipe right to open) and RSS stream news (click to open). Both channel and RSS can be customised in Settings;
- Analysis Interface: This section allows you to interact with the platform according to specific targets, Tasks and filters (such as: Arguments, Geo, Sentiments, Relations);
- Documents/Reports/Knowledge Base.

AMICO was developed to disseminate content to management. Through the platform, it is possible to view the report in the centre, various graphs and maps, while on the sides it is possible to check the news inserted in the task linked to the analysis, having a complete picture of the work carried out and being able to see in a few minutes the information that emerged from the research carried out. The advantage of having a software dedicated entirely to the dissemination of content is its flexibility. In fact, the platform is multi-device and has a very simple interface that can be used by anyone in a

very short time and in any part of the world. For example, travelling sales staff can request an analysis of an issue and have it available a short time later thanks to the work of analysts at headquarters. AMICO reflects the benefits of the cloud in Digital Transformation, facilitating employees' work at any time and in any place.

Thanks to the use of a VPN, QUIPO can also be used anywhere, making it possible to work on the move or in smart working mode. In light of what has happened in recent years, this is a key aspect that should not be overlooked.

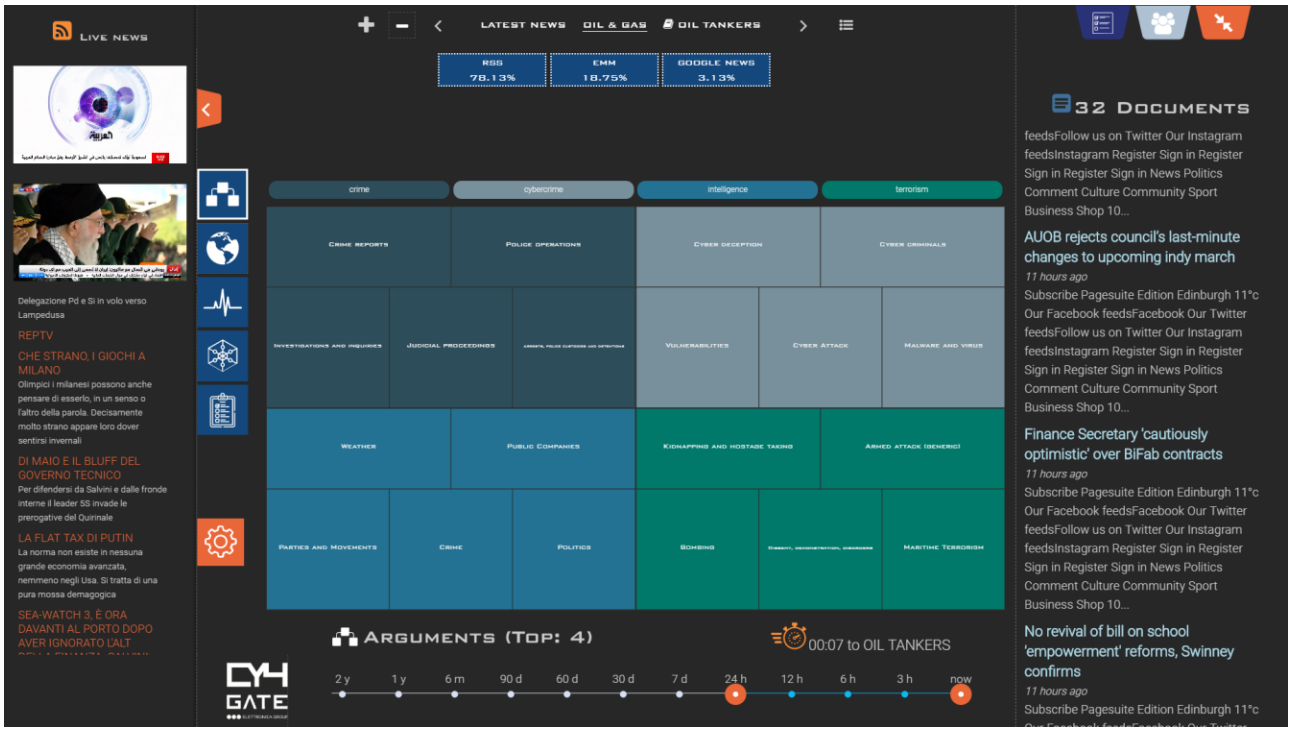


Figure 24: Amico Dashboard.

QUIPO SWOT Analysis

S

The strengths of the platform are closely tied to those of the developer company. Since its listing in 2020, Cy4Gate has embarked on a very ambitious multi-year development plan. The company is receiving numerous investments and developing new contracts with government and corporate entities around the world. In this scenario, the company is betting a lot on its products, investing time and resources in their development. Certainly, one of the products that benefit the most is QUIPO, which can also count on the know-how of Expert Systems, a leading company in the artificial intelligence sector, especially in the field of natural language processing. Thanks to this collaboration, QUIPO boasts one of the best semantic engines in the world, making the platform very attractive in

the marketplace. In addition, like the other companies in the Elettronica Group, Cy4Gate focuses on product excellence and high customization, which, combined with R&D investments, allow the company to meet the needs of its customers.

W

The company is currently going from the start-up phase to the growth phase. This timeframe is very delicate because it could lead to numerous problems if the strategy is not clearly defined and sustainable. For example, if the company is not supported by an adequate workforce, it may fall behind on contracts or miss out on opportunities to increase turnover. The company must make the right choices during this period to accelerate its growth and establish itself in an already crowded market.

O

According to Statista's (2020)[30][31] Digital Economy report, the global pandemic has increased the need for investment in various areas related to Digital Transformation. Cybersecurity is definitely the sector that benefits the most, with a growth in spending of 84%, followed by hybrid cloud infrastructure, automation, and AI with 74%, 66%, and 59% growth, respectively. It is undeniable that COVID-19 has accelerated global digital growth by raising awareness of how fragile the world system can be and that it is imperative to invest in new technologies to support organizations and their employees. Cy4Gate is at a delicate stage during this particular period. If it can make the right moves, it will be able to gain a lot of customers and investors globally and domestically. The opportunities in Italy and Europe in light of the Recovery Fund are driven by implementing digital progress programs at both government and corporate levels. In this scenario, QUIPO would be a game-changer, enabling companies to innovate their analytics and decision-making processes.

T

The threats facing the platform are technical, competitive, and political. First, the risks associated with cyber-attacks are increasing. For a company whose core business is cybersecurity, together with

[30] Global artificial intelligence market revenues 2024 | Statista (Link).

[31] Digital Economy Compass 2020 | Statista https://www.statista.com/study/83121/digital-economy-compass/.

the provision of digital services, a possible attack would represent a series of contractual and, above all, image-related problems. The issue is aggravated by reference to the third-party services that Cy4Gate and Quipo use. Even if in a minor way, a cyber attack on the company's suppliers could be a big problem. Although QUIPO carves out a very specific market niche, it is embedded within a larger context where indirect competitors can also threaten the platform. In fact, as mentioned in the previous paragraphs, even though QUIPO has different functions than some competitors, they can still be considered as substitute assets for some specific needs. In recent years, more companies have been entering the digital marketplace by developing SaaS solutions. There are no high barriers to entry and it is very easy to reorganize the company in case of inadequate performance. A portion of Cy4Gate's revenue comes from relationships with government agencies whose business is regulated by numerous laws and policies. This over-regulation could present problems in moving contracts forward and accelerating business.

CHAPTER IV

Company profile: Elettronica S.p.A.

The Elettronica Group is composed of three companies: Elettronica, headquartered in Rome, specialised in electronic defence; ELT GmbH, headquartered in Germany, 100% owned, created to support some European programs and now become a European centre of excellence in Homeland Security and, finally, CY4GATE, born from a joint venture with Expert System to respond to new cyber threats both in the military and civil sectors. It is also present with several commercial and representative offices in the countries that constitute the major reference markets (GCC, Singapore, Brussels, India). With a turnover of about 200M€ and an order backlog close to 1 billion Euro, the Group is present with its proprietary high-tech systems in the Armed Forces of more than 30 countries in 5 continents. It also participates in the main European programmes such as Eurofighter, NH-90, Horizon and FREMM, thanks to a strategy of cooperation and partnership with the main international industries in the sector. The group's parent company is Elettronica, a high-technology company that for almost 70 years has been a leader in the design, development and supply of strategic surveillance, defence and electronic countermeasures systems for naval, air and land use. The company is today controlled by the Benigni family and Leonardo.

Figure 25: Elettronica Group distribution.

The defence systems designed and produced by Elettronica are mainly aimed at defending platforms (aircraft, ships, land vehicles and now also cyber environments) from threats that may cause their destruction (e.g., missiles guided by radar or other technologies whose task is interception). In essence, the Elettronica systems, in order to guarantee effective protection, must first identify the threat and, based on the knowledge of the same, implement countermeasures aimed at confusing the guidance of the threat itself (e.g. 'blinding' the radars or making them believe that the objects to be protected are in different positions from those in which they actually are). In addition to products, Elettronica is also moving into services, in particular with the Electronic Warfare Academy: a modern and advanced electronic defence education and training centre for members of the Armed Forces worldwide.

Elettronica S.P.A. was founded in 1951 as a manufacturing company for industrial transformers. Immediately after its creation, important partnerships enabled it to acquire very important contracts, thanks to which it focused on precision engineering systems, acquiring significant design skills. During this period, it developed what is now its core business, the construction of electronic defence systems, thanks to which it is now one of the leading companies in Europe, operating worldwide.

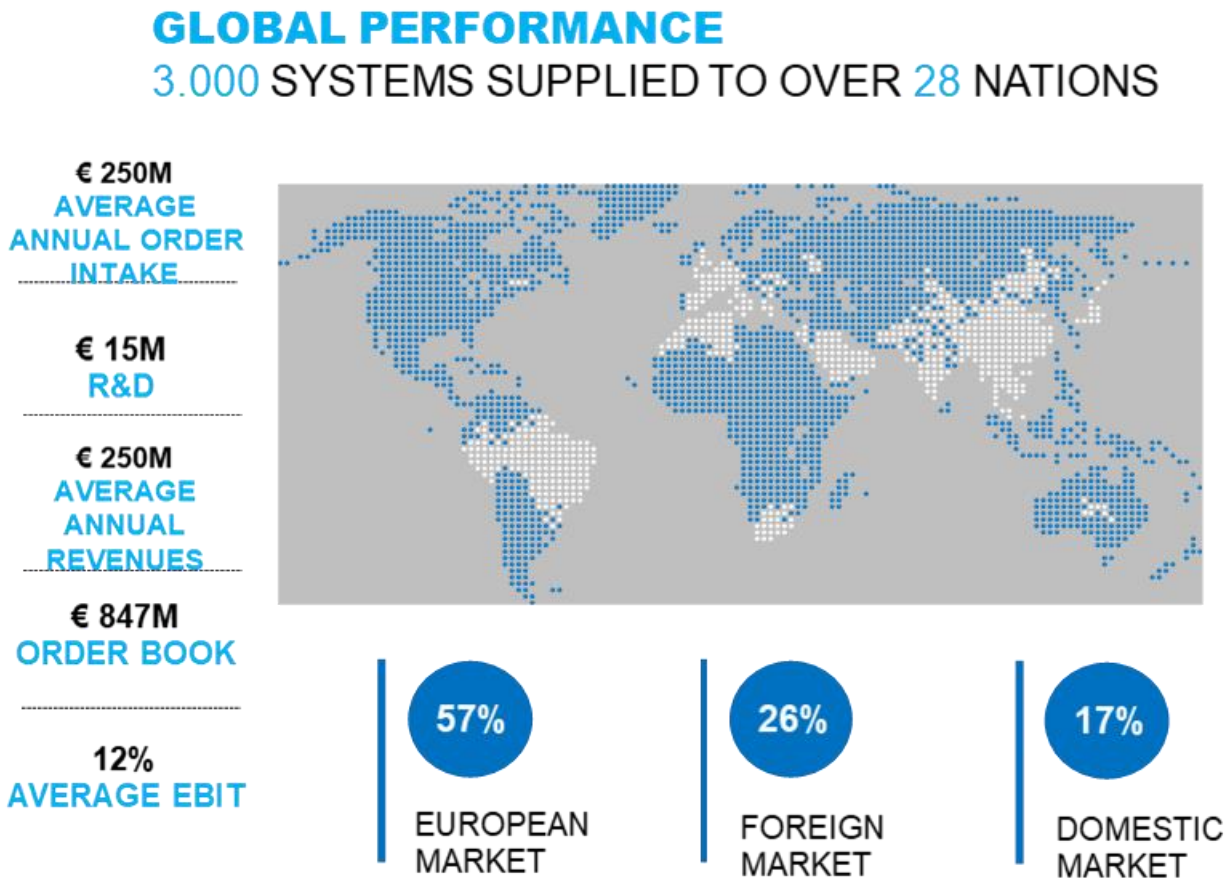


Figure 26: Elettronica Group profile.

QUIPO inside Elettronica S.p.A.

The QUIPO development project is one of Elettronica Group's main objectives for the company's Digital Transformation. In particular, Elettronica S.p.A wants to renew itself through the use of new software and technologies, modernizing its internal structure and operations. Since a few years ago, the company has realized the benefits that the digital revolution is bringing to business and, thanks to its presence in a highly innovative sector such as the defence one, has concluded and is carrying out several projects intending to become a fully digital enterprise. This vision is part of the Aerospace and Defence sector that has driven and continues to drive global innovation, being the starting point for the development and deployment of new technologies for military and civil markets.

In this scenario, Elettronica has a twofold objective: on the one hand, it wants to help Cy4Gate develop its flagship product within a sector that more than others can fully exploit the platform's potential. On the other hand, it has the purpose of inserting a Business Intelligence platform within the business context by expanding its range of use from the Strategy function to all core business functions. For this reason, there are currently three main projects underway that involve the development of QUIPO, all of which are overseen by the Strategy department.

- In collaboration with a consulting firm, the first project called Digital Funnel aims to educate the platform through the development of a modus operandi that allows the Strategy and Sales departments to analyze the market in search of new opportunities for Elettronica.
- The second project deals to help the Procurement department of Elettronica in the decisional choices regarding the business suppliers. In particular, the objectives of the program can be summarized in three points:
    1. Improvement of the effectiveness of the process of scouting suppliers starting from the product or capability
    2. Simplification of the "due diligence" of the key suppliers (actual or potential)
    3. Technological development trend of specific product categories, both to identify future advantages for Elettronica systems and to identify leading-edge or potentially out-of-market suppliers
- The third project is carried out exclusively by the Strategy department within the CIC (Corporate Intelligence Center). This program aims to jointly develop QUIPO functionality with Cy4Gate for inclusion in Elettronica S.p.A.'s operations.

This research is part of the second project mentioned above but also contributed significantly to the third one, which aims to monitor the evolutionary functions of QUIPO and is the basis for all other

projects that have to do with the development of the platform. The project's goal is to create several business cases with a twofold objective: on the one hand, it wants to start exploiting the potential of the platform in the study of the market and the geopolitical situation worldwide. On the other hand, it wants to improve the functioning of QUIPO, solving problems and establishing functions that will have to be implemented in the system for better integration within the company. Two teams working on this project are Elettronica and Cy4Gate. Elettronica is responsible for developing the case throughout the intelligence cycle, while the Cy4Gate team provides technical support and is responsible for resolving and implementing requests that come in from the first team.

QUIPO Procurement Project

In today's highly innovative defence market, it is essential to optimise business processes to implement and run them in the shortest possible time.

This research is part of the procurement department's project to improve technology scouting. In fact, although contributing to the development of the other two programs of Elettronica, the biggest work has been carried out in the context of a precise business case started at the end of January and continued throughout the period of internship at the company. The use case was designed to be the starting point for the implementation, development, and demonstration of QUIPO benefits for supplier scouting. Three companies were analysed in order to find answers to the following questions:

- In which geopolitical scenario do they operate?
- What are their capabilities? And to what extent do they contribute to their revenues?
- Which are the new technologies in the field of electronic warfare in which they are investing?
- How much are they investing in research and development?
- What is their heritage?
- Which are the other defence companies they collaborate with?
- What is their relationship with their government?
- What does an analysis of their financial data show?

This analysis aims to provide an overview of the companies analysed to identify whether they have the products and capabilities of interest and subsequently whether they are suitable to become suppliers. In the latter part, the company's international relationships over time are assessed along with its financial strength.

The project to improve technical scouting using QUIPO aims to partially automate the collection and analysis of data on potential suppliers, assess their stability, and find new technologies of interest

compatible with Elettronica products. This will speed up research at the beginning of supplier scouting, allowing employees to spend more time on more important steps. On the one hand, this has a double benefit, saving time and resources that could be used for other tasks, and on the other hand, the opportunity to discover innovative products or companies that otherwise would not have been considered.

Technology Scouting has the objective of developing Elettronica's portfolio in order to enable the company to gain a competitive advantage by introducing new capabilities, new products and new technologies. The main entity to coordinate and carry out technology scouting is the Technological Coordination Board. QUIPO would fit into the existing scouting process described below.

The Technological Coordination Board is a cross-functional body whose mission is to analyse in a structured and coordinated manner the capabilities and innovative technologies identified within the Long Term Plan (LTP) or from other strategic corporate inputs (e.g. Strategic Agenda, Technology Plan). The task of the Technology Coordination Board is to streamline the Technology Scouting process and make it effective and flexible, to enable the identification of key technologies to be submitted to the company's strategic Make or Buy process. In order to achieve this result, each actor carries out specific activities for each phase of the process.

The actors participating in the Technological Coordination Board refer to the following corporate functions:

Research & Innovation: The technological scouting activity carried out by the Research & Innovation area has its scope of action in the search for new technologies/capabilities/products within the entire sector industry as well as within specialised Universities and Research Centres (Industry & University driven).

Capability Marketing & Scientific Boards: The technology scouting activity carried out by the Capability Marketing & Scientific Boards area has its scope of action for the search of new technologies/capabilities/products in the collection and analysis of the needs expressed by ELT customers (including potential ones) as well as by the sector working tables (Client driven).

Sourcing & Supply Management (S&SM): The technological scouting activity carried out by the Sourcing & Supply Management Function has its scope of action in the search for new technologies/capabilities/products in the collection and analysis of technological innovation trends exclusively linked to the supply market (Suppliers driven).

Corporate Strategy: It has the task of developing business cases on the economic feasibility of the application of the identified technologies/ideas as well as identifying possible sales markets for the

ELT product. It defines the strategic value of the technology by defining the role of Elettronica in the reference market (push vs pull strategy).

Company Management: Not directly involved in technology scouting, it can provide guidance on technological lines of technological development even in the absence of a short-term sales market period.

The supplier scouting stages are:

Stage 1: the Functions/Areas responsible for technology scouting propose to the Technological Coordination Board the analysis of one or more innovative ideas activated on the market. The Corporate Strategy area has the task of identifying the potential market for the proposed innovative ideas.

Stage 2: starts with the identification by the Research & Innovation area of the enabling technologies for the ideas proposed by the players involved in the technology scouting activity. Once the ideas have been translated into enabling technologies, the S&SM department carries out market scouting to define the availability, timing and costs of the technologies identified. The S&SM Function is responsible for identifying the external technological roadmap. Following the identification of technologies on the market, the Research & Innovation area has the task of carrying out the technical analyses (time, costs, quality, reliability and reliability and performance) in order to identify the ELT products for which the technologies are intended. The output of the Technology Scouting process are new capabilities, new technologies, new products that must subsequently be submitted to the Make or Buy Analysis.

Stage 3: If the decision is to buy a certain technology, the S&SM department is responsible for searching for a suitable supplier. The search is carried out by referring to the Elettronica supplier pool and, if not present, through a market study. The best companies found are then contacted and subjected to a suitability process for inclusion in the supplier pool.

The research project is part of the second and third steps. In particular, it aims to develop QUIPO to help sourcing and supply management (also called procurement) and corporate strategy analysts to study the market for possible suppliers. This is the beginning of a larger project aimed at the construction and iterative development of various taxonomies for the general and specific analysis of products, capabilities and technologies. The research and innovation section for the study of new technologies is also part of the project.

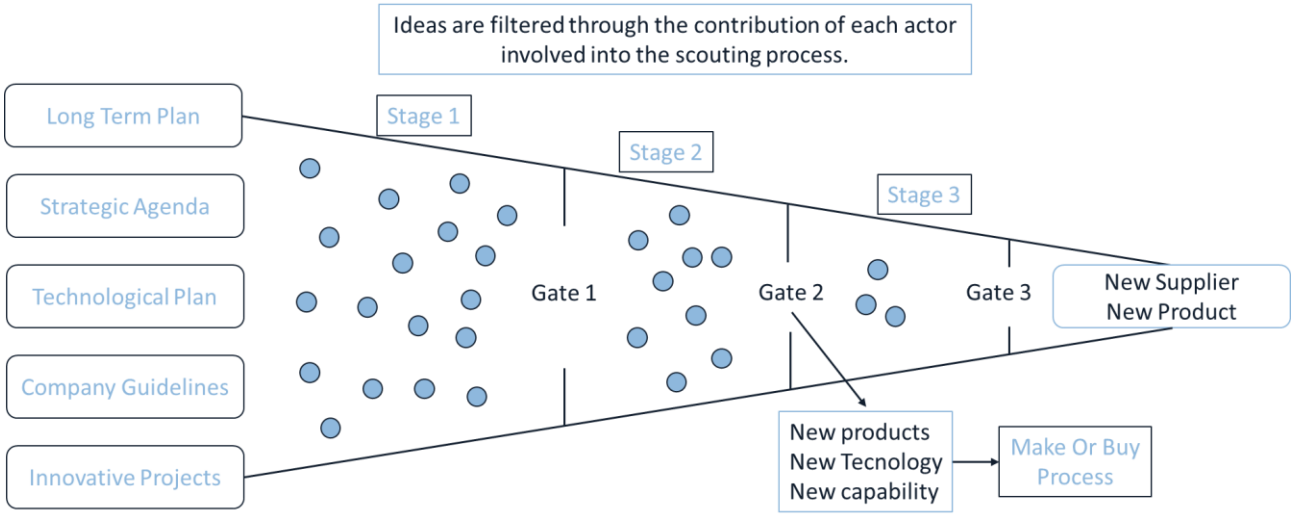


Figure 27: Technological and supplier scouting process.

QUIPO Evolutionary Functions

This research was developed taking into account all the projects described above; however, the evolutionary functions (EF) of QUIPO can be considered as the main project for the development of its usability. The purpose of the business cases carried out on EF was to use QUIPO to analyze competitors and partner firms of Elettronica. Specifically, a multitude of assessments was performed, including: a geopolitical analysis of the country where the companies are headquartered, the industries in which they operate, their capabilities and heritage, their relationship with European companies, their activity in terms of contracts, mergers and acquisitions, joint ventures and partnerships and their relationship with the government of their country. These analyses have been the starting point to reconstruct the business model of these companies together with their modus operandi and their presence in the world, in order to understand the market opportunities and, more precisely more easily:

- The off-limits areas for Elettronica where it is impossible to counter the actions of the three companies.
- Areas that represent an opportunity for Elettronica where the three companies do not sell products in common or where they cannot operate for geopolitical reasons.
- Areas where Elettronica can compete with the three companies.
- The areas where Elettronica can collaborate with the three companies.

The rationale of this project is to develop the platform iteratively, correcting problems that occur and identifying useful functions for its improvement. In addition, the following cases are fundamental for the strategic analysis and integration of QUIPO within other company functions.

Quipo Intelligence Cycle for Elettronica Business Case

The project will be described in all its points in the following paragraphs. It will be given an operational overview of the concepts expressed in the previous chapter by describing in detail all the processes put in place for the development and integration of the platform. Therefore, it will be started from an empirical description of the intelligence cycle, reporting all the actions that have carried to the development of the case. For each part of the cycle, the parallel tasks necessary to make the analysis comprehensible to all stakeholders inside and outside the company will also be explained. Finally, after each paragraph, there will be a small part dedicated to the contribution given in the development of the platform and the functions required for the future, will also be described the possible benefits of these integrations within QUIPO.

Note that the analysed companies are not disclosed for confidentiality.

Taxonomy creation for semantic analysis

A taxonomy is a semantic hierarchy that organises concepts by is-a relations, which exhibits the capability to improve many NLP (Natural Language Processing) tasks (C. Wang, 2017). Its construction is one of the most important operations in the implementation of software capable of semantic analysis. The taxonomies are, in fact, the component that allows the platform to understand what is analysed; otherwise, it would just be a document collector. The moment of their development is, therefore, the most delicate for the success of the analysis. A taxonomy is composed of rules that allow the semantic engine to understand a given taxonomic entity within a document. Their development depends on the domain that has to be analysed and the user requirement, who may have different needs according to his objective[32].

A taxonomy can be created using two types of approach:

- **Categorisation:** a deductive method in which the starting point is a general concept to arrive at specific individual entities. In this approach, a semantic tree is constructed with branches comprising subcategories of more general entities. Categories are generated first by

[32] Part of the following paragraph was written with the help of Cy4Gate engineers and Cogito guides.

formulating their conceptual descriptions and then classifying objects according to the descriptions.

- **Extraction:** The inductive method consists of identifying, aggregating, and normalising specific semantic entities to arrive at more generic concepts.

For both modes, there is a precise cycle that must be followed:

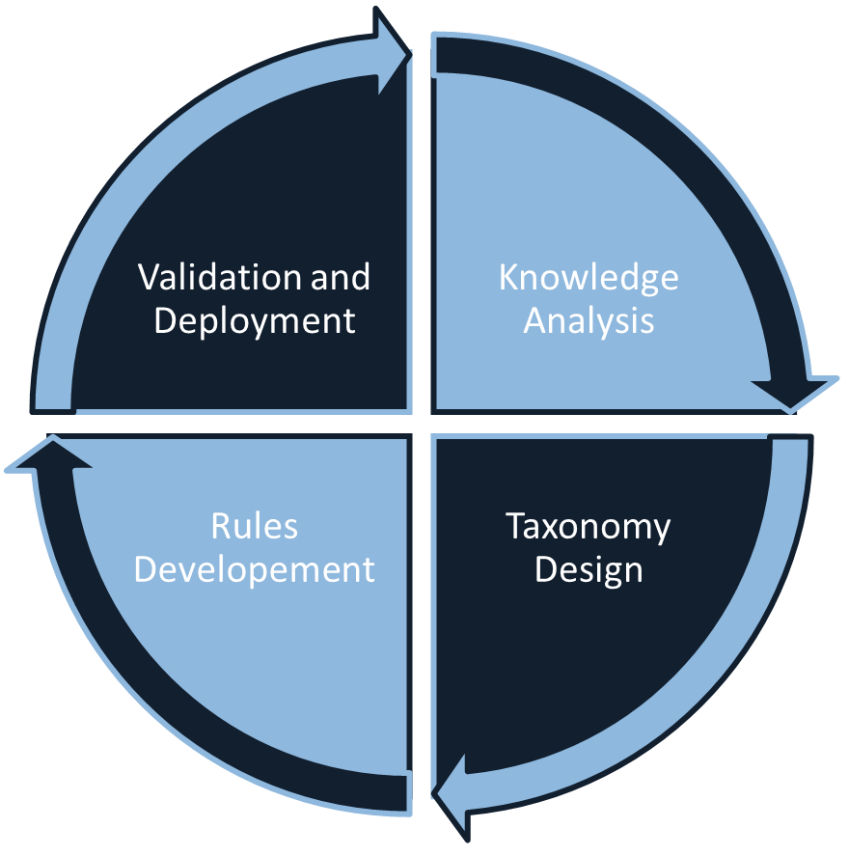


Figure 28: Taxonomy Development Cycle.

Knowledge Analysis: is the first step in a taxonomy tree (also called domain tree) development and consists of analysing the application domain of semantic analysis by studying the requirements for the creation of taxonomies together with the definition of the objectives. At this stage, several aspects have to be considered, including the domain of analysis, the amount of data to be processed, the languages of the documents, the sources and types of documents, the syntactic structure and lexicon of the documents. This is necessary for the preparation of the following steps. For a good analysis, it is also necessary to define the project's objectives, particularly the fundamental aspects of the analysis, the point of view to be kept on the development of the knowledge base and the expected results at the end of the taxonomy development project.

Taxonomy Design: the taxonomy choice or creation is the second step in a categorization project development cycle and a crucial activity. It summarizes and classifies the domain knowledge, clarifies

a specific point of view on the domain knowledge, and is used to attract and aggregate information through categorization rules. The taxonomy definition process goes through iterative cycles, during which the relevant information is selected and organized into several interconnected categories. Usually, several tree drafts are produced and submitted to the judgment of a group of selected final users, who test the trees for consistency, clarity, concision, usability, shareability, scalability, etc. The development team gathers feedback from the final users at each evaluation cycle in order to progressively refine the tree and produce a final, shared version. With the final version ready, each category is given a precise description that will constitute, later, the basis for rule writing. Each taxonomy should then be associated with a library of sample documents representative of the relevant category. This library will be used for accuracy testing.

Rules development: is the third step in a categorization project development cycle. This task can be described as the identification and abstraction/formalization of conditions or parameters that allow relating each domain in the taxonomy to one or several documents. These rules allow to identify the relevant linguistic elements that can be observed in a document written in natural language. More in detail, each rule is univocally associated with a domain and identifies the presence (or absence) of one or more linguistic elements (words, concepts, etc.) in a text or in a portion of it. If the condition that the rule formalizes is verified, the rule assigns a score to the text, which links the latter to the category associated with the rule.

Validation and Deployment: The taxonomy was validated using a benchmark test to establish the correctness of taxonomic rules. Through the use of already analysed libraries, the test can give a percentage on the accuracy of the taxonomy. It is based on an iterative process that allows the rules to be evaluated and modified according to the result. The taxonomies are then exported and provided to the end-user within the platform.

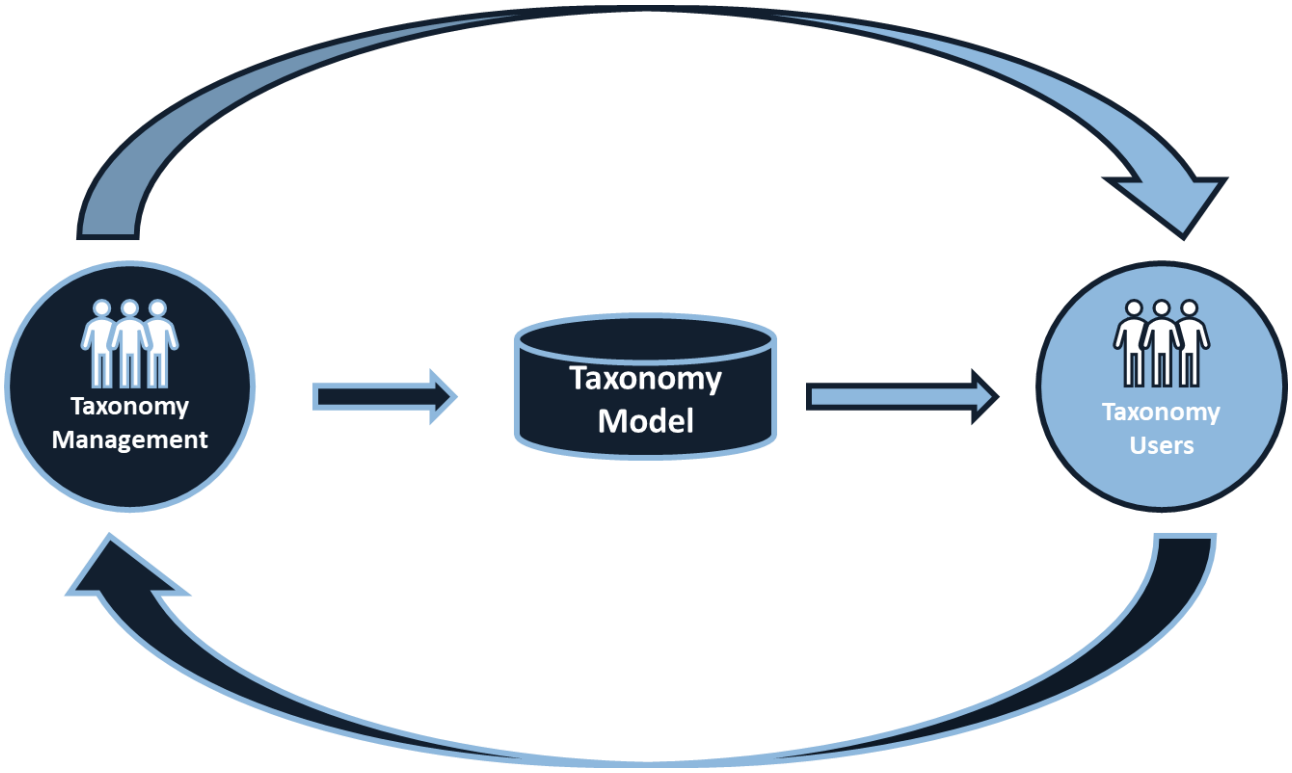


Figure 29: Taxonomy development iterative cycle.

In the Elettronica business case analysis, an entry-level taxonomy of Electronic Warfare (EW) was initially developed to contextualise the possible suppliers within the market, identifying their capabilities and heritage. Therefore, a categorisation approach was used to create a semantic tree composed of general concepts that helped us understand the scope in which a company operates and specific entities to study the differences between players. Within Elettronica, a shared document was created between various experts and analysts who helped to categorise the Electronic Warfare sector from general concepts to more specific capabilities. To date, taxonomies concerning EW macro-areas (i.e. C4ISR, SIGINT, COMINT), which will be briefly described in the next chapter, have been developed. Taxonomies concerning military platforms and EW product categories are under development. Within this research was created a document containing future developments of the procurement project representing the domain for the analysis of 3D Antidrone Radar. The project was developed through a categorisation approach. It started with the concept of 3D radars and then, with the help of experts, all the most important technical and commercial characteristics for the study of these products were decided. The file was then sent to the engineering section of Cy4Gate for development. Future meetings with the devs will be necessary to formalise the rules. Development, as mentioned earlier, is an iterative cycle between the engineers and the end-user that aims at continuous refinement of the rules in order to produce a taxonomy that allows meaningful semantic

analysis. It is, therefore, essential that the parties within the project talk to each other continuously for the development of the platform.

1 Planning

The first phase of the project involved planning the work and creating all the archives for the data lake and report compilation. First of all, a new archive was created using the function within the QUIPO inner circle (chapter 3). This new archive was used to contain all the items collected in the data gathering process, which we will see in the next paragraphs. This space contains the data lake of all the articles that are analyzed only by the semantic engine of the platform and that will require a deeper investigation by the analyst. The creation of the archive was a key step for the project since, thanks to the collected data, it was possible to build a first dashboard able to give an overview of the quantity, the substance and the significance of the collected data. In order to carry out an analysis as complete as possible, several archives can be used within a case. This was vital because it is possible to retrieve information prior to the start date of the analysis that the platform has already collected. Then the task was created in which all the significant articles were inserted after being viewed and validated by the analyst. Similar to the archives, the tasks are spaces where data are stored. We could define them as second-level archives subject to two analysis processes: automatic and manual. In their creation, it is necessary to specify their name, the people who will work on them, the implementation times and the priority with which the case must be carried out. This, along with the purpose of the analysis and the specifics of the business case, is discussed in a meeting and reported back into the platform. After this first step, the data gathering process begins.

Target

At this stage, it would be very useful to be able to migrate archives from other cases as well. Since QUIPO was conceived for use in defence and legal companies, this function was not considered necessary because of the confidentiality of certain types of analysis even within the same organization. However, over the years, it has become clear that QUIPO can also be applied in other areas where the ability to have a single archive that includes all the documents collected and then redistributed within the platform is essential. This function is currently being developed in the second version of QUIPO, also designed for corporate use.

2 Data Gathering

The data-gathering phase is one of the most delicate moments within the analysis. It is essential to perform a first skimming of the collected articles in this step, doing searches as relevant and intelligent as possible. In fact, the risk is to conduct summary searches for keywords with very broad meanings and to find the archive filled with tens of thousands of documents that are completely useless for the analysis and that increase the analyst's difficulty in recognizing the articles to be included in the task. It is necessary to have a thorough knowledge of the platform and its reasoning method for data ingestion at this stage. In this way, it is very easy to make more precise research that allows a clear overview of the phenomenon that is being observed at first glance.

In the business case analysis, all of the methods made available in QUIPO for data collection were used to test their potential and problems. Within the news box, the following were employed:

- RSS source to include within it the collection of information from various specialized sites dealing with defence & aerospace to have an archive of reliable sources. The problem is given by the variety of documents that do not always refer to the subject of the topic, but the authoritativeness of the header added to the possibility of using the articles for other analyses make the use of RSS sources essential.
- The Google News source allows the analyst to search for articles by topic, keyword or country. In order to avoid having too many misleading results, a keyword search method has been used. In fact, the names of companies or precise events considered to be of key importance for the analysis were searched. On the one hand, this allowed to have a cleaner working environment and, on the other hand, to exploit the power of the semantic engine to analyze the collected documents as soon as possible. The presence of Google News was fundamental to obtain precise information on the contracts and partnerships of the three companies.
- The Europe Media Monitor source works in a similar way to Google News, so much so that the same data collection procedures were used.

Regarding social research, the following sources were used:

- The source Twitter, due to its inherent characteristics, was definitely the most useful social for the research. Like other organizations operating in the defence market, the three companies have a greater presence on Twitter or Linkedin than on other social networks. In addition to following the companies' official accounts, a brief search was done to find relevant accounts that talked about the reference sector of the research, specifying some hashtags considered interesting.

- Instagram and Facebook were used in the same way. On the one hand, it was tried to include, where present, the official pages of the companies of reference, on the other hand, to search for influential profiles on the subject that could have interesting information. Instagram was not used much because it was not considered essential to the research. In fact, the social network is indicated more for tasks that require the analysis of images for the recognition of people or objects. Also, the function that allows images to be collected from documents for analysis was never performed in the collection of other documents. Even if, in this business case, the image analysis function is not useful to find relevant information, in the design of the procurement function, it is instead essential to analyze and recognize certain products and consequently educate the platform to find similar ones.

Within the Web section, the following sources were used:

- Through the web crawler function, it was possible to enter the information contained in a particular website. The advantage over a normal search is the possibility to enter a time range in which the platform automatically collects news within a particular web page together with the possibility to select a depth of information collection. Quipo is able to collect data with depth 0 (only the information contained within the web page), 1 (links within the page) and with depth 2 (links within the links of the main page). This method of ingestion was mainly used to insert articles from the press releases of the three companies under analysis and the reports.
- The Wikipedia function was used to insert information on political events into the archive that was fundamental to outlining the geopolitical framework of our analysis. Data on reference organisations were also collected in order to have more information on past activities (all information was collected with 0 depth).
- The Search Engine feature was used to refine our search by seeing if documents not present in our archive could be taken from a search.

In the Other section, the following sources were used:

- Through the options Internal DB and Email, all company documents were entered to investigate certain topics. The Email function is necessary, for example, for the ingestion of the Elettronica press release, which is fundamental for acquiring information on international relations between companies operating in the sector. Thanks to the platform is also possible to filter the topics of our interest. The two options are very useful because they allow the semantic engine to analyse company data that would otherwise take a long time to process.

In the construction of our data lake, were used as sources: news from major defence websites and magazines, Google News, Europe Media Monitor, the press releases of the analysed companies and related organisations, ministerial websites, social media (Twitter and Facebook), the press release of Elettronica and from archives of documents already present in the platform from previous searches to have prior data to our research. In choosing the sources and methods of data gathering applied, the NATO Open-Source Intelligence Handbook guidelines were respected. (Gibson et al. 2016)[33]:

1. The authority of the source
2. The accuracy (by validating it against other sources)
3. The objectivity of the source
4. The currency (i.e., the provision of a timestamp for publication and the presence of an author)
5. The coverage (the degree of relevancy)

The data gathering process was conducted by searching for very precise keywords that included the name of the companies under study, their products or geopolitical concepts related to the research. In three months, the specific archive for the analysis had 30,000 articles, to which were added other archives containing news from magazines and defence fairs on the major players in the market (500,000), the press release of Elettronica (200,000), news on investment banks and related companies (20,000) and geopolitical information (1,500,000). The data lake, at the time of analysis, consisted of 2,250,000 articles analysed by the platform. The next step was to process them by filtering the documents within the data lake according to taxonomic concepts and the companies of interest. In the selection of the significant articles, the same rules for the identification of sources were applied. It was necessary to repeat the process because some media such as Google News, Facebook and Twitter require more careful analysis to avoid the ingestion of unreliable news. After filtering the articles of interest, reliability (low, medium, high) was assigned to the most interesting ones. Finally, an archive containing all these documents was created and compared with the attainable information in our possession.

Target

The data gathering function is a core component of the intelligence cycle. Through the correct construction of the data lake, it is possible to accelerate, simplify and improve the analysis and the resulting information. Although there are many functions within the data gathering part, it is important to work in two parts:

[33] NATO (2001) NATO open-source intelligence handbook.

- Improving the data collection with new customised tools that allow the end-user to use an essential source for their research or to perform a specialized search. Through this improvement, the scope of use can be further extended, allowing the platform to specialise in specific tasks or to communicate with other services that are widely used in a company. Within Elettronica, for example, one of the objectives is to include in the collection process the possibility of having direct access to the journal databases to which the company subscribes to have more authoritative and precise data in less time to complement open-source research.
- Making existing tools more elastic so that that data collection rules can be further refined. This is very useful for the automatic alerting function, where articles can be inserted automatically within a task. In addition, thanks to the ability to customise searches, it will be possible to greatly reduce the amount of data collected, speeding up analysis times.

In addition to this, work needs to be done on integrating new socials into the platform such as Linkedin, for example, could be a key part of the procurement project, to understand the direction of companies on new technologies and to search for advertisements for new professionals. Reddit would also be suitable for inclusion in the platform's social park. The social network has a very good reputation and a large user base that talks about various topics.

3 Analysis

The QUIPO analysis section is the heart of the platform and the analysis process and is divided into two modules:

- Integrated: where there is the explore function that incorporates all the other modules (Special) and from which you can consult all the documents collected.
- Special: where there is the structured analysis able to display the dashboards created, the link analysis to automatically create graphs, the darknet section used to display information collected on the dark web and the image analysis that uses machine learning to recognise faces or objects and to tag images.

Within the business case, the explore, link analysis and structured analysis functions were mainly used.

First part: data lake analysis through dashboards, link analysis and taxonomies

The explore module was of paramount importance to filter the documents collected in the previous data gathering step. As seen in the previous chapter, this section shows in the middle the news, posts,

images and tweets collected, while on the left are the various taxonomies, which the semantic engine Cogito uses to enrich documents and find concepts within them, along with archives and tasks. Finally, at the top, there are various filtering possibilities concerning the source, the collection time, the type (image or document), the language, the reliability (assigned by the analyst), the task and the reading status (read / unread). Above the filters, there is a search bar where you can search or exclude one or more keywords within the text or title.

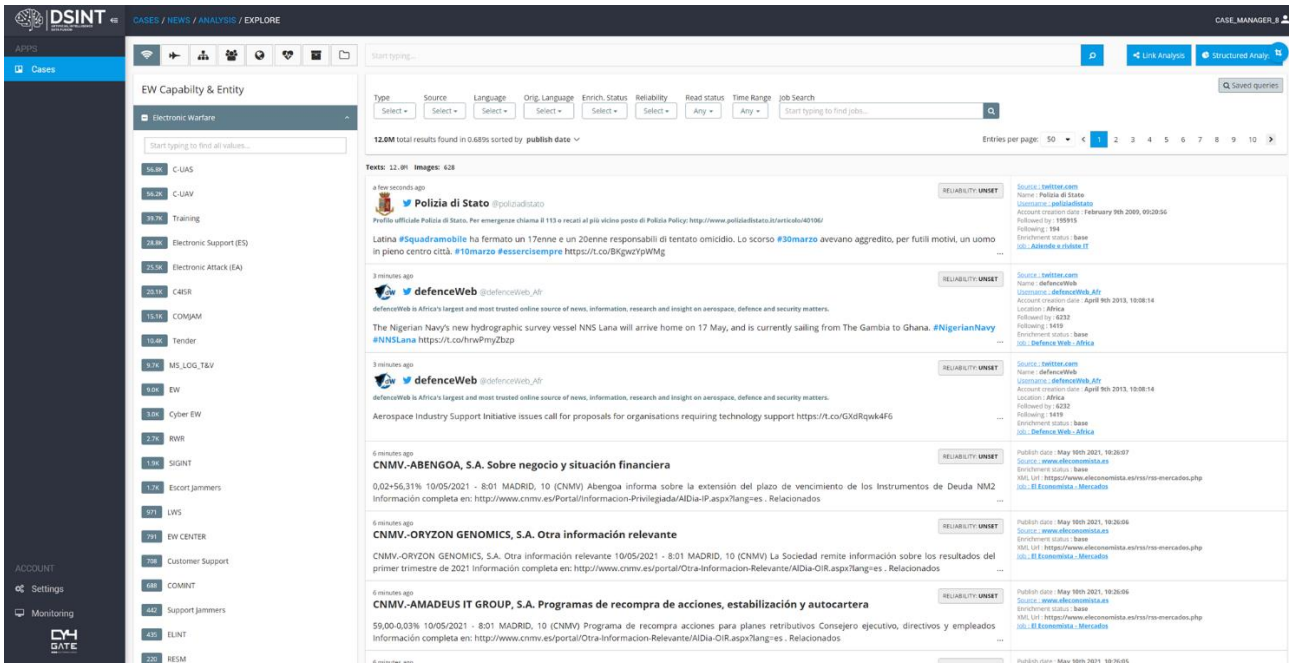


Figure 30: Explore Module with documents (centre), EW taxonomy (left) and filters (above).

To fully exploit the potential of the platform, a dashboard has been created to provide an overview of the data collected in the various archives. The interactivity of the charts allows cross-searching using taxonomies. In addition, it is possible to get information from QUIPO's Heterogeneous Data Analytics module, which aggregates document metadata into graphical representations.

Since the first dashboard was mainly designed for filtering work, word cloud plots together with heatmaps were used. The word clouds were constructed to identify occurrences of the number of intelligence, electronic warfare and organisation concepts indicating the frequency with which a particular concept was detected. Heatmaps were created to cross-reference the analysed company data with the various taxonomic concepts or other organisations to identify the presence of relationships.

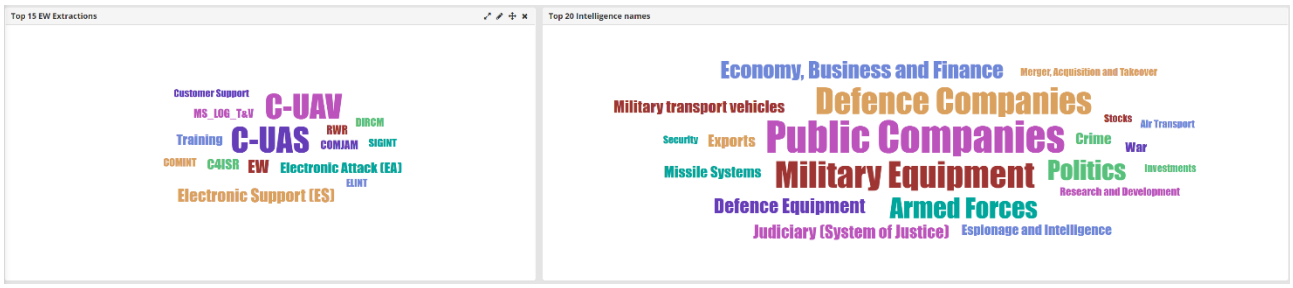


Figure 31: EW and Intelligence Wordcloud.

During the analysis, filters were very important to identify relevant search items among more than 11 million documents. Thanks to the normalisation rules, a category including the companies under analysis was created within the taxonomy dedicated to the organisations' recognition. This made the search process much faster. The rationale behind the analysis in this section was to filter by topics in a very targeted manner, to analyse the documents present each time and, if deemed reliable and useful, to include them in the task.

Searches were conducted by filtering each time the companies analysed (potential suppliers) with the intelligence taxonomies for the concepts of:

- **Merger & Acquisition:** to see the acquisition activities of the companies, understand their worldwide distribution and which technologies the organisations are investing in.
- **Export:** to see the degree to which companies are exporting and the products they are choosing to share abroad.
- **Joint Ventures and Partnerships:** to see the collaboration with other companies and the technologies they choose to focus on.
- **Investment and Research & Development:** to get an overview of the company's investments and the most important technologies.

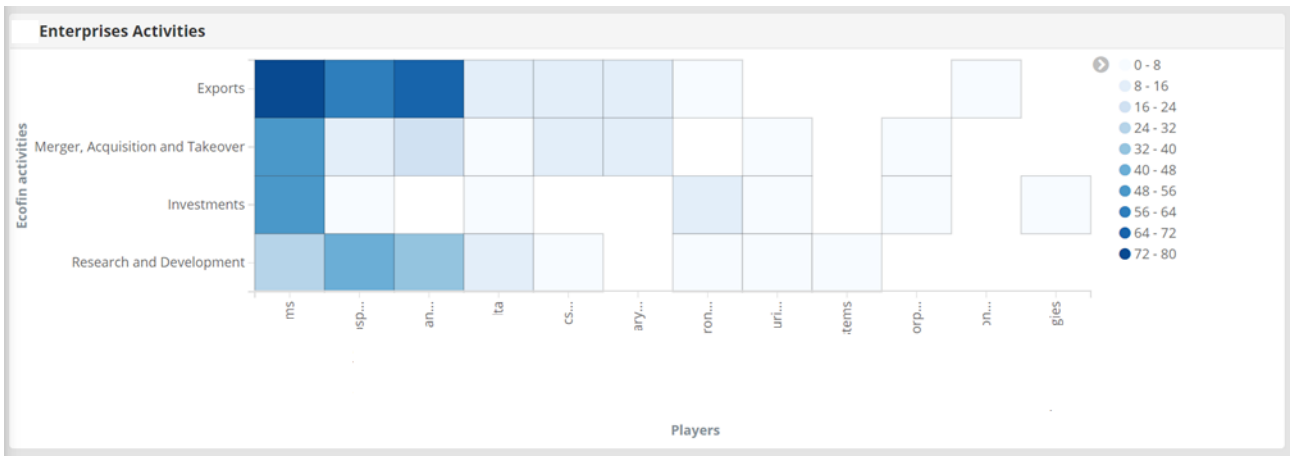


Figure 32: Heatmap representing Intelligence vs Organisation Taxonomies.

The taxonomies created internally by Elettronica concerning EW (Electronic Warfare) capabilities were then used. EW is defined as all the actions related to the control of the radiofrequency electromagnetic spectrum to allow an electronic attack or prevent it. The taxonomy in question was essential to analyse the capabilities and to find the products related to a specific company. It also allowed to filter the data in a solid way allowing to carry out an analysis in a short time focusing only on the features of interest. In this category, the most interesting capabilities were selected for the analysis of the supplier base.

1. C4ISR: Command, Control, Communication, Computer Intelligence, Surveillance and Reconnaissance include the complete fusion of several sensors to provide a battlespace awareness, supporting cooperative engagement of enemy using its operational plans and tactical behaviour.
2. DIRCM: Directional Infrared Counter Measures is a system installed on a flying platform designed to counter infrared ground to air threats.
3. Training: Are the activities that teach the operators to use EW systems.
4. Customer Support: This is the activity devoted to support customers in the operative and logistic use of EW Systems.
5. RECM: Radio Frequency Electronic Counter Measures is an EW system devoted to jam and deceive enemy landed, platform installed and missile radars.
6. COMINT: Communication Intelligence is an EW System devoted to receive and decrypt communications.
7. RESM: Radiofrequency Electronic Support Measurements is the EW system devoted to detect, intercept, identify, measure and record all the radiofrequency emission present in a specific operational area in order to recognize an immediate threat.
8. SIGINT: Signal Intelligence is an EW system devoted to collect and analyse signals, both radio or electronic, in a large scenario.
9. ELINT: Electronic Signal Intelligence is an EW system devoted to detect, measure, deeply analyse and collect electronic signal (generated by radars).

All documents were then assigned a degree of trustworthiness that made it possible to distinguish documents from the most authoritative sources from those of dubious origin. More precisely, the following tags were assigned:

- Reliability High: the most significant articles from the most reputable defence publications and world-renowned newspapers. Documents from the press releases of the companies analysed and from Elettronica were also included.

- Reliability Medium: this section includes relevant information from blogs and social networks. It was decided to proceed in this way to compare the content of these articles with those included in the high-reliability cluster to see if the data contained in both groups led to the same results.
- Reliability Low: In this section were included the less important articles (however useful for the research) and the more relevant articles, but coming from sources that could be conditioned by the international relations of their government.

Similar to the dashboard described above, the link analysis function was also used to inspect the relationships between organisations, companies, EW capabilities, companies and intelligence concepts. With a dynamic graph available, it was possible to find links that had not been taken into account. Through this function, it is also possible to select a certain node and underline all its arcs in order to see all its relations with the various taxonomic concepts.

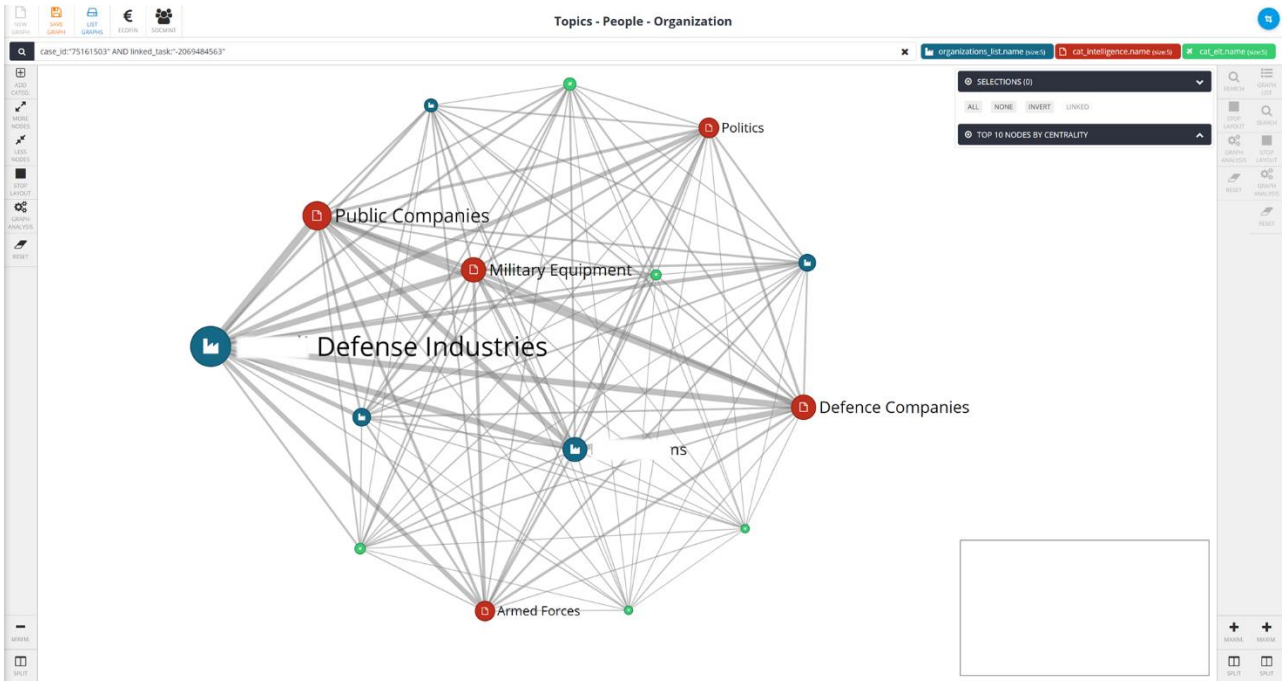


Figure 33: QUIPO Link Analysis with EW, Intelligence and Organisation taxonomies.

Second part: analysis of the task for compiling the report and creation of the dashboard for presenting the results and validating the EW taxonomy

After analysing the relevant articles within the data lake, a dashboard was created containing 932 documents that were entered into the task to compile the report. The construction of the graphs helped once again to filter out the most important content for analysis. With a clean archive of data, it was possible to graphically represent some of the insights from the research, such as:

The geographical activity of the companies divided by contracts, merger & acquisition (M&A) and joint ventures and the count of documents related to these activities. This was very useful in understanding the international relationships of the companies along with the geopolitical landscape in which they operate. By analysing the following graphs, relationships emerged on the companies' governance, investments received, investments made and relationship with the governments of the countries in which they operate. The graphs used are:

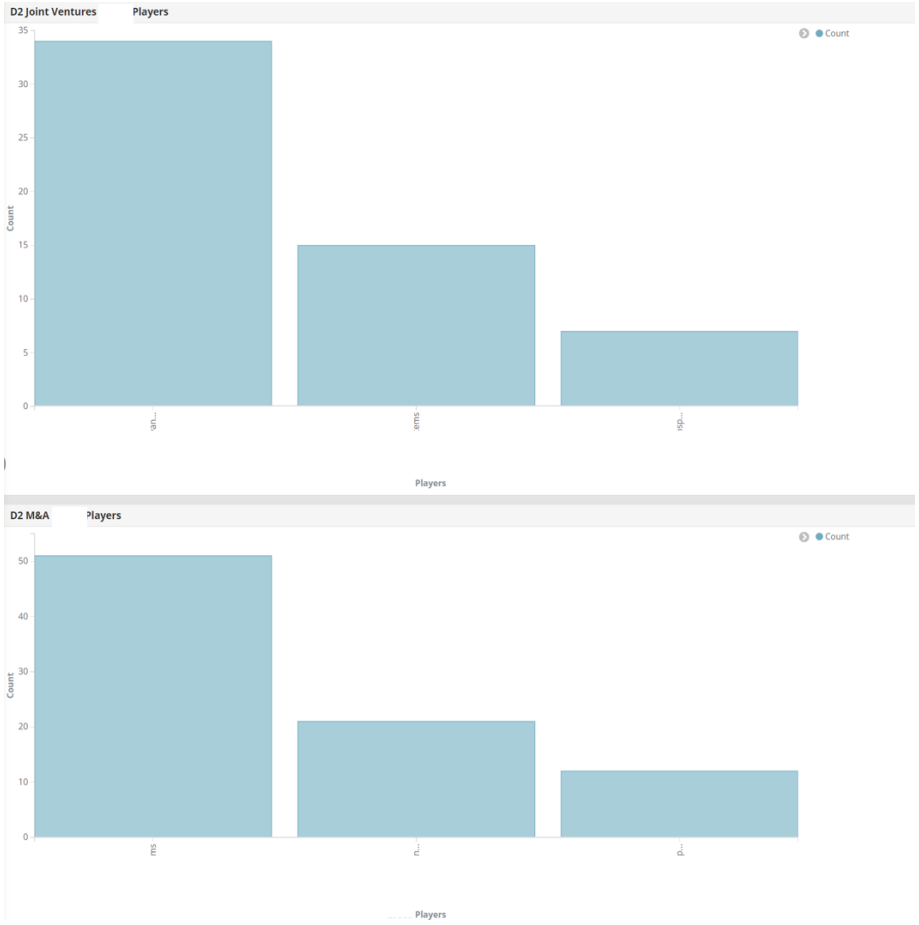


Figure 34: Bar Plot: count of articles containing intelligence concepts.

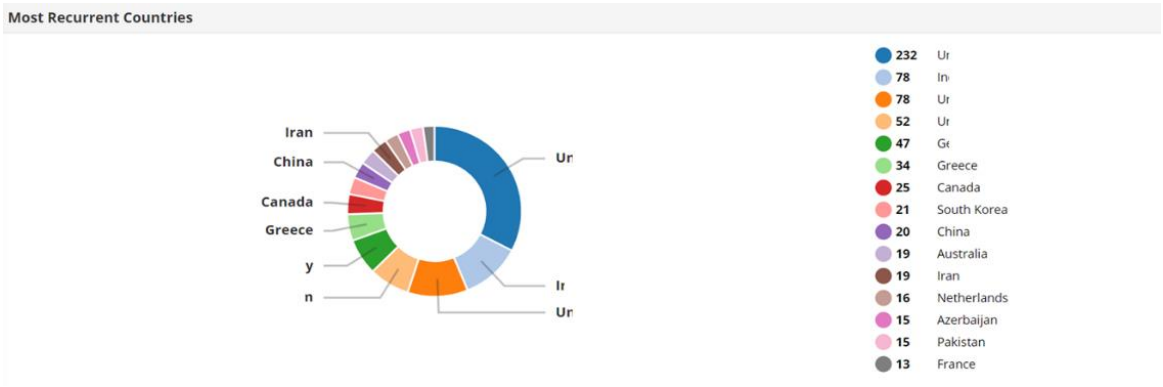


Figure 35: Most recurred countries per contract.

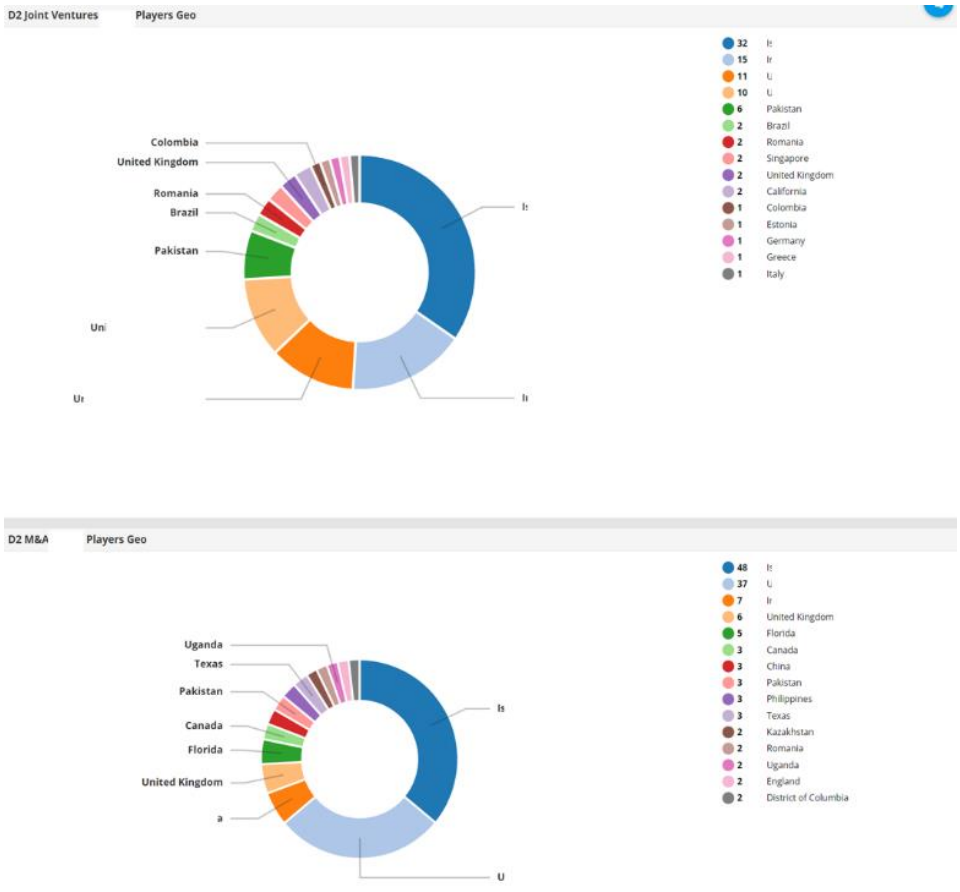


Figure 36: Donut Chart: division of M&A activities, contracts and joint ventures by country.

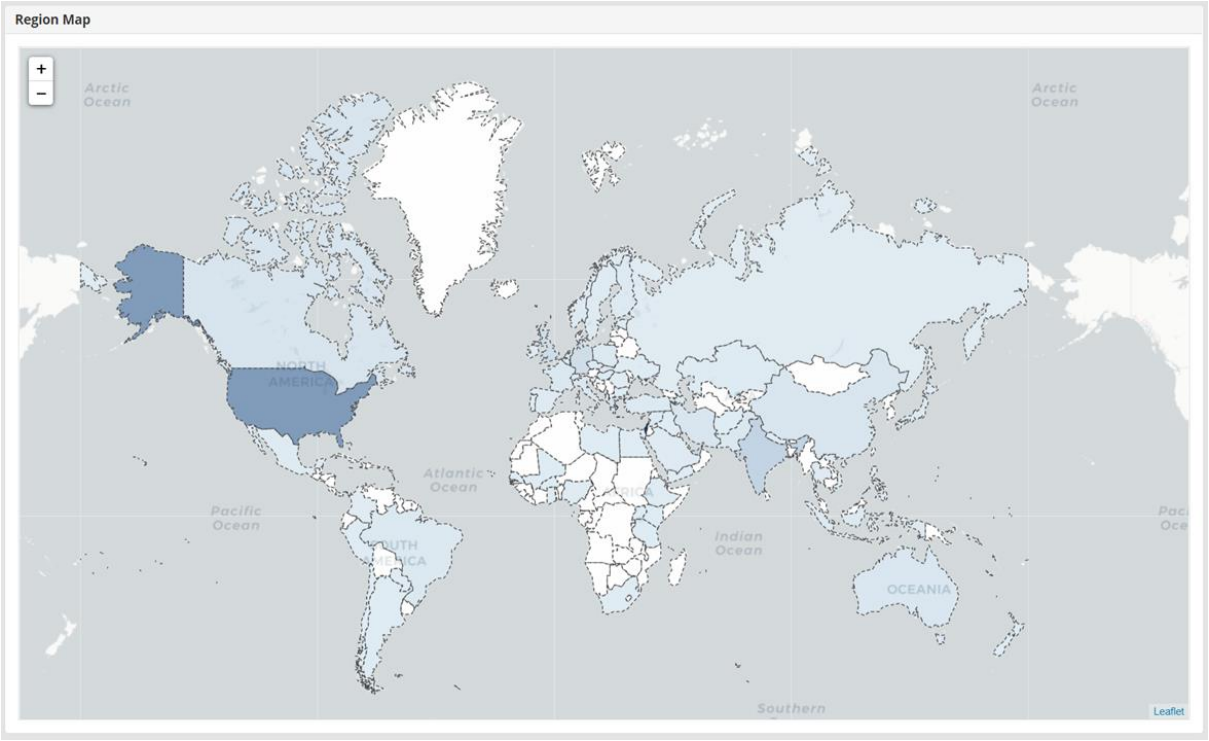


Figure 37: Geographic Map: count of documents within the task that refer to a specific country.

Since in the defence market, G2G contracts are of paramount importance for defence companies. It was tried to find out within the task how often the country's government agencies, where the analysed companies have their headquarters, were mentioned by counting the number of times a certain word appeared in a document. The relationship between government and companies in this country is widely acknowledged. The main objective for creating these metrics was to see if the platform could capture this message in the absence of a suitable taxonomy.

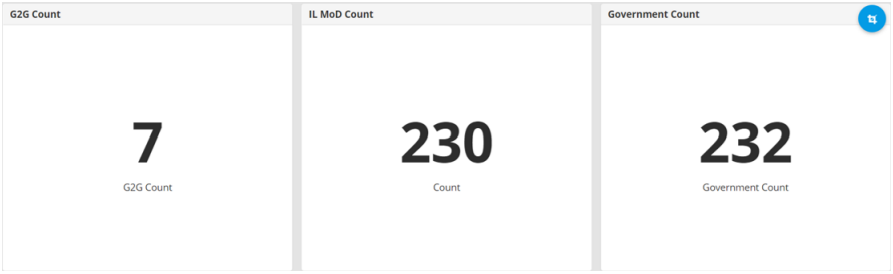


Figure 38: Word count metric to establish government presence on documents.

As was done in the first part of the data lake analysis, heatmaps were built to cross-reference data concerning:

- Capabilities and Companies: to understand whether a specific company owns a certain product that interests to Elettronica
- Capabilities and Partners: to understand in which sectors there is a collaboration with other companies[34]
- Companies and Partners: to understand the relationships between the analysed companies and other players in the defence market. Necessary to know about contracts, partnerships, acquisitions and investments.
- Companies and Platforms: to understand on which platforms the companies operate most[35][36]

The construction of these heatmaps was essential in answering the research question. Through these graphs, it was possible to more efficiently compare data derived from OSINT platform analysis with data derived from business intelligence. Furthermore, the dynamic graphs allowed a faster and more in-depth analysis of the items contained within the task, allowing for greater accuracy when compiling the report. The capabilities analysis was also the most important part in terms of the interests of Elettronica's procurement function. The project's main objective is to acquire information on potential suppliers and their technologies to find better solutions for the company's products and have as much

[34] The focus was on European defence companies.
[35] At the time this thesis was written the taxonomies regarding platforms were in an experimental phase, therefore they were not taken into account for the analysis.
[36] Only the airborne taxonomy is shown in the graph.

information as possible on the companies' due diligence. The four main heatmaps for analysis are shown below. The first chart was the most important because it contains the internally developed taxonomies that will be subject to mathematical and empirical validation.

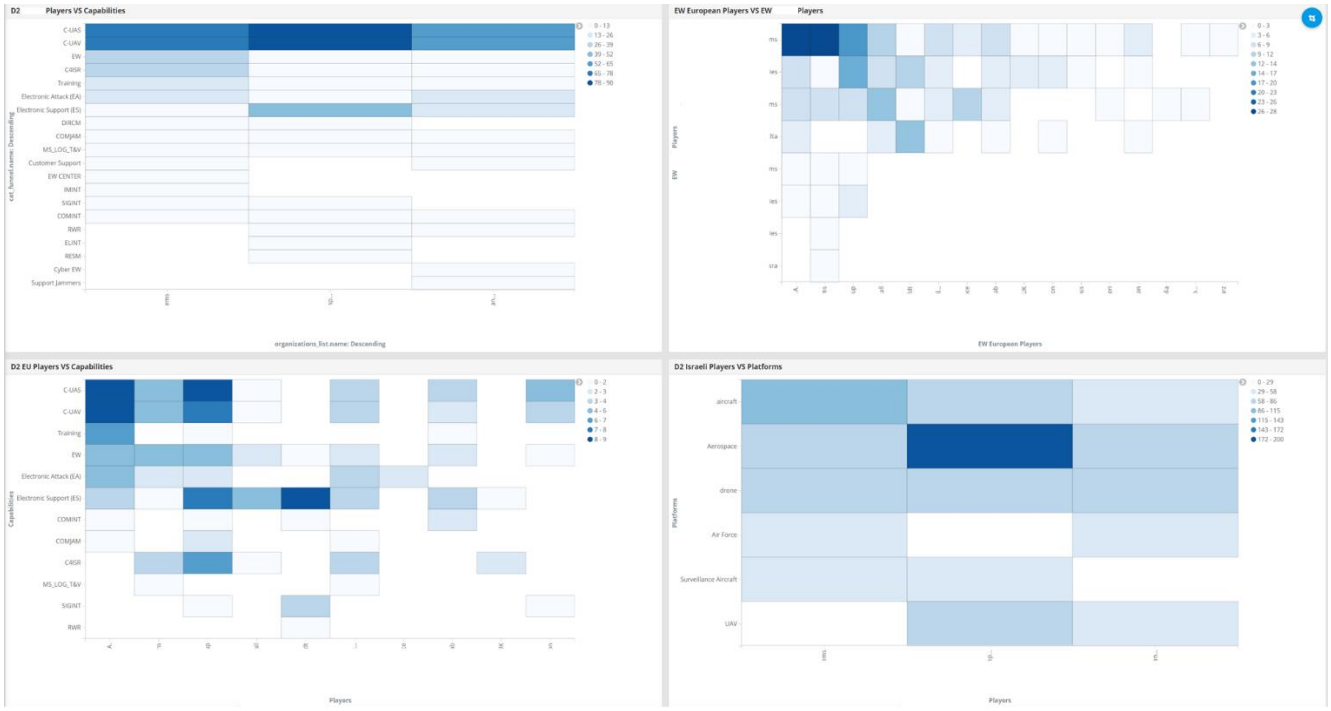


Figure 39: Heatmap for cross-data representation.

Similar to the first part of the analysis, the dashboard, taxonomies and link analysis were used to further analyse the documents within the task. The presence of underlined taxonomy concepts in the text avoided the need to read long documents and made the whole process much faster. After finding all the necessary information, the report was compiled.

Target

The analysis phase took almost three months to complete and was the longest part of the whole case, so the Explore and Analysis functions were the ones where most time was spent. In this section, several aspects were solved and implemented to improve the search and filtering of documents within QUIPO. The following functions have been requested

- Reported and solved the ingestion block of Elettronica press release.
- Reported and solved search or exclusion of a keyword limited to the title of the article: practical example, a limitation of the data collected through Google News and search engines is to collect market report articles in which certain concepts are enclosed in bulk. These documents give no added value and are confusing, so it was possible to eliminate them by searching for all data that do not have the word market in the title.

- Reported and solved filter retention when switching from one analysis function to another: practical example, during the analysis in the structured analysis section, if you switched to the explore or link analysis functions, some filters were not retained, such as the temporal filter.
- Reported and solved improvement in keyword research: practical example, sometimes it could happen that a particular word searched for was not present in the document analysed or indicated part of a compound word.
- Reported and solved insertion of new taxonomic categories for the creation of graphs in the link analysis function: this is essential for analysing data with EW's internal taxonomies.
- Reported and solved ingestion problems in the web functions of the platform: practical example, the problem that did not allow the ingestion of web pages with 0 depth was solved. A blockage in the ingestion of internal documents has also been resolved.
- The absence of a tick (article read) when switching from one article to another via the next button has been solved.
- Request for the possibility of migration of the archives between different cases: in this way, if it is necessary to create a new case, the articles already collected by the platform can be used.
- Request for a new label to be added to the reliability one to establish the relevance of the document: this would make it easier to divide up the articles within a task.
- Request for a specific way of displaying metadata on the source, job, number of similar articles and the possibility of displaying statistics for each source.
- Request to be able to insert keywords in link analysis.
- Request for the integration of ORBIS Bureau Van Djik for the financial analysis of organisations into QUIPO: This last point could strengthen QUIPO commercially and at the same time increase its flexibility of use. Thanks to the integration of similar apps, it is possible to quickly find financial data, ownership structure, financial risk and key employees of a company and visualise everything through link analysis. A practical example is in the analysis performed. It was necessary to know the distribution of the analysed companies' subsidiaries to understand how much and how they were present outside their borders. This analysis was carried out off the platform due to the lack of a connector to make the two applications communicate. The implementation of ORBIS within QUIPO is currently under development and expected in the coming months. An entire afternoon's work was required to create the following graph. With the presence of the database within the platform, it would only take a few minutes.

Figure 40: Subsidiary distribution of one company analysed (Source: ORBIS Bureau Van Dijk, April 2021).

The work that was done to develop the platform was very important for the development of QUIPO within Elettronica and Cy4Gate. On one side, it will allow Elettronica to optimise the business processes of its core functions by automating time-consuming analysis steps. By looking at the procurement scouting process, the platform has more than halved the time needed to analyse the supplier base, allowing employees to focus more on the selection and contractual processes. QUIPO takes time to be trained to perform certain tasks, but once this first part is over it is able, through automatic alerting functions, to signal the presence of interesting information at any time, allowing analysts to receive insights faster than before and, consequently, speeding up business operations.

On the other hand, this development project will also help Cy4Gate, allowing it to test the platform in different contexts and enabling it to correct some problems faster than before. In addition, the requests made during this period could be useful for enhancing QUIPO's business strategy. For example, the addition of services such as ORBIS (financial data/company ownership) or the integration of specific databases could make the platform more attractive to the market. In the end, data is the resource on which QUIPO and all business intelligence, in general, is based.

4 Report

After the analysis part, a report was compiled containing a summary of all the information from the previous step. According to a Top-Down logic, the document was divided, starting with an overview of geopolitical events in the relevant markets. Then, the merge & acquisition activities, partnerships, joint ventures, orders backlog of the analysed companies and their subsidiaries were analysed in detail. A section was also dedicated to the analysis of financial data for the calculation of financial strength and solvency.

Thanks to this data, we were able to cross-reference various types of information and understand the context in which these companies operate and their relationship with their government. Then, thanks to the information on activities, it was understood the degree of innovation, the direction they are taking in the development of new technologies, who their partners and customers are, what their strengths are and in which sector they are most developed. Through the analysis of the subsidiaries, it was understood how the three companies are distributed worldwide. This information was essential for a 360-degree assessment of the potential supplier, which included:

- Geopolitical information to understand international relations
- Activity data to identify capabilities and product heritage
- Investment information to identify new technologies and investment opportunities
- Financial data to assess economic stability
- International distribution

The report is closely linked to the process used within the Electronica procurement department. One of the objectives is to automate the first part of the scouting process to identify potential suppliers of core products for the company's operations. The report was then disseminated through the AMICO platform, which will be described in the next section.

Target

In this section, compatibility with a more complete text editor could be included, which would also allow the creation of presentations and better image management.

5 Dissemination, AMICO Platform

The report, together with the task news, were automatically inserted into AMICO. The graphs in this section are partly created using QUIPO's knowledge base function, which allows the development of customised graphs that summarise the report's contents and partly generated automatically by the platform on the basis of the information contained in the report or articles in the task. Within AMICO, it is also possible to view images and videos collected during the analysis together with the origin of the sources through a dashboard that summarises useful quantitative information about the task data. The platform was used to present the state of work to the company management in order to show the progress of the company's digitalisation project. Through AMICO, it was also possible to show the development progress to the technical and commercial management of Cy4Gate in order to illustrate the status of the work and to request the addition of the new functions described at the end of each step above.

Study Validation: Process and Results

For this research, the focus was on the first cluster of taxonomies available in the procurement project concerning firm Electronic Warfare capabilities in the defence sector by examining three firms that belong to the same capability cluster. Several reports available in the company were inspected and a file containing information about the contracts, merge & acquisition activities, partnership and joint ventures of the organisations under study was compiled. The file containing 170 occurrences was then validated by the company functions. At the same time, was analysed the documents contained in the QUIPO data lake, resulting in a final archive of 932 items. The nine taxonomic concepts concerning Electronic Warfare capabilities within the documents were selected and divided into clusters all the articles for each specific concept. Then calculated the percentage of each cluster was calculated with respect to the total number of documents in the archive to normalise the data. The same it was done for the data inside QUIPO. Next, was studied the existing correlation between the two matrices created through the following formula for the calculation of the Pearson Correlation Coefficient:

$$r = \frac{\sum[(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{\sum(x_i - \bar{x})^2 * \sum(y_i - \bar{y})^2}}$$

The results (see fig. 41, table 1) indicate a highly positive correlation (Mukaka, 2012) between the two matrices ($r = 0.927$, $R^2 = 0.86$). That denotes a correspondence between analysis based solely on OSINT techniques using a business intelligence platform and data validated by the firm's experts. An empirical correlation also emerged when compiling the report, noting a sufficient match between the concepts extracted from the platform and those present in the reports analysed.

In light of the following results, the following conjecture was formulated:

"In the development of a Business Intelligence platform, which uses analysis techniques based on the OSINT intelligence cycle, the creation of a data lake following precise rules that point to correct use of sources together with the creation of an appropriate taxonomy to the specific domain, allows the number of occurrences on a given entity to be very close to reality".

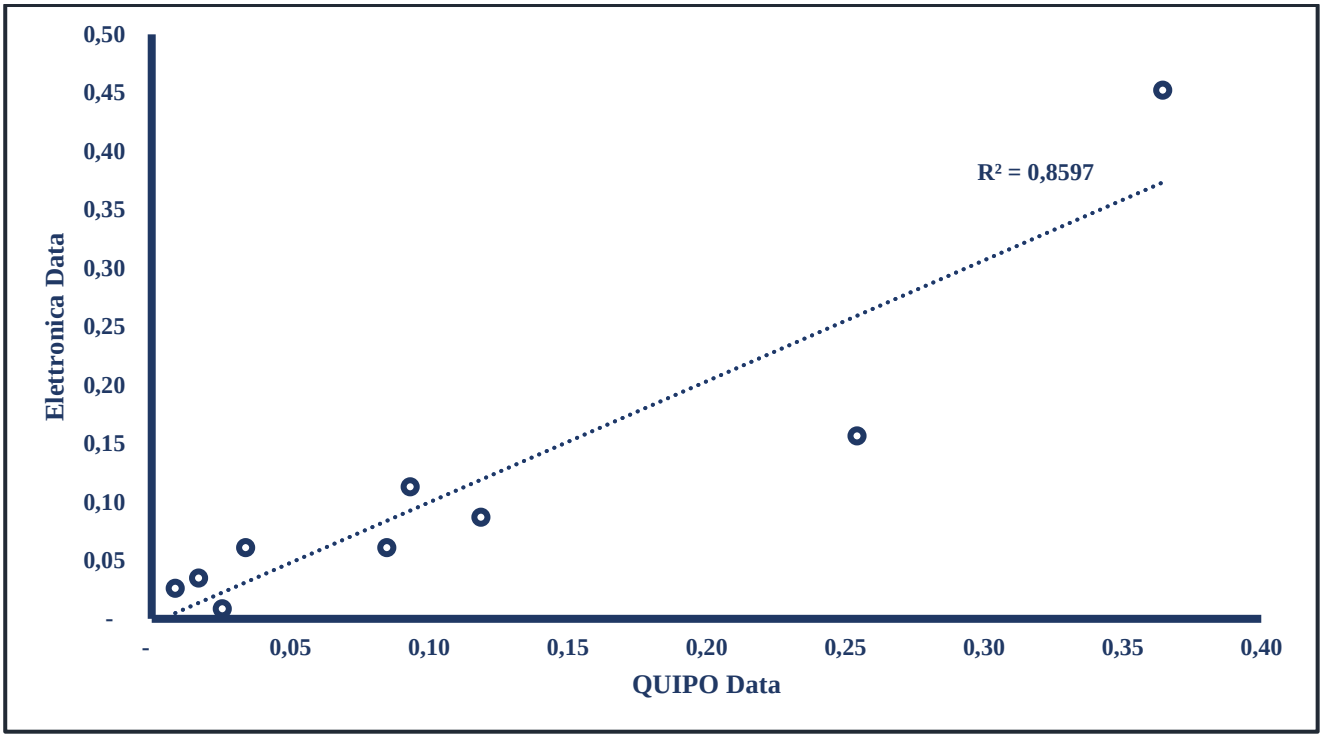


Figure 41: Graphical representation of the correlation between the two matrices (EW Taxonomy).

Table 1: Capabilities examined and their percentages on total documents (EW Taxonomy).

Capabilities	QUIPO Data	ELT Data
C4ISR	0,36	0,45
DIRCM	0,09	0,11
Training	0,25	0,16
Customer Support	0,03	0,06
RECM	0,01	0,03
COMINT	0,12	0,09
RESM	0,02	0,03
SIGINT	0,08	0,06
ELINT	0,03	0,01

After the conclusion of the analysis, the same kind of study was done on the geography data in order to see if it was necessary to build a new geographical taxonomy to add to the default one developed by Expert Systems. The analysis was done by choosing the five most representative countries within the platform and the in house analysis. Again, the data are encouraging, although not as good as those concerning the EW taxonomy. In this case, the matrices are highly correlated ($r = 0.694$ $R^2 = 0.482$). Considering that the taxonomies used for the semantic analysis concerning geographical places are very complex to construct due to the high probability of error, the result is to be considered good. In

fact, it happens very often that within an article, there are places mentioned that have no relation with the news reported, such as the place where the document was written or the possibility that is ingested a piece of information that is not part of the article (links and titles on the page). Consequently, it is very likely that the semantic engine will report misleading geographical data.

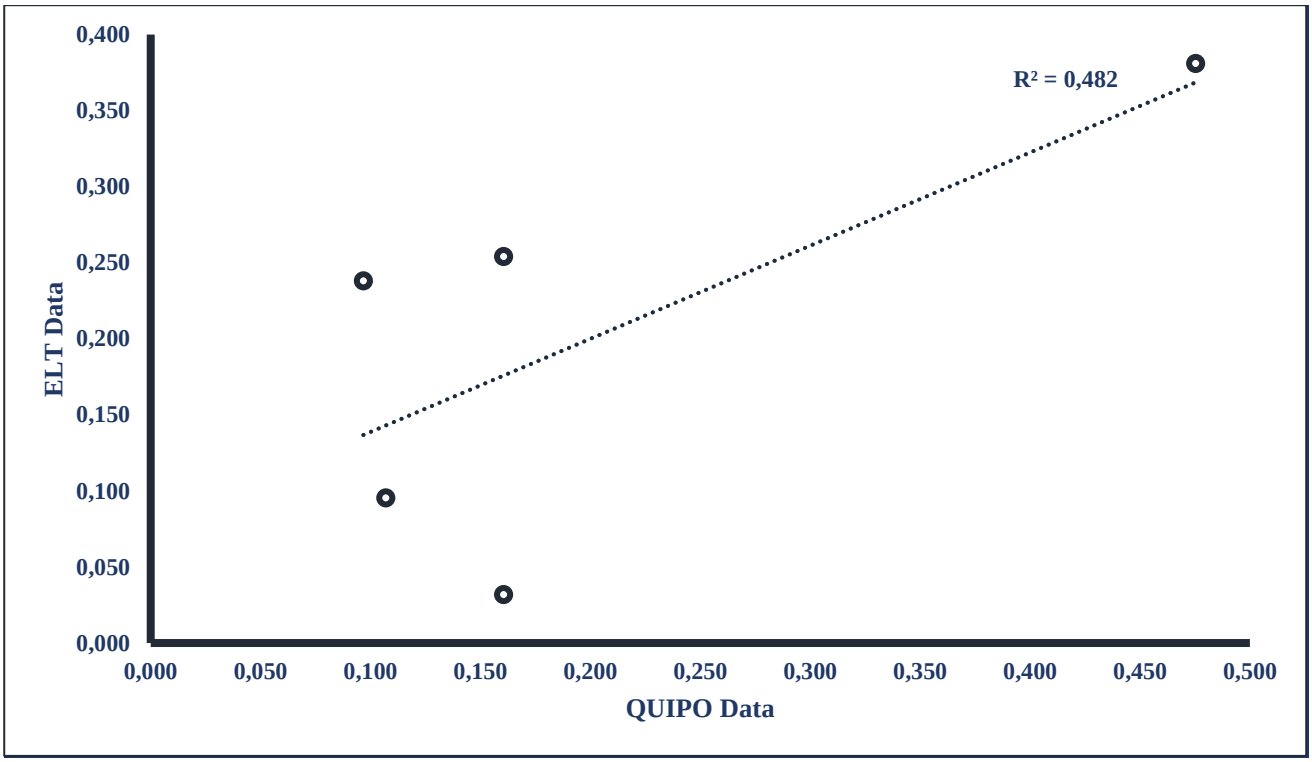


Figure 42: Graphical representation of the correlation between the two matrices (Geography Taxonomy).

Table 2: Capabilities examined and their percentages on total documents (Geography Taxonomy).

Geography	QUIPO Data	ELT Data
U**	0,475	0,381
I****	0,160	0,254
U**	0,160	0,032
U*	0,107	0,095
G******	0,097	0,238

The result obtained is very important as a starting point for the implementation of QUIPO within Elettronica. It was found that in terms of content, the analysis carried out on the platform corresponding to the company's internal intelligence regarding EW capabilities, the company's core business. This made it possible for stakeholders within the company to begin to understand the platform's potential. First of all the management, whose interest is fundamental to carry out the

project, but also the employees of the interested functions whose contribution is fundamental to develop new taxonomies and achieve new goals. In both cases, the validity of the semantic analysis is confirmed by the empirical analysis carried out during these months. Also surprising was the ability to find information not identified by the firm's departments about the companies analysed, which was very useful for an overview of the market in which Elettronica operates.

Summary of the Analysis Process

The process used for the research carried out can be summarised as follows:

1. **Planning:** analysts, timeframe and object of the research.
2. **Data gathering:** data lake construction referring to the rules dictated by the NATO Open Intelligence Handbook in order to have an archive as clean and meaningful as possible.
3. **Analysis:** divided into the first part of data lake analysis and task development and the second part of task analysis and creation of explicative dashboards.
4. **Report production:** report writing and preparation of the knowledge base for the dissemination phase.
5. **Dissemination:** upload the report, articled and graphs on AMICO

In addition to these steps, the processing phase is done entirely by the QUIPO semantic engine without the need for an analyst.

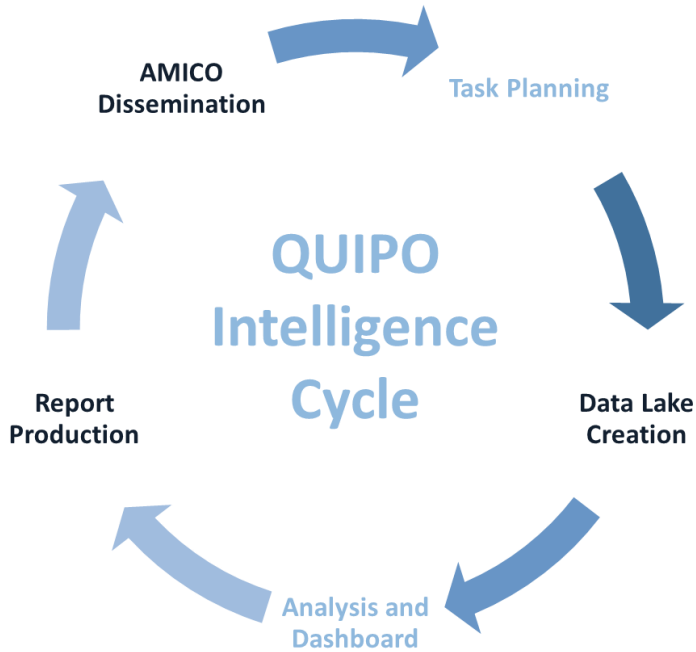


Figure 43: Summary of Quipo Business Case Intelligence Cycle Analysis.

CHAPTER V

Conclusion and Practical Implications

The defence market has always been characterised by its high degree of confidentiality, with most revenues coming from contracts with the military authorities of the various countries who rarely give out information on the agreements. Until a few years ago, the only way for companies to make themselves known was through trade fairs. Today with the evolution of the media, this has changed. Documents and specialized news are now available at short notice, analysts and specialised magazines in the sector publish in-depth analysis and news through their channels, market and financial reports can be easily found, some states are increasingly starting to share information on military contracts and, finally, even job advertisements contain information on the type of technologies being developed. This explains why OSINT analysis is even more essential than in the past for companies operating in this sector. Through an in-depth and methodical study of the news that we find on the web every day, it is possible to obtain insights that allow the company to understand various factors such as the geopolitical landscape, the development of new technologies and the market direction. This is all reinforced by the fact that the defence industry has always been about innovation. In fact, many of the technologies we use today are derived from military projects. Therefore, it is necessary for all companies operating in this sector to use any method to have a competitive advantage over other organisations.

This research is part of the initial phase of a much larger project within Elettronica and Cy4Gate. Their close collaboration has many benefits for both:

- On the one hand, it allows Elettronica to push on the corporate digitisation process to become a data-driven company.
- On the other it allows Cy4Gate to test their products within the defence market, one of the core sectors for the company. The feedback given during these months will be very useful for the technical and commercial development of QUIPO and will enable the expansion of its scope of use through the introduction and incorporation of new services.

The QUIPO implementation project at Elettronica is a clear demonstration of this. The aim is to extend the use of the platform to the core departments of strategy, procurement and sales to make internal processes more efficient and allow employees to spend more time on more delicate and demanding tasks. The success of this project is crucial for the future of the company, which cannot afford to be delayed in its digitisation process, so it is necessary to counteract any obstacles along the way with a clear and transparent implementation phase that involves as many people as possible and

lets them see first-hand how important this tool is and the benefits it brings. The biggest risks are people's resistance to change or the fact that employees do not feel directly involved in digitisation projects because it is outside their scope. Therefore, Elettronica is trying to create several parallel projects linked to different and complementary business functions. In fact, what has been done in this research is as fundamental for procurement as for sales or strategy since all the problems solved and the functions required will be integrated within the platform for all.

This research was the start of the procurement-related project and is a first reading of the potential of QUIPO within Elettronica. In response to the main research question, the study findings were able to reproduce more in-depth results than a previous study conducted entirely on a manual basis. There were many insights into the activities of the companies studied that revealed the existence of relationships and products unknown to the company. It should also be noted that these types of projects aim to develop and educate the platform for very specific semantic analysis. In fact, in the implementation of the platform, the most time-consuming tasks are the creation of the taxonomies and the approach to the platform, in other words, the understanding of the dynamics and functioning of the application. The real added value of this study and the parallel ones is indeed the creation of a shared process within the company that will make these two aspects much simpler and faster in the future. In addition, when it becomes necessary to carry out a supplier scouting again, all the old and new information can be consulted quickly and easily, thus considerably reducing the average analysis time. In fact, the advantage of the platform is the continuity of the analysis. The input given in the data-gathering phase and the semantic analysis of the platform continue to work overtime, guaranteeing the inclusion of the latest news on a given topic. After refining the data collection phase, the automatic alerting function can also be activated. This function inserts documents that respond to certain ingestion rules automatically into the tasks (and AMICO). Once trained in a certain function, QUIPO enables the retrieval of news on a topic in a very short time without the need to conduct a new analysis from scratch. This is very important for Elettronica, as company researches are repeated at intervals as required. Automatically finding articles that the platform has already analysed saves a great deal of time that could be used for other tasks.

In conclusion, in the use of an OSINT Business Intelligence platform, the results state that: through the creation of a structured process that includes the creation of taxonomies appropriate to the reference domain and the correct setting of the intelligence cycle starting from an extensive data lake, the information obtained reflects reality.

From a research standpoint, we have seen how these applications should be developed to give added value to the company in terms of research time. Through the creation of shared processes, it is possible

to significantly reduce the time needed to create taxonomies and to analyse specific searches, while automatic and continuous ingestion and processing allow searches that have already been carried out to be easily retrieved and enriched with new data, reducing the effort needed for a manual search. This allows employees to pay more attention to other important tasks.

A business intelligence platform using AI tools was also analysed, which is very important given the wide acceptance of these types of software. It was seen how some functions within the intelligence cycle can be automated and also the role that the analyst and technician have in guiding the processes of the software.

The research also shows that, due to the very large database, new information and insights can be gained through the use of business intelligence platforms that give a competitive advantage to the company by allowing it to have a detailed and real-time overview of the market situation and help it to make more effective decisions.

Limitation and Future Research

This research had some limitations as it was the initial part of the project mentioned before. First of all, some required functions, such as the integration of some databases, were not available at the time of the study, it will be interesting to see if and to what extent they will improve the intelligence cycle phases of the platform. In addition, the taxonomies, the main part around which the success of such an analysis revolves, were not fully developed in order to extend the boundaries of the research. In fact, the only taxonomy developed internally for semantic analysis was related to electronic warfare capabilities. This resulted in a very subject-specific result that allowed targeted conclusions to be drawn. Future studies must seek to scientifically validate the research conjecture with a broader analysis that aims to compare more complex matrices through a superposition model. It will be necessary to study further the implementation and design processes of BI Platforms that exploit OSINT techniques to create a specific model that allows for easy and rapid application in the field. It is also needed to describe the process of taxonomy construction more analytically by formalising it through the use of validated models for the implementation of information systems artefacts. Finally, it is very important to continue to investigate the results of the application of Business Intelligence software in corporate and governmental environments.

APPENDIX: Conference paper

An Industrial Application of Business Intelligence Approach to the Electronic Defence Sector

Abstract. In the age of digital transformation, the availability of data is growing exponentially leading companies to struggle in processing big data while not missing out useful insights to focus on their business development strategy. In this scenario, always more often, companies are making use of Business Intelligence platforms that could allow them to collect, analyses and disseminate data in real-time to face the dynamic of the market. This paper aims to apply a Business Intelligence approach that adopts OSINT (open-source intelligence) and SOCMINT (Social Media Intelligence) techniques to Defence Electronics Market to analyse how this technology could facilitate Companies decision-making process by providing them with a distinct competitive advantage. In this frame we used QUIPO intelligence platform for an industrial scenario analysis in the Defence Electronics sector. This is an initial research to study the correlation between the experimental OSINT analysis carried out by the intelligence platform and the information based on the internal experience and know-how of the company for the use case study.

Keywords: Business Intelligence & Analytics, Defence, OSINT & SOCMINT, Digital Transformation.

1 Introduction

In Today's world, enterprises are under huge competitive pressures [12], markets are changing faster and faster and companies are subject to Digital Transformation challenges. According to McAfee [19], there is a need to adopt new data sources called Big Data which is defined as extremely large data sets in volume, velocity, variety, and veracity [15]. Big Data could help enterprises collect data and convert it into competitive advantages in the global market [19]. La Valle [14] states that the enterprises using data analytics to convert data into useful information outperform their competitors, by helping them understanding their business more deeply and improving the decision-making process [1]. As a consequence, data-driven enterprises such as Amazon, Microsoft, and Apple in less than ten years have replaced oil companies becoming the biggest firms for capitalization and profits

worldwide[37]. Big Data is also getting bigger as information is coming from a larger number of devices and sources. In the last 2/3 years, humanity has generated more data than ever in its history[38]. It is no wonder that organization are forced to use only a selection of the vast amount of data available to be effective [25]. Considering such relevance of Big Data analysis, many companies must face constant challenges to keep up with digital transformation and become a data-driven company is one of them.

During this Digital Revolution, many companies rely on Business Intelligence (BI) software that can help analysts and management to more easily analyse and understand the primary and secondary data collected, facilitating the extraction of meaningful information for decision making.

Regarding BI Software, the two most important factors for its development and integration within a company are the creation of a meaningful data lake and correct taxonomic rules. According to Gibson [2, 9] without data there is nothing to analyse and collecting data in a wrong way could lead to inadmissible failures; so it is imperative to collect data for a right reason and in the right amount. The constitution of a data lake following precise data gathering rules is essential to increase the performance of the software, to considerably speed up the manual analysis of the collected or inserted documents, to limit the semantic engine errors when analysing unstructured data and, especially in OSINT research, to facilitate the process of assigning the items reliability. Due to the exponential increase in data, it is becoming increasingly difficult to monitor all the information around us. In this scenario, semantic analysis can help process and visualise online information and extract useful knowledge that helps people make informed decisions. For this to happen it is necessary to build new taxonomies each time the domain of interest changes [3]. To develop a new taxonomy, it is also essential to have a strong knowledge base on the specific domain [6]. The creation of structured taxonomic rules is fundamental to increase the software's analysis accuracy and speed up the examination of documents. Analysing the literature, the study of the impact of Business Intelligence is of paramount importance. More research is needed to verify the value of Business Intelligence, from both strategic and managerial aspects [15]. In this frame it will be fundamental to investigate the growing application of AI to BI platforms development [11].

1.1 Purpose

This research is a starting point to study the validation process of a specific “use case” resulting from BI platform implementation, by answering, through the formulation of a conjecture, to the following research question: *"In the frame of the implementation of an OSINT Business Intelligence Platform*

[37] *Le prime 10 aziende al mondo negli ultimi 25 anni, Il sole 24 Ore 2018.* https://lab24.ilsole24ore.com/aziende-top/

[38] A Day in Data, Raconteur. https://www.raconteur.net/infographics/a-day-in-data/

Analysis in the defence market, does the creation of a meaningful Data Lake together with an appropriate Taxonomy structure make the experimental analysis results comparable and coherent to equivalent analysis validated by internal experience and know-how of the firm ?".

In this frame a segmentation of the BI Platform Market is performed in order to give an overview of the positioning of the QUIPO, a BI platform developed by Cy4Gate S.p.A.

1.2Business Intelligence Literature Review

There are different definitions of Business Intelligence; one of the oldest and most exhaustive is Hans Peter Luhn's [16] a researcher at IBM: "an automatic system... developed to disseminate information to the various sections of any industrial, scientific, or government organisation". Business intelligence is the process of transforming data into information, information into knowledge and knowledge into intelligence; with Artificial intelligence software, all this process is cyclical and automated. Karbhari [23] stated in his research that business intelligence is a broad category of application and technologies for collecting, storing, analysing and providing access to data to help the organisation to make a better decision. Chen et Al. [5] and Davenport [7] divide the evolution of BI software into three periods. The 1.0 phase begun in the 70s and concerns primarily the analysis of structured data with a focus on extraction, transformation and loading (ETL) processes to select decision-relevant data from transactional systems and bring them into the proper format for analyses [4]. To store these data, companies mostly used relational database management systems and data warehouses. Data were analysed mainly with the statistical method. With the spread of Internet in the 2000s, user-generated content and web analytics, collected through Web applications [22], are drivers for the second phase of BI. Business Intelligence 2.0 opened new frontiers for data analysis with more complex analysis techniques such as web and text mining and social network analysis [5]. With BI 3.0, we get into digital transformation, with data coming from mobile and IoT devices and the analysis of sensor-generated data [5]. With A Business intelligence software with OSINT (OpenSource Intelligence), SOCMINT (Social Media Intelligence), AI tools, and limitless possibilities given by the Internet, we can access a large set of open information that provides us with a constant updating intelligence to improve decision making at all levels. Literature during these years has proven the benefits of BI, Torres et Al. [26] explain that Business intelligence helps companies in their Digital Transformation and improve organisational outcomes. BI also leads to efficiency improvement, process optimisation, time and cost reduction; taking in exam prior studies, we can also see an improvement in profitability, market share, and customer satisfaction [8]. According to Fink [8], BI assets generate value via two parallel mechanisms, operational and strategic, and these two

dimensions not always are aligned but still lead to value creation. Business Intelligence also help the stakeholders identify the best solution for their firm and develop custom KPIs that best suit them [27].

1.3 OSINT Analysis Literature Review

In a 2011 document published by the Office of the Director of National Intelligence, OSINT was defined as: "intelligence produced from publicly available information that is collected, exploited, and disseminated in a timely manner to an appropriate audience for the purpose of addressing a specific intelligence requirement".

OSINT is distinguished into two data types:

- Open-source data are published by individuals or groups without privacy restrictions. They have little value when taken individually but, when put together, can give a detailed overview of the situation being analysed.
- Open-source information (OSIF), on the other hand, includes all documents that can be obtained legally through request or purchase by citizens. They are usually more in-depth documents.

Through the use of OSINT tools, one can locate, extract and analyse documents from various sources such as blogs, social media, geolocation data, IP addresses, government documents etc. in order to produce intelligence [24]. OSINT analysis has had a lot of luck over the years because it is a quick and efficient way to carry out in-depth analysis [11]. It is widely used by many government agencies and in recent years also at corporate level. In fact, many companies that develop open-source intelligence tools are creating customised software for companies applications. OSINT is used to analyse links between various people, organisations, devices, and many other entities [17]. In the last two decades, there has been a shift from analysing to find hidden information to analysing to find relevant information [21]. The growth of data has created an environment where we are constantly flooded with information. It has become very difficult if not impossible for analysts to conduct manual searches without the help of this kind of software. In this scenario, artificial intelligence and machine learning algorithms are increasingly being used in Business Intelligence projects to partially manage various stages of the intelligence cycle. Although there are various interpretations of the intelligence cycle, which is commonly defined in 5 steps[39]:

- **Planning** the purpose and the activities of the analysis
- **Data Collection** to create the Data Lake
- **Processing** the documents gathered in the previous step (in AI-powered software, this stage is fully automated)

[39] The intelligence process/cycle (Source: "Joint Publication 2-0, Joint Intelligence". Defense Technical Information Center (DTIC). Department of Defense. February 2013)

- **Analysis** of the documents
- **Dissemination** of the intelligence (Report or Graphic Dashboard)

These stages include the acquisition and validation of information, the identification of the value of the information and the distribution of the information to the client [2, 10, 18].

1.4 Methodology

In the first part of the research, we employed data from internal company databases and companies' websites to identify competitors to the QUIPO platform and to segment the market. In relation to the research objectives, we tried to segment the market through the construction of a two-variable positioning map. We examined a sample of 14[40] companies operating in the business intelligence market and we positioned them on the map according to their focus on the use of OSINT data and the possibility of customising taxonomies. In the second part, we used an applied research method to empirically validate the use of the BI platform when applied to a specific "use case study". According to Kothari [13], Applied research *"aims at finding a solution for an immediate problem facing a society, or an industrial/business organisation, whereas fundamental research is mainly concerned with generalisations and with the formulation of a theory"*.

Specifically, we designed a use case within the company following the five phases of the intelligence cycle, paying particular attention to the data gathering phases for the data lake construction and the taxonomy development. Once the analysis was complete, we compared the data collected by the platform with the data from in-house intelligence and expertise to find a match between the experimental analysis performed by the QUIPO platform and the company's knowledge.

2 Business Intelligence Market Segmentation

The two variables for the market segmentation have been selected by studying the literature on OSINT techniques and Business Intelligence. The dimensions relate to:

- **Focus on OSINT Analysis**, representing the degree to which the platform focuses on collecting and analysing external data available on the Internet.
- **Focus on TAXONOMY Customization**, represents the degree to which the platform focuses on taxonomies customisation according to the application cases.

Analysing the results deriving from the positioning map (see fig. 1), we noticed the presence of four market niches corresponding to the quadrants outlined by the map plan.

[40] Note: Company names are not disclosed for confidentiality

Specific In-House. includes platforms that have a rigid semantic structure and that refer to internal non open-source data. These software collect data provided by the developing company or customer databases. They are used for specific purposes within the organisation that cannot be extended to other business functions. Since these are programs developed to solve a specific problem, the taxonomy structure is very rigid, focusing only on their scope of application such as CRM applications. The advantages of these systems come from their specialisation. As they are used for very specific tasks, they require less time to be implemented and benefit from more established taxonomies in their field of application.

Multi-Purpose In-House. includes platforms whose primary purpose is to use in-house data, but unlike the previous group, they have the ability to modify taxonomies depending on their field of application. They are used for various purposes and are very adaptable to various markets and various business functions, which is precisely the reason why they need to adapt the taxonomy structure depending on the type of analysis required. The advantages of these software lie in their greater elasticity of use, making them attractive in many sectors and for the most various problems.

Specific OSINT. includes platforms that focus on the collection and analysis of open-source data and have a limited ability to customise taxonomies for semantic analysis. Like the programs in the first quadrant, these software are also developed for very specific functions within organisations that require analysis of externally sourced data. The taxonomies are often fixed due to the high specificity of the task, even though platforms dedicated to OSINT research are very widely developed in their scope (e.g. this group includes cyber intelligence software). These platforms are able to manage the data they collect more quickly and accurately because of their knowledge of the primary sources from which they collect information and their highly developed semantic analysis around the main function.

Multi-Purpose OSINT. In this niche there are the Business Intelligence platforms that make open-source data collection their core business and allow developers and designers to create new taxonomies for semantic analysis depending on the platform's field of application. The tools in this quadrant have various features that make them suitable for various tasks, and their flexibility makes them usable by multiple functions within the company. Although they focus on OSINT research,

these software are, in most cases, capable of handling internal data to combine the benefits of external and internal research. Since the strength of these platforms is their adaptability to various environments, they require ad hoc taxonomic development to reflect the client's needs. While implementation can be slightly time-consuming, once embedded in business operations, it provides a high degree of flexibility at all levels.

While aiming to help organisations in the collection, management, and analysis of data, BI platforms are very different from each other and can be used for different purposes. QUIPO, the BI Platform under study, is positioned within the Multi-Purpose OSINT industry market with its related key features to be considered for the adoption.

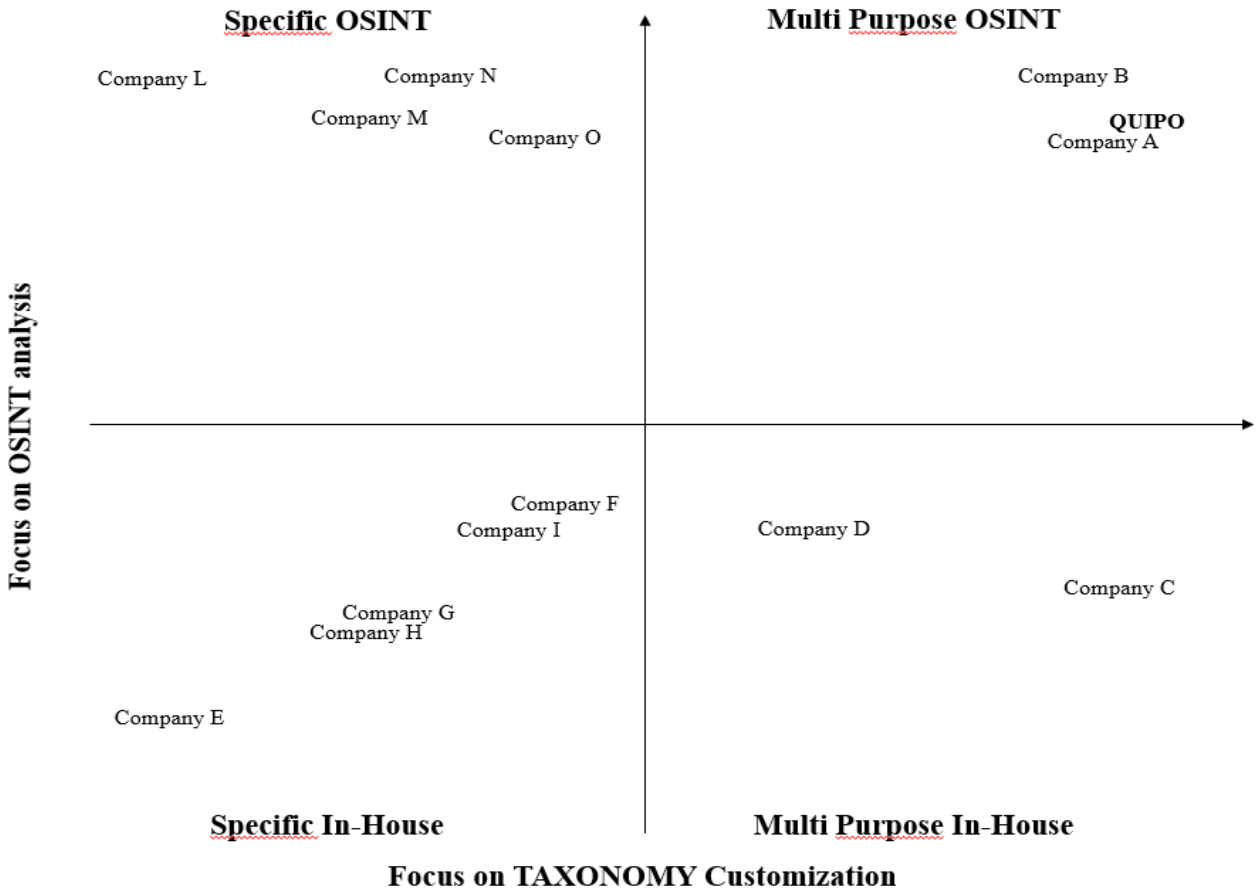


Fig. 1. Business Intelligence Positioning Map

3 QUIPO Application Case

The objective is to evaluate the benefits and opportunities that OSINT analysis can bring to the Company organisation by creating several use cases. This study is part of the company's project to integrate QUIPO into the procurement function in order to achieve the following objectives:

- Improvement of the effectiveness of the suppliers scouting process
- Simplification of the suppliers "due diligence" (actual or potential).
- Technological development trend of specific product categories, by identifing leading edge or out-of-market suppliers.

During our research, we followed the intelligence cycle described above by focusing on product capability to determine if the information we found during the analysis of the documents collected by the platform were correlated to the information held in the company. In the use case, we have examined three potential suppliers, scouting their capabilities to explore potential collaboration opportunities.

3.1 QUIPO Structure

QUIPO consists of three main modules linked together by tools that exploit artificial intelligence and machine learning algorithms to analyse documents:

- The first module consists of a non-relational database with the task of indexing documents from internal and external sources. An archive is then created containing all the documents, enriched with the first metadata that establishes their uniqueness and allows the internal search engine to retrieve them quickly. After being saved and indexed in the database, the data is analysed by the platform's semantic engine or, in the case of images, by image and face recognition tools. In this phase, the documents are assigned additional metadata, which is very important for the representation and analysis by the user.

- The second module deals with the representation of the documents collected in the database using the metadata produced in the previous phases. Through the aggregation of this data, the module allows the creation of dashboards and graphs with which we have obtained the data to formulate our conjecture.

- The third module is the platform's GUI, which allows the end-user to manage all the tasks necessary to carry out the analysis. It can be considered as the point of contact between the backend tools and the analyst. Closely linked to this module is the data dissemination platform that allows the textual and graphical display of the work done.

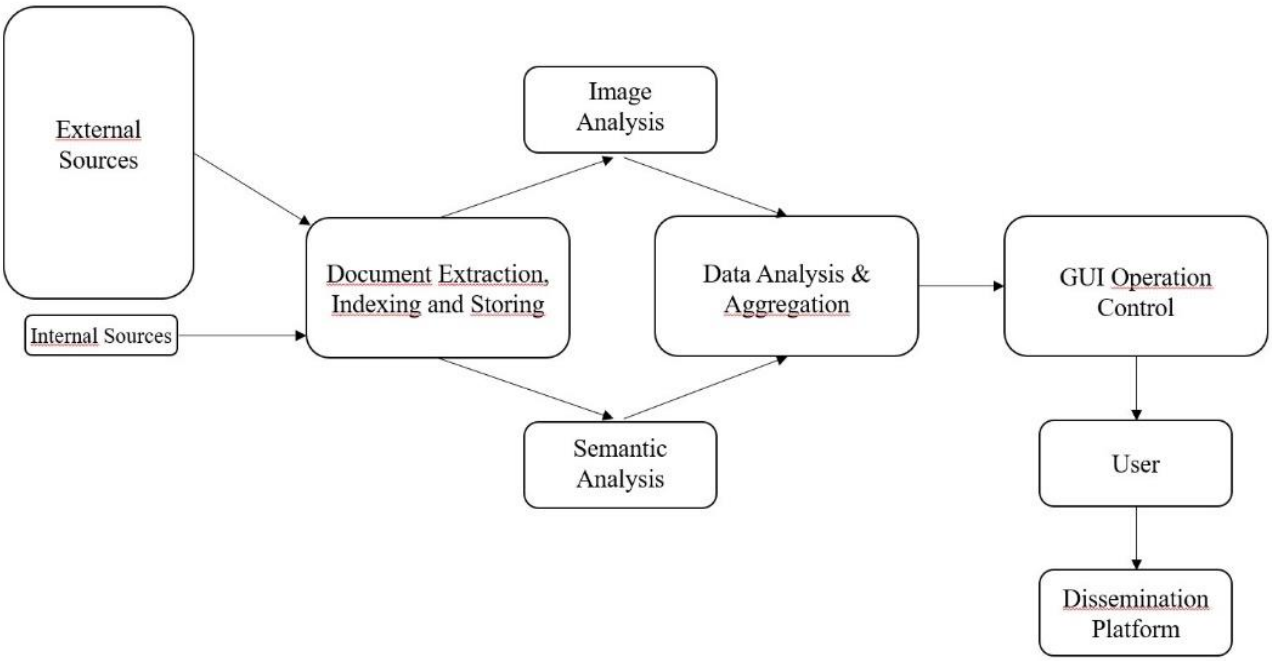


Fig. 2. Visual representation of Quipo Structure

3.2 Data Lake Construction

The data-gathering phase is one of the most delicate moments of this kind of analysis. In this step it is essential to perform a first skimming of the collected articles, doing searches as relevant and intelligent as possible. In fact, the risk is to conduct summary searches for keywords with very broad meanings and find the archive filled with tens of thousands of documents that are completely useless for the analysis by increasing the analyst's difficulty in recognising the articles that must be included in the task. At this stage, it is necessary to have a thorough knowledge of the platform and its reasoning method for data ingestion. In this way, it is very easy to make more detailed research that allows having a clearer overview of the phenomenon observed at first glance.

In the construction of the data lake, we source news from major defence websites and magazines, Google News, Europe Media Monitor, the press releases of the analysed companies and related organisations, ministerial websites, social media (Twitter and Facebook), the press release, archives of documents already present in the platform from previous searches.

In choosing the sources and methods of data gathering applied, we respected the NATO Open-Source Intelligence Handbook guidelines[41] [2, 9]:

1. The authority of the source
2. The accuracy (by validating it against other sources)

[41] NATO (2001) NATO open-source intelligence handbook.

3. The objectivity of the source
4. The currency (i.e., the provision of a timestamp for publication and the presence of an author)
5. The coverage (the degree of relevancy)

The data gathering process was conducted by searching for very precise keywords that included the name of the companies under study, their products or geopolitical concepts related to the research. In three months, we had 30,000 articles in our specific archive, to which we added other archives containing news from magazines and defence fairs on the major players in the market (500,000), the press release (200,000), news on investment banks and related companies (20,000) and geopolitical information (1,500,000). Our data lake, at the time of analysis, consisted of 2,250,000 articles analysed by the platform. The next step was to process them by filtering the documents within the data lake according to taxonomic concepts and the companies we were interested in. In the selection of the significant articles, the same rules for the identification of sources were applied. It was necessary to repeat the process because some media such as Google News, Facebook and Twitter require more careful analysis to avoid the ingestion of unreliable news. After filtering the articles of interest, reliability (low, medium, high) was assigned to the most interesting ones. Finally, an archive containing all these documents was created and compared with the attainable information in our possession.

3.3 Taxonomy Creation

A taxonomy is a semantic hierarchy that organises concepts by is-a relations, which exhibits the capability of improving many NLP task [6]. Its construction is one of the most important operations in the implementation of software capable of semantic analysis. A taxonomy is composed of rules that allow the semantic engine to understand a given taxonomic entity within a document. A taxonomy development depends on the domain that has to be analysed and the user requirement, who may have different needs according to his objective. A taxonomy can be created using two types of approach:

- Categorisation: a deductive method in which the starting point is a general concept to arrive at specific individual entities. In this approach, a semantic tree is constructed with branches comprising subcategories of more general entities. Categories are generated first by formulating their conceptual descriptions and then classifying objects according to the descriptions.
- Extraction: The inductive method consists of identifying, aggregating, and normalising specific semantic entities to arrive at more generic concepts.

After a knowledge analysis on the reference domain and the choice of the approach to be used, rules are developed and associated to a single domain. These rules allow the platform to be able to distinguish ambiguous words and concepts within a context, making data analysis more precise.
In our analysis, we needed to initially develop an entry-level taxonomy of Electronic Warfare (EW). Our objective was to contextualise the possible suppliers within the market, identifying their capabilities and heritage. Therefore, a categorisation approach was used to create a semantic tree composed of general concepts that helped us understand the scope in which a company operates and specific entities to study the differences between players. The taxonomy was validated using a benchmark test to establish the correctness of taxonomic rules. Through the use of already analysed libraries, the test can give a percentage on the accuracy of the taxonomy. It is based on an iterative process that allows the rules to be evaluated and modified according to the result.

3.4 Proceedings and Results

For this research, we focused on the first cluster of taxonomies available in the procurement project concerning firm Electronic Warfare capabilities in the defence sector, by examining three firms that belongs to the same capability cluster. We inspected several reports available in the company and compiled a file containing information about the contracts, merge & acquisition activities, partnership and joint ventures of the organisations under study. The file containing 170 occurrences was then validated by the company functions. At the same time, we analysed the documents contained in the QUIPO data lake, resulting in a final archive of 932 items. We selected the ten taxonomic concepts concerning Electronic Warfare capabilities within the documents and divided into clusters all the articles for each specific concept. We then calculated the percentage of each cluster with respect to the total number of documents in the archive to normalise the data. Next, we studied the existing correlation between the two matrices through the following formula for the calculation of the Pearson Correlation Coefficient:

$$r = \frac{\sum[(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{\sum(x_i - \bar{x})^2 * \sum(y_i - \bar{y})^2}}$$

The results (see fig.3, table 1) indicate a highly positive correlation [20] between the two matrices (r = 0.927, R2 = 0.86). That denotes a correspondence between analysis based solely on OSINT techniques by using a business intelligence platform and data validated by firm's experts. An empirical correlation also emerged when compiling the report, noting a sufficient match between the concepts extracted from the platform and those present in the reports analysed.

In light of the following results, we can formulate the following conjecture:

"In the development of a Business Intelligence platform, which uses analysis techniques based on the OSINT intelligence cycle, the creation of a data lake following precise rules that point to correct use of sources together with the creation of an appropriate taxonomy to the specific domain, allows the number of occurrences on a given entity to be very close to reality".

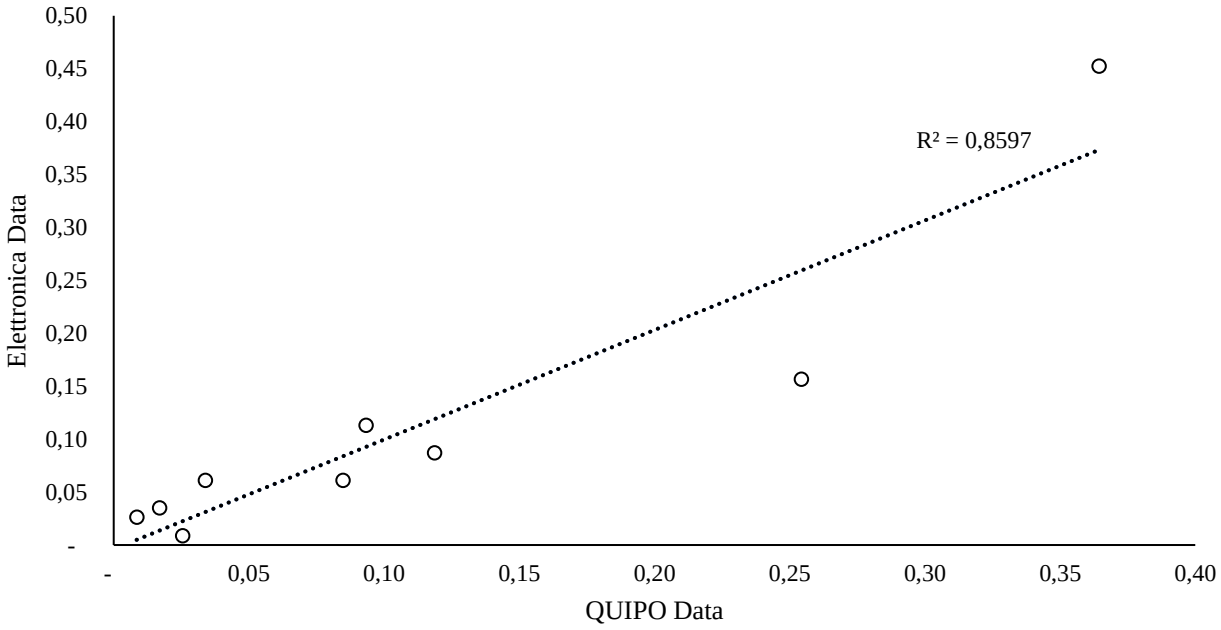


Fig. 3. Graphical representation of the correlation between the two matrices.

Table 1. Capabilities examined and their percentages on total documents.

Capabilities	QUIPO Data	ELT Data
C4ISR	0,36	0,45
DIRCM	0,09	0,11
Training	0,25	0,16
Customer Support	0,03	0,06
RECM	0,01	0,03
COMINT	0,12	0,09
RESM	0,02	0,03
SIGINT	0,08	0,06
ELINT	0,03	0,01

4 Conclusion and Future Work

In the following research, we have described a technique for the construction of a data lake and the creation of a taxonomy in a specific domain through the use of a practical case. We have also seen how the information extracted by the BI platform in the frame of the use case analysis is correlated to the information available and validated by company expertise. This study is only the starting point for a broader project, aiming to expand the platform's scope through the creation of various taxonomies in the procurement domain and to continue research into the application of Business Intelligence platforms powered by AI tools in areas not yet studied.

References

1. Ahi, A., G. Baronchelli, O. Kuivalainen, and M. Piantoni. International market entry: How do small and medium-sized enterprises make decisions? Journal of International Marketing 25 (1): 1–21 (2017).
2. Babak, Akhgar, P. Saskia, Bayerl, Fraser, Sampson, Open-Source Intelligence Investigation From Strategy to Implementation. 1st edn. Springer International Publishing AG (2016).
3. Carrion, B., Onorati, T., Díaz, P. et al. A taxonomy generation tool for semantic visual analysis of large corpus of documents. Multimed Tools Appl 78, 32919–32937 https://doi.org/10.1007/s11042-019-07880-y (2019).
4. Chaudhuri, S., U. Dayal, and V. Narasayya: An overview of business intelligence technology. Communications of the ACM 54 (8):88. https://doi.org/10.1145/1978542.1978562 (2011).
5. Chen, H., R.H.L. Chiang, and C.S. Veda: Business intelligence and analytics: from big data to big impact. MIS Quarterly 36 (4): 1165–1188. https://doi.org/10.2307/41703503 (2012).
6. Chengyu Wang, Xiaofeng He, Aoying Zhou: A Short Survey on Taxonomy Learning from Text Corpora: Issues, Resources and Recent Advances. Conference on Empiri-cal Methods in Natural Language Processing, pages 1190–1203 (2017).
7. Davenport, T.H. Analytics 3.0: In the new era, big data will power consumer products and services. Harward Business Review 91: 64–72 (2013).
8. Fink,L. Yogev,N. Even: A Business intelligence and organisational learning: An empirical investigation of value creation processes. Information & Management 54 38–56 (2017).
9. Gibson H., Ramwell S., Day T. Analysis, Interpretation and Validation of Open Source Data. In: Akhgar B., Bayerl P., Sampson F. (eds) Open Source Intelligence Investigation. Advanced Sciences and Technologies for Security Applications. Springer, Cham. https://doi.org/10.1007/978-3-319-47671-1_7 (2016).
10. Heather J. Williams, Ilana Blum: Defining Second Generation Open Source Intelligence (OSINT) for the Defense Enterprise. RAND National defense Research Institute ISBN: 978-0-8330-9883-2 (2018).

11. João Rafael Gonçalves Evangelista, Renato José Sassi, Márcio Romero & Domingos Napolitano: Systematic Literature Review to Investigate the Application of Open Source Intelligence (OSINT) with Artificial Intelligence, Journal of Applied Security Research, DOI: 10.1080/19361610.2020.1761737 (2020).
12. Knight, Gary A., and Peter W. Liesch: Internationalisation: From incremental to born global. Journal of World Business 51 (1): 93–102 (2016).
13. Kothari, C. R., Research Methodology, Methods and Techniques (2nd ed., pp. 109-110). New Delhi: New Age Inter- national (P) Limited (2008).
14. LaValle, S., E. Lesser, R. Shockley, M.S. Hopkins, and N. Kruschwitz: Big data, analytics and the path from insights to value. MIT sloan management review 52 (2): 21 (2011).
15. Liang, T.P., and Y.H. Liu: Research landscape of business intelligence and big data analytics: A bibliometrics study. Expert Systems with Applications 111: 2–10 (2018).
16. Luhn H.P., A business intelligence system IBM Journal of Research and https://doi.org/10.1147/rd.24.0314 (1958).
17. Maciołek, P., & Dobrowolski, G.: Cluo: Web-scale text mining system for open source intelligence purposes. Computer Science, 14(1), 45–62. https://doi.org/10.7494/ csci.2013.14.1.45 (2013).
18. Marzell, Laurence: OSINT as Part of the Strategic National Security Landscape. 10.1007/978-3-319-47671-1_4 (2016).
19. McAfee, A., E. Brynjolfsson, T.H. Davenport, D.J. Patil, and D. Barton: Big data: The management revolution. Harvard Business Review 90 (10): 60–68 (2012).
20. M.M Mukaka, Statistics Corner: A guide to appropriate use of Correlation coefficient in medical research Malawi Medical Journal; 24(3): 69-71 September (2012).
21. Nicart, E., Zanuttini, B., Gilbert, H., Grilheres, B., & Praca, F.: Building document treatment chains using reinforcement learning and intuitive feedback [Paper presentation]. 2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI) (pp. 635–639). IEEE. https://doi.org/10.1109/ICTAI.2016.0102 (2016).
22. O'Reilly, T.: What Is Web 2.0? Design Patterns and Business Models for the Next Generation of Software. https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html (2020).
23. P. Karbhari, "SOA trends: Service oriented business intelligence," White Paper. [Online]. Available: www.technologyexecutivesclub.com (2006).
24. Quick, D., & Choo, K. K. R.: Digital forensic intelligence: Data subsets and Open Source Intelligence (DFINTþOSINT): A timely and cohesive mix. Future Generation Computer Systems, 78, 558–567. https://doi.org/10.1016/j.future.2016.12.032 (2018).

25. Thomas C. Redman, Does your Company Know What to Do with All Its Data? Harvard Business review. https://hbr.org/2017/06/does-your-company-know-what-to-do-with-all-its-data (2017).
26. Torres,R. Sidorova,,A. Jones,M: Enabling firm performance through business intelligence and analytics: A dynamic capabilities perspective. Information & Management 1 (2019) 1–18 (2018).
27. Tripathi, Anuj and Bagga, Teena, Leading Business Intelligence (BI) Solutions and Market Trends: Proceedings of the International Conference on Innovative Computing & Communications (ICICC) 2020, Available at SSRN: https://ssrn.com/abstract=3568414 or http://dx.doi.org/10.2139/ssrn.3568414 (April 4, 2020).

REFERENCES

- Admiral William Studeman, Teaching the Giant to Dance: Contradictions and Opportunities in Open Source Within the Intelligence Community, Open Source Solutions, Inc., December 1992 https://fas.org/irp/fbis/studem.html (1992).
- Ahi, A., G. Baronchelli, O. Kuivalainen, and M. Piantoni. International market entry: How do small and medium-sized enterprises make decisions? Journal of International Marketing 25 (1): 1–21, (2017).
- Babak, Akhgar, P. Saskia, Bayerl, Fraser, Sampson, Open-Source Intelligence Investigation From Strategy to Implementation. 1st edn. Springer International Publishing AG (2016).
- B. O. Hostmann, "Business intelligence trends, with Gartner analyst". [Online]. Available: http://searchdatamanagement.techtarget.com/ (2007).
- Carrion, B., Onorati, T., Díaz, P. et al. A taxonomy generation tool for semantic visual analysis of large corpus of documents. Multimed Tools Appl 78, 32919–32937 https://doi.org/10.1007/s11042-019-07880-y (2019).
- Chaudhuri, S., U. Dayal, and V. Narasayya. 2011. An overview of business intelligence technology. Communications of the ACM 54 (8):88. https://doi.org/10.1145/1978542.1978562, (2011).
- Chen, H., R.H.L. Chiang, and C.S. Veda. Business intelligence and analytics: from big data to big impact. MIS Quarterly 36 (4): 1165–1188. https://doi.org/10.2307/41703503 (2012).
- Chengyu Wang, Xiaofeng He, Aoying Zhou: A Short Survey on Taxonomy Learning from Text Corpora: Issues, Resources and Recent Advances. Conference on Empiri-cal Methods in Natural Language Processing, pages 1190–1203 (2017).
- Davenport, T.H. Analytics 3.0: In the new era, big data will power consumer products and services. Harward Business Review 91: 64–72, (2013).
- Erinc Dikici, Gerhard Backfried, Jurgen Riedler. The SAIL LABS Media Mining Indexer and the CAVA Framework. INTERSPEECH 2019: Show & Tell Contribution 4630 – 4631, (2019).

- Fink,L. Yogev,N. Even,A(2017). Business intelligence and organisational learning: An empirical investigation of value creation processes. Information & Management 54 38–56, (2017).
- Gayane Shalunts, Gerhard Backfried, Katja Prinz. Sentiment Analysis of German Social Media Data for Natural Disasters, 11th International ISCRAM Conference – University Park, Pennsylvania, USA, May (2014).
- Geissbauer Reinhard, Vedso J., Shrauf S., PWC, 'Industry 4.0: Building the digital enterprise', (2016).
- Gerhard Backfried, Christian Schmidt, Mark Pfeiffer, Gerald Quirchmayr, Johannes Göllner. Open-Source Intelligence for Traditional - and Social Media Sources, The 10th International Conference on e-Business (iNCEB 2015).
- Gibson H., Ramwell S., Day T. Analysis, Interpretation and Validation of Open Source Data. In: Akhgar B., Bayerl P., Sampson F. (eds) Open Source Intelligence Investigation. Advanced Sciences and Technologies for Security Applications. Springer, Cham. https://doi.org/10.1007/978-3-319-47671-1_7, (2016).
- Heather J. Williams, Ilana Blum: Defining Second Generation Open Source Intelligence (OSINT) for the Defense Enterprise. RAND National defense Research Institute ISBN: 978-0-8330-9883-2 (2018).
- Intelligence Reform and Terrorism Prevention Act of 2004, Sec. 1052, pp. 166–167 https://www.govinfo.gov/content/pkg/PLAW-108publ458/pdf/PLAW-108publ458.pdf, (2004).
- Kothari, C. R., Research Methodology, Methods and Techniques (2nd ed., pp. 109-110). New Delhi: New Age Inter- national (P) Limited (2008).
- J. Niles Riddel, Remarks at the First International Symposium, 'National Security and National Competitiveness: Open Source Solutions,' Open Source Solutions, Inc., December 2, (1992).
- Jaiswal, and M. Mita, Management Information Systems, Oxford University Press, pp. 382-398, (2004).
- João Rafael Gonçalves Evangelista, Renato José Sassi, Márcio Romero & Domingos Napolitano: Systematic Literature Review to Investigate the Application of Open Source

Intelligence (OSINT) with Artificial Intelligence, Journal of Applied Security Research, DOI: 10.1080/19361610.2020.1761737, (2020).

- Joseph E. Roop, Foreign Broadcast Information Service History. Part I: 1941–1947, Central Intelligence Agency, April, p.7, 8, 50, (1969).
- Knight, Gary A., and Peter W. Liesch. Internationalisation: From incremental to born global. Journal of World Business 51 (1): 93–102, (2016).
- LaValle, S., E. Lesser, R. Shockley, M.S. Hopkins, and N. Kruschwitz. Big data, analytics and the path from insights to value. MIT sloan management review 52 (2): 21, (2011).
- Liang, T.P., and Y.H. Liu. Research landscape of business intelligence and big data analytics: A bibliometrics study. Expert Systems with Applications 111: 2–10, (2018).
- Maciołek, P., & Dobrowolski, G. Cluo: Web-scale text mining system for open source intelligence purposes. Computer Science, 14(1), 45–62. https://doi.org/10.7494/csci.2013.14.1.45, (2013).
- Maltego Technology GMBH. How Maltego helps accelerate the i investigation of insider threat alerts provided by User Behavior Analysis (UBA) systems, (2021).
- McAfee, A., E. Brynjolfsson, T.H. Davenport, D.J. Patil, and D. Barton. Big data: The management revolution. Harvard Business Review 90 (10): 60–68, (2012).
- M.M Mukaka, Statistics Corner: A guide to appropriate use of Correlation coefficient in medical research Malawi Medical Journal; 24(3): 69-71 September (2012).
- Munn, L. Seeing with software : Palantir and the regulation of life. Studies In Control Societies, 2(1). Retrieved from https://studiesincontrolsocieties.org/seeing-with-software/, (2017).
- NoorUl Aina, Giovanni Vaia, William H. DeLone, Mehwish Waheed: "Two decades of research on business intelligence system adoption, utilisation and success, A systematic literature review, Decision Support Systems 125 113113, (2019).
- Nicart, E., Zanuttini, B., Gilbert, H., Grilheres, B., & Praca, F. Building document treatment chains using reinforcement learning and intuitive feedback [Paper presentation]. 2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI) (pp. 635–639). IEEE https://doi.org/10.1109/ICTAI.2016.0102, (2016).

- O'Reilly, T. What Is Web 2.0? Design Patterns and Business Models for the Next Generation of Software. https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html, (2020).
- P. Karbhari, "SOA trends: Service oriented business intelligence," White Paper. [Online]. Available: www.technologyexecutivesclub.com (2006).
- Quick, D., & Choo, K. K. R. Digital forensic intelligence: Data subsets and Open Source Intelligence (DFINTþOSINT): A timely and cohesive mix. Future Generation Computer Systems, 78, 558–567. https://doi.org/10.1016/j.future.2016.12.032, (2018).
- Renu Jakhar and Chhavi Krishna, Business Intelligence: As a Strategic Tool for Organization Development, ANWESH: International Journal of Management and Information Technology 5 (1) 44-46, March (2020).
- Samer Al khateeb, Nitin Agarwal. Social cyber forensics: leveraging open source information and social network analysis to advance cyber security informatics, Springer Science+Business Media, LLC, part of Springer Nature, Published online: 19 July (2019).
- Sumera Ahmad; Miskon, S.; Alabdan, R.; Tlili, I. Towards Sustainable Textile and Apparel Industry: Exploring the Role of Business Intelligence Systems in the Era of Industry 4.0. Sustainability, 12, 2632. https://doi.org/10.3390/su12072632, (2020).
- Tatiana Genzorova, Tatiana Corejova, Natalia Stalmasekova, How digital transformation can influence business model, Case study for transport industry, Transportation Research Procedia, Volume 40, Pages 1053-1058, ISSN 2352-1465, https://doi.org/10.1016/j.trpro.2019.07.147 (2019).
- Ting-Peng Liang, Yu-His Liu: "Research Landscape of Business Intelligence and Big Data analytics: A bibliometrics study", Science Direct, Expert Systems With Applications 111 2–10, http://www.elsevier.com/locate/eswa, (2018).
- The European House Ambrosetti & Elettronica Group, a cura di M.C. Carrozza, M. Sideri, A.S. Vicentelli, 'Digital Transformation: Nuovi confini, crescita e sicurezza del paese', (2017).
- The European House Ambrosetti & Elettronica Group, 'Geopolitica del Digitale: Nuovi confini, crescita e sicurezza del paese, II Edizione', Novembre (2018).
- Thomas C. Redman, Does your Company Know What to Do with All Its Data? Harvard Business review, (2017).

- Torres,R. Sidorova,,A. Jones,M(2018). Enabling firm performance through business intelligence and analytics: A dynamic capabilities perspective. Information & Management 1 1–18 (2019).
- Tripathi, Anuj and Bagga, Teena, Leading Business Intelligence (BI) Solutions and Market Trends (April 4, 2020). Proceedings of the International Conference on Innovative Computing & Communications (ICICC), Available at SSRN: https://ssrn.com/abstract=3568414 or http://dx.doi.org/10.2139/ssrn.3568414 , (2020).
- Z. Sheng, "Measuring index system and method of organisational intelligence," School of Economics, University of Jinan, (2005).

SITOGRAPHY

- A Day in Data, Raconteur. https://www.raconteur.net/infographics/a-day-in-data/.
- Airbus believes its AI can eliminate flight delays: https://venturebeat.com/2018/10/22/airbus-believes-its-ai-can-eliminate-flight-delays/.
- Alberto Nasciuti ci parla di Kpi6: Dati e insights per la consumer market research (startup.info) https://startup.info/it/alberto-nasciuti-kpi6/
- Application Areas - HENSOLDT Analytics (hensoldt-analytics.com) https://www.hensoldt-analytics.com/application-areas/
- Business intelligence solution penetration by industry worldwide, as of 2018, Statista. https://www.statista.com/statistics/873143/business-intelligence-penetration-by-industry/
- Cloud business intelligence (BI) importance worldwide as of 2020, by function, Statista. https://www.statista.com/statistics/1016209/business-intelligence-penetration-by-function/.
- Come è cambiato il traffico su Internet durante l'emergenza coronavirus (agi.it) https://www.agi.it/innovazione/news/2020-04-29/aumento-traffico-dati-internet-coronavirus-lockdown-8471456/.
- Does Your Company Know What to Do with All Its Data? HBS 2017. https://hbr.org/2017/06/does-your-company-know-what-to-do-with-all-its-data
- Digital Economy Compass 2020 | Statista https://www.statista.com/study/83121/digital-economy-compass/.
- Digital Transformation opens up the High Road to GDPR: https://towardsdatascience.com/digital-transformation-opens-up-the-high-road-to-gdpr-8a1b11788e19
- Gesher, Ari (2010-03-08). "Friction in Human-Computer Symbiosis: Kasparov on Chess". Palantir tech. Archived from the original on 2012-04-14. Retrieved 2012-01-30. https://web.archive.org/web/20120414224548/http://blog.palantirtech.com/2010/03/08/friction-in-human-computer-symbiosis-kasparov-on-chess/.
- Global artificial intelligence market revenues 2024 | Statista https://www.statista.com/statistics/694638/worldwide-cognitive-and-artificial-intelligence-revenues/.

- Il traffico internet è cresciuto con la pandemia, la rete ha tenuto grazie al cloud - Il Sole 24 ORE https://www.ilsole24ore.com/art/il-traffico-internet-e-cresciuto-la-pandemia-rete-ha-tenuto-grazie-cloud-ADqZk1HB.
- HENSOLDT Acquires SAIL LABS - European Security & Defence (euro-sd.com) https://euro-sd.com/2021/02/news/cyber-coms/20935/hensoldt-sail-labs/
- HENSOLDT Analytics System - HENSOLDT Analytics (hensoldt-analytics.com) https://www.hensoldt-analytics.com/hensoldt-analytics-system/
- Homepage - Maltego https://www.maltego.com/.
- Le prime 10 aziende al mondo negli ultimi 25 anni, Il sole 24 Ore 2018. https://lab24.ilsole24ore.com/aziende-top/.
- Le terre rare, ovvero la lotta per l'egemonia e l'Europa evanescente: https://www.econopoly.ilsole24ore.com/2019/07/11/terre-rare-egemonia/?refresh_ce=1
- Leaked Palantir Doc Reveals Uses, Specific Functions And Key Clients: https://techcrunch.com/2015/01/11/leaked-palantir-doc-reveals-uses-specific-functions-and-key-clients/.
- Leaked Palantir S-1 shows company has 125 customers after 17 years: https://techcrunch.com/2020/08/21/leaked-palantir-s-1-shows-company-has-125-customers-after-17-years/.
- Leveraging Big Data In A Sales Organization, Frobes 2021. https://www.forbes.com/sites/forbesbusinessdevelopmentcouncil/2021/01/27/leveraging-big-data-in-a-sales-organization/?sh=5457d55a6cd3.
- Nei prossimi 5 anni cambieranno 6 lavori su 10. Ecco come: https://www.ilsole24ore.com/art/ecco-come-prossimi-cinque-anni-cambieranno-sei-lavori-dieci--AB67zUaB (Sole 24 Ore).
- Palantir Apollo: Powering SaaS where no SaaS has gone before: https://blog.palantir.com/palantir-apollo-powering-saas-where-no-saas-has-gone-before-7be3e565c379.
- Palantir Homepage - https://www.palantir.com.

- Palantir Technologies financial data: https://finance.yahoo.com/quote/PLTR/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAANgPvv7h2L2oyGpWrIbXs9f49HEh-ikemvS32qbZBFtk4XHkN-dap0m_xwECDItvKsuoxMe3m5-X2rjbjqcgTVo5mIPlNXhQTLCLxN4KGjDYh1u4MpPnw-OSdNVTA4bNdqyDz4zGP_-xf3ibZ_8KDqQbSsU9igEY9g1JQBhaXURd.

- Palantir Technologies Inc. ANNUAL REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1 0001193125-21-060650 (d18rn0p25nwr6d.cloudfront.net).

- Peter Waldman, Lizette Chapman, Jordan Robertson, Peter Thiel's data-mining company is using War on Terror tools to track American citizens. The scary thing? Palantir is desperate for new customers, Bloomberg Businessweek, (2018) https://www.bloomberg.com/features/2018-palantir-peter-thiel/.

- Size of the business intelligence and analytics software application market worldwide, from 2019 to 2024, Statista. https://www.statista.com/statistics/590054/worldwide-business-analytics-software-vendor-market/.

- Structure of the Internet: https://teachcomputerscience.com/structure-of-the-internet/.

- What you need to know about Palantir's direct listing: https://www.protocol.com/palantir-ipo-what-you-need-to-know?rebelltitem=4#rebelltitem4.

- Why is Business Intelligence (BI) important to your company?: https://www.linkedin.com/pulse/why-business-intelligence-bi-important-your-company-yass.

PRIVATE DATABASES

- Big Data Dossier 2020 available at: https://www.statista.com/study/14634/big-data-statista-dossier/.
- Digital economy as percentage of the total economy (GDP) in the United States from 2005 to 2018 https://www.statista.com/statistics/961982/digital-economy-gdp-share-usa/.
- Distribution of rare earth metal exporting countries worldwide in 2019 https://www.statista.com/statistics/702689/global-rare-earth-metal-export-share-by-country/.
- Dassault Systems SA Market report 21 Feb 2021, Marketline Report.
- Ipo in vista per Palantir, la società tech di dati sensibili finanziata dalla Cia: https://24plus.ilsole24ore.com/art/ipo-vista-palantir-societa-tech-dati-sensibili-finanziata-cia-ADiKDed.
- Open Source Intelligence Market Size, Share & Trends | Forecast - 2026 (alliedmarketresearch.com) https://www.alliedmarketresearch.com/open-source-intelligence-market.
- Open Source Intelligence (OSINT) Market Growth, Share, Trade Analysis, Current Trends 2021 – 2027 - MarketWatch https://www.marketwatch.com/press-release/open-source-intelligence-osint-market-growth-share-trade-analysis-current-trends-2021-2027-2021-03-18.
- ORBIS, Data Update 209,001 (02/04/2021) © Bureau van Dijk 2021 Hensoldt Analytics Complete book.
- ORBIS, Data Update 209,001 (02/04/2021) © Bureau van Dijk 2021 Maltego Technologies GMBH Complete book.

LUISS

Dipartimento
di Impresa e Management

Laurea Magistrale in Analisi e Metriche di Marketing
Cattedra: Customer Intelligence e Logiche di Analisi dei Big Data

THESIS SUMMARY

The impact of a Business Intelligence & Analytics Software on an Electronic warfare Company inside the Defence Sector:
The development of the QUIPO Platform to enhance the firm's supplier Scouting.

Prof. Luigi Laura
RELATORE

Prof. Simone Ciccarone
CORRELATORE

Marco Smacchia - Matr. 721671
CANDIDATO

Anno Accademico 2020/2021

Complete Summary

Chapter I

Introduction

Enterprises are under substantial competitive pressures (Knight and Liesch 2016), markets are changing faster and faster, and companies are subject to Digital Transformation challenges. According to McAfee (2012), there is a need to adopt a new source of data called Big Data which is defined as extremely large data sets in volume, velocity, variety, and veracity (Liu 2018). Big Data could help enterprises collect data and convert it into competitive advantages in the global market (McAfee 2012). La Valle (2011) states that the enterprises using data analytics to transform data into useful information outperform their competitors by helping them understanding their business more deeply and improving the decision-making process (Ahi et al. 2017). Consequently, data-driven enterprises such as Amazon, Microsoft, and Apple in less than ten years have replaced oil companies becoming the biggest firms for capitalisation and profits worldwide (Sole 24 Ore). Big Data is also getting bigger as information is coming from a larger number of devices and sources. In the last 2/3 years, humanity has generated more data than ever in its history (Raconteur). It is no wonder that organisation are forced to use only a selection of the vast amount of data available to be effective (Redman 2017).

Business Intelligence Literature Review

During this Digital Revolution, many companies rely on Business Intelligence (BI) software that can help analysts and management to more easily analyse and understand the primary and secondary data collected, facilitating the extraction of meaningful information for decision making. Regarding BI Software, the two most important factors for its development and integration within a company are the creation of a meaningful data lake and correct taxonomic rules. According to Gibson (Babak, 2016 Gibson et Al. 2016), without data, there is nothing to analyse and collecting data in a wrong way could lead to inadmissible failures; so it is imperative to collect data for the right reason and in the right amount. Due to the exponential increase in data, it is becoming increasingly difficult to monitor all the information around us. In this scenario, semantic analysis can help process and visualise online information and extract useful knowledge that helps people make informed decisions. For this to happen, it is necessary to build new taxonomies each time the domain of interest changes (Carrion, 2019). To develop a new taxonomy, it is also essential to have a strong knowledge base on the specific domain (C. Wang, 2017). The creation of structured taxonomic rules is fundamental to increase the software's analysis accuracy and speed up the examination of documents.

There are different definitions of Business Intelligence; one of the oldest and most exhaustive is Hans Peter Luhn’s a researcher at IBM: “an automatic system... developed to disseminate information to

the various sections of any industrial, scientific, or government organisation". Business intelligence is the process of transforming data into information, information into knowledge and knowledge into intelligence; with Artificial intelligence software, all this process is cyclical and automated. P. Karbhari (2006) stated in his research that business intelligence is a broad category of application and technologies for collecting, storing, analysing and providing access to data to help the organisation to make a better decision. Business intelligence helps extract crucial facts from a vast amount of structured and unstructured data and transform them into useful information that makes organisation strategic decisions, improved operational efficiency and business productivity (Jackar et al. 2020). BI also helps to understand in effective, efficient and automatic way micro and macro/environmental factors. B.O. Hostmann (2007) stated in his research that business intelligence facilitates the functional organisation such as processing offline analysis, data mining, business analysis, business research network, and knowledge management and other related activities. His purpose is to help managers and other company functions to make better and informed business decisions. Organisation uses business intelligence to cut cost, identify new business opportunities and spot inefficient process (Jaiswal et al. 2003). Z. Sheng et al. (2005) stated in their research that business intelligence refers to the tools, technology, application, practices use to collect, integrate, analyse and present organisation raw data into actionable business information.

OSINT and the Intelligence Cycle Literature Review

In a 2011 document published by the Office of the Director of National Intelligence, OSINT was defined as: "intelligence produced from publicly available information that is collected, exploited, and disseminated in a timely manner to an appropriate audience for the purpose of addressing a specific intelligence requirement". Through the use of OSINT tools, one can locate, extract and analyse documents from various sources such as blogs, social media, geolocation data, IP addresses, government documents etc. in order to produce intelligence (Quick, 2018). From OSINT analysis come other types of intelligence disciplines used for a variety of purposes. OSINT is used to analyse links between various people, organisations, devices and many other entities (Maciołek, 2013). In the last two decades, there has been a shift from analysing to find hidden information to analysing to find relevant information (Nicart et al. 2016). The growth of data has created an environment where we are constantly flooded with information. It has become very difficult, if not impossible, for analysts to conduct manual searches without the help of software. In this scenario, artificial intelligence and machine learning algorithms are increasingly being used in Business Intelligence projects to partially manage various stages of the intelligence cycle. Although there are various interpretations of the intelligence cycle, it is commonly defined in 5 steps (USA Department of Defense): **planning**, **collection**, **processing**, **analysis** and **dissemination**. These stages include the acquisition and

validation of information, the identification of the value of the information and the distribution of the information to the client. According to Evangelista's study (2020), the main areas of application for OSINT software are Information Security, Digital Marketing and Artificial Intelligence. This last technology is used in software development through the adoption of Machine Learning and Natural Language Processing that allows to make a first analysis in a fully automatic way.

Research Aim

This research is a starting point to study the validation process of a specific "use case" resulting from BI platform implementation by answering, through the formulation of a conjecture, to the following research question:

"In the frame of the implementation of an OSINT Business Intelligence Platform Analysis in the defence market, does the creation of a meaningful Data Lake together with an appropriate Taxonomy structure make the experimental analysis results comparable and coherent to equivalent analysis validated by internal experience and know-how of the firm ?".

In this frame, a segmentation of the BI Platform Market is performed in order to give an overview of the positioning of the QUIPO, a BI platform developed by Cy4Gate S.p.A. The potential growth, scalability and adaptability of the QUIPO application will be assessed to extend its use to other business areas.

Research Methodology

In order to study what mentioned above, is employed data from internal company databases and companies' websites to identify competitors to the QUIPO platform and to segment the market. In relation to the research objectives, will be made a market segmentation through the construction of a two-variable positioning map. The aim is to examine a sample of 14 companies operating in the business intelligence market and position them on the map according to their focus on the use of OSINT data and the possibility of customising taxonomies.

In the second part of the study, an applied research method will be used to empirically validate the use of the BI platform when applied to a specific "use case". According to Kothari (2008), Applied research "aims at finding a solution for an immediate problem facing a society, or an industrial/business organisation, whereas fundamental research is mainly concerned with generalisations and with the formulation of a theory" (Kothari, 2008). More specifically, the study will go through the five intelligence cycle stages to develop the experimental research on the QUIPO platform. Each phase will be procedurally and analytically analysed and issues will be reported along with suggestions for desired changes.

Then, to validate the platform's performance, it will be studied the correlation between a manual OSINT study based on Market and Company Reports together with data available on the web and an experimental analysis performed entirely with QUIPO. Variables will be created based on the information contained in the QUIPO dashboards and compared with those obtained from the manually compiled Excel file using Pearson's correlation index. The research will be conducted with a dedicated team within the Company.

Chapter II

Business Intelligence Market Segmentation

In the competitor study, 14 business intelligence companies were examined and studied in detail to understand their market positioning. The most significant platforms were described in more depth to understand how they operate. The two variables for the market segmentation have been selected by studying OSINT techniques and Business Intelligence literature. The dimensions relate to:

- Focus on OSINT Analysis, representing the degree to which the platform focuses on collecting and analysing external data available on the Internet.
- Focus on TAXONOMY Customization represents the degree to which the platform focuses on taxonomies customisation according to the application cases.

Analysing the results deriving from the positioning map (see fig. below), four market niches were noticed corresponding to the quadrants outlined by the positioning map.

Specific In-House: includes platforms that have a rigid semantic structure and that refer to internal non open-source data. These software collect data provided by the developing company or customer databases. They are used for specific purposes within the organisation that cannot be extended to other business functions. Since these are programs developed to solve a specific problem, the taxonomy structure is very rigid, focusing only on their scope of application, such as CRM applications. The advantages of these systems come from their specialisation. As they are used for very specific tasks, they require less time to be implemented and benefit from more established taxonomies in their field of application. In this section, there are well-known applications of CRM and patent analysis. Although they are very comprehensive software, they are highly specialised in their field of use, unlike Palantir, which is used for several purposes.

Multi-Purpose In-House: includes platforms whose primary purpose is to use in-house data, but unlike the previous group, they have the ability to modify taxonomies depending on their field of application. They are used for various purposes and are very adaptable to various markets and various

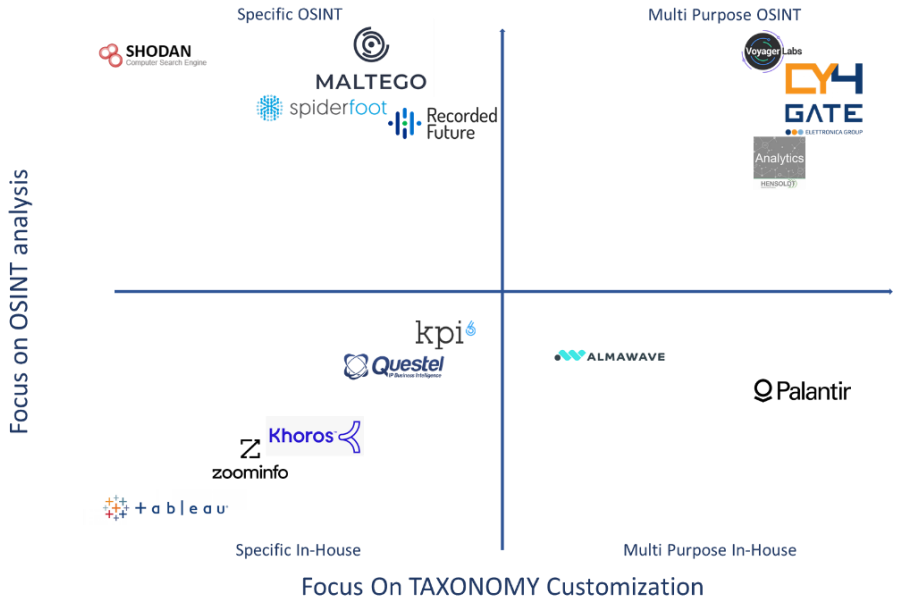
business functions, which is precisely the reason why they need to adapt the taxonomy structure depending on the type of analysis required. The advantages of these software lie in their greater elasticity of use, making them attractive in many sectors and for the most various problems. While being able to analyse external documents, Palantir focuses more on the analysis of internal primary data, ensuring maximum flexibility of use. Almawave, instead, is a company that allows a highly customisable platform development and can adapt to many markets. Both platforms have a wide scope of use, which makes them easily customisable to the customer's needs.

Specific OSINT includes platforms that focus on collecting and analysing open-source data and have a limited ability to customise taxonomies for semantic analysis. Like the programs in the first quadrant, these software are also developed for very specific functions within organisations that require analysis of externally sourced data. The taxonomies are often fixed due to the high specificity of the task, even though platforms dedicated to OSINT research are very widely developed in their scope (e.g. this group includes cyber intelligence software). These platforms are able to manage the data they collect more quickly and accurately because of their knowledge of the primary sources from which they collect information and their highly developed semantic analysis around the main function. In this quadrant, we find platforms such as Maltego and Spiderfoot, known for their specialisation in investigation and cyber threat intelligence. Over the years, these applications have been developing more and more, expanding their scope of use.

Multi-Purpose OSINT: In this niche, the Business Intelligence platforms make open-source data collection their core business and allow developers and designers to create new taxonomies for semantic analysis depending on the platform's field of application. The tools in this quadrant have various features that make them suitable for various tasks, and their flexibility makes them usable by multiple functions within the company. Although they focus on OSINT research, these software are, in most cases, capable of handling internal data to combine the benefits of external and internal research. Since the strength of these platforms is their adaptability to various environments, they require ad hoc taxonomic development to reflect the client's needs. While implementation can be slightly time-consuming, once embedded in business operations, it provides a high degree of flexibility at all levels. In this section, Cy4Gate, Hensoldt and Voyager Labs are in very close competition with each other. Quipo, Hensoldt Analytics and Voyager Analytics focus on external analysis of unstructured documents using artificial intelligence and machine learning. While the three platforms have differences, they are on the same end of the market.

While aiming to help organisations in the collection, management, and analysis of data, BI platforms are very different from each other and can be used for different purposes. QUIPO, the BI Platform

under study, is positioned within the Multi-Purpose OSINT industry market with its related key features to be considered for adoption.



OSINT Competitors Positioning Map.

Chapter III

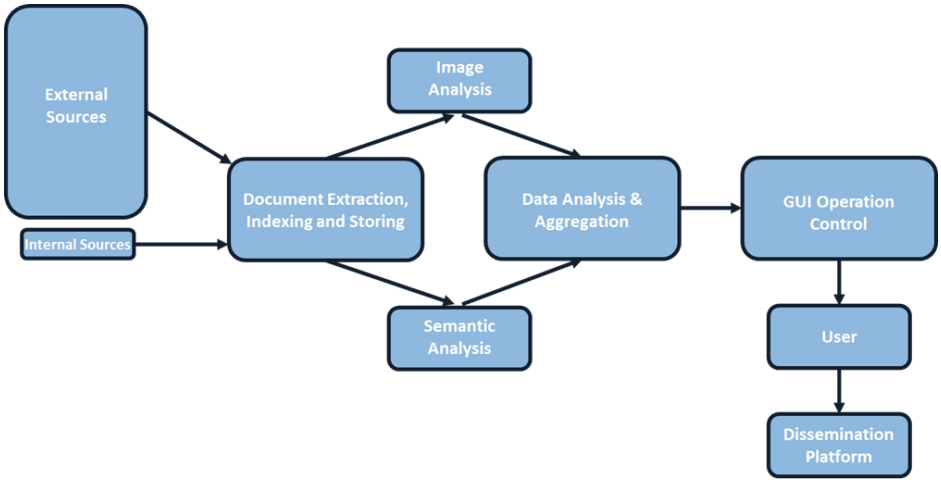
QUIPO Structure and SWOT Analysis

Quipo is Cy4Gate's proprietary Business Intelligence platform focused on supporting and optimising decision making. The software supports the analyst throughout the intelligence cycle by collecting, analysing, and representing OSINT data. Born in 2017 from the collaboration between Cy4Gate and Expert System, Quipo (formerly known as D-Sint) uses the integration of the Cogito semantic engine leading in semantic understanding and natural language processing by leveraging artificial intelligence and machine learning algorithms.

QUIPO consists of three main modules linked together by tools that exploit artificial intelligence and machine learning algorithms to analyse documents:

- Extract, Transform and Load (ETL): the first module consists of a non-relational database with the task of indexing documents from internal and external sources. An archive is then created containing all the documents, enriched with the first metadata that establishes their uniqueness and allows the internal search engine to retrieve them quickly. After being saved and indexed in the database, the data is analysed by the platform's semantic engine or, in the case of images, by image and face recognition tools. In this phase, the documents are assigned additional metadata, which is very important for the representation and analysis by the user.

- Heterogeneous Data Analytics (HDA): the second module deals with the representation of the documents collected in the database using the metadata produced in the previous phases. By aggregating this data, the module allows the creation of dashboards and graphs with which we have obtained the data to formulate our conjecture.
- Operation Control Platform (OCP): the third module is the platform's GUI, which allows the end-user to manage all the tasks necessary to carry out the analysis. It can be considered as the point of contact between the backend tools and the analyst. Closely linked to this module is the data dissemination platform that allows the textual and graphical display of the work done. The user interface of QUIPO is inspired by the intelligence cycle. Through the platform, it is possible to follow the standard analysis process to which are added the functions of creation of dashboards, link analysis, normalisation of taxonomic entities, creation of archives, automatic setting of the reliability of sources and automatic alerting.



Visual representation of Quipo Structure.

Strengths: The company is receiving numerous investments and developing new contracts with government and corporate entities worldwide. In this scenario, the company is betting a lot on its products, investing time and resources in their development. Certainly, one of the products that benefit the most is QUIPO, which can also count on the know-how of Expert Systems, a leading company in the artificial intelligence sector, especially in the field of natural language processing.

Weakness: The company is currently going from the start-up phase to the growth phase. This timeframe is very delicate because it could lead to numerous problems if the strategy is not clearly defined and sustainable.

Opportunities: According to Statista's (2020) Digital Economy report, the global pandemic has increased the need for investment in various areas related to Digital Transformation. Cybersecurity is the sector that benefits the most with a growth in spending of 84%, followed by hybrid cloud infrastructure, automation, and AI with 74%, 66%, and 59% growth, respectively.Cy4Gate is at a delicate stage during this particular period. However, if it can make the right moves, it will be able to gain a lot of customers and investors globally and domestically.

Threats: The threats facing the platform are technical, competitive, and political. First, the risks associated with cyber-attacks are increasing and a cyber-attack on the company's suppliers could be a big problem. Second, even though QUIPO has different functions than some competitors, they can still be considered as substitute assets for some specific needs. In recent years, more companies have been entering the digital marketplace by developing SaaS solutions. There are no high barriers to entry and it is easy to reorganize the company in case of inadequate performance. Finally, over-regulation could present problems in moving contracts forward and accelerating business.

Chapter IV

Elettronica Procurement Project

Elettronica has a twofold objective: on the one hand, it wants to help Cy4Gate develop its flagship product within a sector that more than others can fully exploit the potential of the platform. On the other hand, it has the purpose of inserting a Business Intelligence platform within the business context by expanding its range of use from the Strategy function to all core business functions.

This research is part of the procurement department's project to improve technology scouting that deals to help Elettronica in the decisional choices regarding the business suppliers. In particular, the objectives of the program can be summarized in three points:

1. Improvement of the effectiveness of the process of scouting suppliers starting from the product or capability.
2. Simplification of the "due diligence" of the key suppliers (actual or potential).
3. Technological development trend of specific product categories, both to identify future advantages for Elettronica systems and to identify leading edge or potentially out-of-market suppliers.

The project to improve technical scouting using QUIPO aims to partially automate the collection and analysis of data on potential suppliers, assess their stability, and find new technologies of interest

compatible with Elettronica products. This will speed up research at the beginning of supplier scouting, allowing employees to spend more time on more important steps.

This analysis aims to provide an overview of the companies[42] analysed to identify whether they have the products and capabilities of interest and subsequently whether they are suitable to become suppliers. In the latter part, the company's international relationships over time are assessed along with its financial strength.

Taxonomy Development

A taxonomy is a semantic hierarchy that organises concepts by is-a relations, which exhibits the capability of improving many NLP task (C. Wang, 2017). A taxonomy is composed of rules that allow the semantic engine to understand a given taxonomic entity within a document. A taxonomy development depends on the domain that has to be analysed and the user requirement, who may have different needs according to his objective. A taxonomy can be created using two types of approach:

- Categorisation: a deductive method in which the starting point is a general concept to arrive at specific individual entities. In this approach, a semantic tree is constructed with branches comprising subcategories of more general entities. Categories are generated first by formulating their conceptual descriptions and then classifying objects according to the descriptions.
- Extraction: The inductive method consists of identifying, aggregating, and normalising specific semantic entities to arrive at more generic concepts.

In the Elettronica business case analysis, an entry-level taxonomy of Electronic Warfare (EW) was initially developed to contextualise the possible suppliers within the market, identifying their capabilities and heritage. Therefore, a categorisation approach was used to create a semantic tree composed of general concepts that helped us understand the scope in which a company operates and specific entities to study the differences between players. Within Elettronica, a shared document was created between various experts and analysts who helped to categorise the Electronic Warfare sector from general concepts to more specific capabilities. To date, taxonomies concerning EW macro-areas (i.e. C4ISR, SIGINT, COMINT).

Business Case Intelligence Cycle

Planning. The first phase of the project involved planning the work, along with creating all the archives for the data lake and report compilation.

[42] Note: the analysed companies are not disclosed for confidentiality.

Data Gathering. In the construction of our data lake, were used as sources: news from major defence websites and magazines, Google News, Europe Media Monitor, the press releases of the analysed companies and related organisations, ministerial websites, social media (Twitter and Facebook), the press release of Elettronica and from archives of documents already present in the platform from previous searches to have prior data to our research. In choosing the sources and methods of data gathering applied, the NATO Open-Source Intelligence Handbook guidelines were respected. (Gibson et al. 2016) : The authority of the source, The accuracy (by validating it against other sources), The objectivity of the source, The currency (i.e., the provision of a timestamp for publication and the presence of an author), The coverage (the degree of relevancy). The data lake, at the time of analysis, consisted of 2,250,000 articles analysed by the platform. The next step was to process them by filtering the documents within the data lake according to taxonomic concepts and the companies of interest. **Target:** Improving the data collection with new customised tools that allow the end-user to use an essential source for their research or to perform a very special search. Making existing tools more elastic, so that data collection rules can be further refined.

Analysis: The analysis phase is divided into two parts. The first one concerns the analysis of the data lake through the creation of a first dashboard, link analysis and the developed taxonomies. In this first part, graphs were constructed to explore the distribution and topic of the collected data. Then, through the taxonomic entities, the documents of interest were filtered, analysed individually and inserted into a task (archive for the report's compilation with all the most significant articles, each with a reliability category). In the second part, the task for the compilation of the report was analysed. A dashboard was created for the graphic representation of the results. The EW taxonomy was validated both empirically and through a correlation function with an analysis conducted manually and validated by experts in the field. In the construction of the charts, much attention was paid to the representation of the fundamental aspects of the research, firstly the capabilities, and then the activities of the companies studied (merge & acquisition, joint ventures, partnerships and contracts). A primary focus was given to relations with European defence companies. Finally, the link analysis was created to inspect the relationship between organisations and capabilities better. **Target:** solved some platform issues and requested integration with third-party services (trade magazines and application for financial analysis of companies).

Report. After the analysis part, a report was compiled containing a summary of all the information from the previous step. According to a Top-Down logic, the document was divided, starting with an overview of geopolitical events in the relevant markets. Then, the merge & acquisition activities, partnerships, joint ventures, orders backlog of the analysed companies and their subsidiaries were

analysed in detail. A section was also dedicated to analysing financial data for the calculation of financial strength and solvency. Thanks to this data, it was possible to cross-reference various types of information and understand the context in which these companies operate and their relationship with their government. Then, thanks to the information on activities, it was understood the degree of innovation, the direction they are taking in the development of new technologies, who their partners and customers are, what their strengths are and in which sector they are most developed.

Dissemination. The report, together with the task news, were automatically inserted into AMICO. The graphs in this section are partly created using QUIPO's knowledge base function, which allows the development of customised graphs that summarise the report's contents and partly generated automatically by the platform on the basis of the information contained in the report or articles in the task.

Study Validation. For this research, the focus was on the first cluster of taxonomies available in the procurement project concerning firm Electronic Warfare capabilities in the defence sector by examining three firms that belong to the same capability cluster. Several reports available in the company were inspected and a file containing information about the contracts, merge & acquisition activities, partnership and joint ventures of the organisations under study was compiled. The nine taxonomic concepts concerning Electronic Warfare capabilities within the documents were selected and divided into clusters all the articles for each specific concept. Then the percentage of each cluster was calculated with respect to the total number of documents in the archive to normalise the data. The same it was done for the data inside QUIPO. Next, was studied the existing correlation between the two matrices created through the following formula for the calculation of the Pearson Correlation Coefficient:

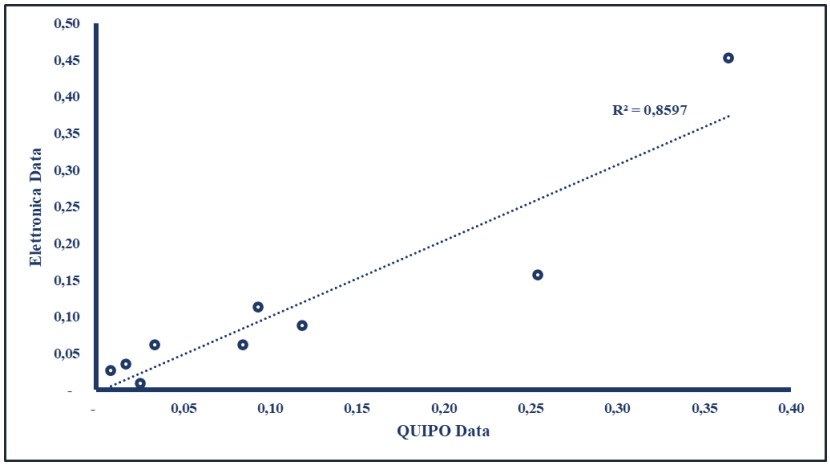
$$r = \frac{\sum[(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{\sum(x_i - \bar{x})^2 * \sum(y_i - \bar{y})^2}}$$

The results (see fig. below) indicate a highly positive correlation (Mukaka, 2012) between the two matrices ($r = 0.927$, $R^2 = 0.86$). That denotes a correspondence between analysis based solely on OSINT techniques using a business intelligence platform and data validated by the firm's experts. An empirical correlation also emerged when compiling the report, noting a sufficient match between the concepts extracted from the platform and those present in the reports analysed.

In light of the following results, the following conjecture was formulated:

"In the development of a Business Intelligence platform, which uses analysis techniques based on the OSINT intelligence cycle, the creation of a data lake following precise rules that point to correct use

of sources together with the creation of an appropriate taxonomy to the specific domain, allows the number of occurrences on a given entity to be very close to reality".



Graphical representation of the correlation between the two matrices (EW Taxonomy).

Capabilities	QUIPO Data	ELT Data
C4ISR	0,36	0,45
DIRCM	0,09	0,11
Training	0,25	0,16
Customer Support	0,03	0,06
RECM	0,01	0,03
COMINT	0,12	0,09
RESM	0,02	0,03
SIGINT	0,08	0,06
ELINT	0,03	0,01

Capabilities examined and their percentages on total documents (EW Taxonomy).

The result obtained is very important as a starting point for the implementation of QUIPO within Elettronica. It was found that in terms of content, the analysis carried out on the platform corresponding to the company's internal intelligence regarding EW capabilities, the company's core business. This made it possible for stakeholders within the company to begin to understand the platform's potential.

Chapter V

Conclusion and Future Research

This research is part of the initial phase of a much larger project within Elettronica and Cy4Gate. Their close collaboration has many benefits for both:

- On the one hand, it allows Elettronica to push on the corporate digitisation process to become a data-driven company.

- On the other it allows Cy4Gate to test their products within the defence market, one of the core sectors for the company. The feedback given during these months will be very useful for the technical and commercial development of QUIPO and will enable the expansion of its scope of use through the introduction and incorporation of new services.

The QUIPO implementation project at Elettronica is a clear demonstration of this. The aim is to extend the use of the platform to the core departments of strategy, procurement and sales to make internal processes more efficient and allow employees to spend more time on more delicate and demanding tasks. The success of this project is crucial for the future of the company, which cannot afford to be delayed in its digitisation process, so it is necessary to counteract any obstacles along the way with a clear and transparent implementation phase that involves as many people as possible and lets them see first-hand how important this tool is and the benefits it brings. Therefore, Elettronica is trying to create several parallel projects linked to different and complementary business functions. In fact, what has been done in this research is as fundamental for procurement as for sales or strategy since all the problems solved and the functions required will be integrated within the platform for all. This research was the start of the procurement-related project and is a first reading of the potential of QUIPO within Elettronica. In response to the main research question, the study findings were able to reproduce more in-depth results than a previous study conducted entirely on a manual basis. There were many insights into the activities of the companies studied that revealed the existence of relationships and products unknown to the company. It should also be noted that these types of projects aim to develop and educate the platform for very specific semantic analysis. In fact, in the implementation of the platform the most time-consuming tasks are the creation of the taxonomies and the approach to the platform, in other words the understanding of the dynamics and functioning of the application. The real added value of this study and the parallel ones is indeed the creation of a shared process within the company that will make these two aspects much simpler and faster in the future. In addition, when it becomes necessary to carry out a supplier scouting again, all the old and new information can be consulted quickly and easily, thus considerably reducing the average analysis time. In fact, the advantage of the platform is the continuity of the analysis. The input given in the data-gathering phase and the semantic analysis of the platform continue to work overtime, guaranteeing the inclusion of the latest news on a given topic.

In conclusion, in the use of an OSINT Business Intelligence platform, the results state that: through the creation of a structured process that includes the creation of taxonomies appropriate to the reference domain and the correct setting of the intelligence cycle starting from an extensive data lake, the information obtained reflects reality.

From a research standpoint, we have seen how these applications should be developed to give added value to the company in terms of research time. Through the creation of shared processes, it is possible to significantly reduce the time needed to create taxonomies and to analyse specific searches, while automatic and continuous ingestion and processing allow searches that have already been carried out to be easily retrieved and enriched with new data, reducing the effort needed for a manual search. This allows employees to pay more attention to other important tasks. A business intelligence platform using AI tools was also analysed, which is very important given the wide acceptance of these types of software. It was seen how some functions within the intelligence cycle can be automated and also the role that the analyst and technician have in guiding the processes of the software. The research also shows that, due to the very large database, new information and insights can be gained through the use of business intelligence platforms that give a competitive advantage to the company by allowing it to have a detailed and real-time overview of the market situation and help it to make more effective decisions. This research had some limitations as it was the initial part of the project mentioned before. First of all, some required functions, such as the integration of some databases, were not available at the time of the study, it will be interesting to see if and to what extent they will improve the intelligence cycle phases of the platform. In addition, the taxonomies, the main part around which the success of such an analysis revolves, were not fully developed in order to extend the boundaries of the research. In fact, the only taxonomy developed internally for semantic analysis was related to electronic warfare capabilities. This resulted in a very subject-specific result that allowed targeted conclusions to be drawn. Future studies must seek to scientifically validate the research conjecture with a broader analysis that aims to compare more complex matrices through a superposition model. It will be necessary to further study the implementation and design processes of BI Platforms that exploit OSINT techniques to create a specific model that allows for easy and rapid application in the field. It is also needed to describe the process of taxonomy construction more analytically by formalising it through the use of validated models for the implementation of information systems artefacts. Finally, it is very important to continue to investigate the results of the application of Business Intelligence software in corporate and governmental environments.