This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement n° 700367

Detecting and ANalysing TErrorist-related online contents and financing activities

D2.3 - Recommendations for LEAs: improving counter-terrorism intelligence on the web

WP number and title	WP2 - Analysis of terrorist-related activities and recommendations
Lead Beneficiary	RISSC
Contributor(s)	RISSC
Deliverable type	Report
Planned delivery date	28/02/2019
Last Update	20/03/2019
Dissemination level	Public

DANTE Project
H2020-FCT-06-2015 - *Law Enforcement capabilities 2: Detection and analysis of terrorist-related content on the Internet*
Grant Agreement n°: 700367
Start date of project: 1 September 2016
Duration: 30 months

Disclaimer

This document contains material, which is the copyright of certain DANTE contractors, and may not be reproduced or copied without permission. All DANTE consortium partners have agreed to the full publication of this document. The commercial use of any information contained in this document may require a license from the proprietor of that information.

The DANTE Consortium consists of the following partners:

	Partner Name	**Short name**	**Country**
1	Engineering Ingegneria Informatica S.p.A.	ENG	Italy
2	Expert System Iberia SL	ESI	Spain
3	Ethniko Kentro Erevnas Kai Technologikis Anaptyxis	CERTH	Greece
4	Ciaotech S.r.l.	CTECH	Italy
5	RISSC - Centro Ricerche e Studi su sicurezza e criminalità	RISSC	Italy
6	PROMT Gmbh	PROMT	Germany
7	Vocapia Research	VOC	France
8	United Technology Research Centre Ireland, Limited	UTRC	Ireland
9	Agnitio SL	AGNI	Spain
10	AIT Austrian Institute of Technology GmbH	AIT	Austria
11	Trilateral Research and Consulting LLP	TRI	UK
12	Fundación Deusto	FD	Spain
13	Pragsis Technologies	PRG	Spain
14	KU Leuven	KUL	Belgium
15	Home Office	CAST	UK
16	Ministerio da Justiça - Polícia Judiciária	PJ	Portugal
17	Ministerio del Interior - Guardia Civil	GUCI	Spain
18	Ministero della Difesa - Carabinieri	ITCC	Italy

To the knowledge of the authors, no classified information is included in this deliverable

Document History

VERSION	DATE	STATUS	AUTHORS, REVIEWER	DESCRIPTION
V0.1	09/11/2018	Draft	RiSSC	Draft Table of contents
V0.2	28/02/2019	Completed version	RISSC, in cooperation with ENG, ESI, VOC, AGNI, AIT, FD, PRG, CTECH, KUL	Completed version integrated with revisions from partners and inputs from final events in Rome (February 2019)
V1.0	20/03/2019	Final	RISSC	Final version integrated with the final input of the partners and the reviewers

Table of contents

List of Figures

Executive Summary

The role of technology is becoming crucial in policing activities, as it provides LEAs with a set of tools able to strengthen their analytical and investigative skills and expand their capacity to handle huge amounts of data, deriving from monitoring and surveillance operations. This is particularly true when dealing with counter-terrorism (CT), as intelligence agencies are required to put greater attention on tracking terrorist financing, monitoring propaganda activities and the dissemination of training materials, both on the Surface and Deep/Dark Web/Net – among other priorities. This can be considered as the starting point, upon which the DANTE project built and elaborated a set of recommendations for law enforcement intelligence and investigation practices on the Web.

Based on DANTE, the **main purpose** of the present study is, in fact, to provide police officers, analysts, investigators, but also other relevant stakeholders in the CT field – such as policy makers, governmental authorities and researchers – with a set of recommendations to contribute at developing and improving some of the existing practices in intelligence and investigation on both the Surface and Dark Web. The reference framework settles within the EU counter-terrorism strategy and its four strands: *prevent, protect, pursue and respond*.

Before entering into details, an overview on the **methodology** adopted in elaborating the recommendations and writing this report is provided.

The approach used was very pragmatic and developed through inductive reasoning, as the final aim was to elaborate recommendations with practical implications. As generally known, inductive reasoning makes broad generalizations from specific observations. Thus, the starting point of the observations was the **set of services** designed, developed and integrated in a platform for automatically detecting, analysing and monitoring a huge amount of terrorist-related online contents and activities, with the final aim of providing concrete support to the LEAs across Europe. For the purpose of this report, specific attention was devoted to some services/modules which are representative of the 'DANTE philosophy'. They are related to different inputs-level of the platform, and specifically: 1) crawling services, as basic tools to start with the investigative activities; 2) first level input – basic mono modal technologies, which include tools for audio-video analysis, text analysis and video-image analysis; 3) second level input – platform enabling technology, mainly aiming at preserving the chain of custody and chain of evidence; 4) third level input – 'connecting the dots', including tools for visual analytics and trend analysis, social graph analysis and financial networks transactions.

It is important to reiterate at this stage, that the aim of this report is not to technically describe the functioning of the entire platform and/or its functionalities, while it rather aims at assessing the potentialities and the concrete application in the investigation and intelligence field, from a cross-cutting perspective, thus elaborating inputs also for further developments.

The recommendations are based also on the dynamic interaction between:

- the **inter-disciplinary and criminological analysis** developed in the framework of WP2,
- the definition of the **scenarios**, the **use-cases** and the **users' requirements** implemented by WP4,
- the **pilots** conducted within WP11.

It was a process that combined the knowledge-generation on the areas of interest of the DANTE project, with the continuous assessment of the LEAs standpoint and the technological development.

Accordingly, the inductive reasoning developed for the sake of the present report is strongly influenced by the **overall project experience** and by the **interactions with the consortium**.

For example, the elaborations presented in Part II of the report, dealing with a set of technical modules pertaining to the above-mentioned areas and their potentialities *vs* current limits in supporting LEAs in CT, are the results of a joint assessment, carried out in cooperation with the end users and the technological developers of DANTE.

Furthermore, most inputs on the DANTE platform have been collected in cooperation with the LEAs during the pilot sessions of the DANTE project held in Rome (25 October 2018), Lisbon (22 November 2018), Madrid (16 January 2019) and Rome (18 February 2019), mainly through an external observation of the demos of the modules to the end users, as well as through direct interviews. The elaboration of the collected feedback has been then shared with the technological partners of DANTE, and thus reflected/incorporated into the final release of the DANTE system.

Finally, both the comments collected during the trainings delivered by the DANTE project to the LEAs, and the overall ex-post evaluation (based on questionnaires disseminated to the participants), provided relevant interesting inputs to develop the recommendations included in the present report.

Accordingly, considering the DANTE outcomes and the efforts to address the key challenges in CT stated by the EC-funding scheme, **FIVE KEY AREAS** have been identified to strategically work towards an improvement of LEAs capabilities in counter-terrorism on the Web.

The key-areas include:

- knowledge,
- technology,
- training and operational capacity building,
- transnational and inter-regional cooperation and ethics,
- privacy and fundamental rights.

In general terms, the overall experience of DANTE confirmed the theorical framework initially designed around the project concept. In fact, it clearly emerged that the role of **technology** was – and still is - crucial because LEAs need to be provided with automated tools for improving detection and monitoring of terrorist related contents and activities, especially online. Moreover, tools are needed by the LEAs to further enhance the handling of huge amount of heterogeneous (big) data, and information about the terrorist threats and the possible interdependencies with other criminal activities. In this view, the project implementation substantiated that a continuous exchange between the end users (LEAs) and the technology providers is of fundamental importance to develop truly relevant technological tools for preventing and countering terrorism.

This stated, the innovative element which emerged as the silver thread throughout the project implementation was the importance of an **inter-disciplinary approach** and the involvement of professionals with different and complementary skills. In fact, this was the key to facilitate the dialogue and the mutual understanding between the LEAs and the technology providers, which presented some challenges due to the complexities of both domains.

Alongside the interdisciplinary approach, also the **knowledge-based approach** has proved to be decisive. In fact, the DANTE project confirmed that the development of technological tools, in order to be effective, must be based not only on the LEAs needs, but also on the characteristics of the phenomenon to be faced. The creation and sharing of knowledge, based on the inter-disciplinary approach mentioned above, was crucial under many respects. Some of them are hereby listed as a matter of example.

It definitely supported the dialogue and cooperation among the end users and the IT developers. Based on the criminological analysis, along with the relevant inputs from the legal, privacy and ethical assessment, the LEAs were facilitated in explaining the procedures and activities they usually implemented to investigate terrorist related crime, as well as to monitor the online environment. On the other side, the technological partners have been facilitated in understanding the different constraints affecting the enforcement environment as well as in proposing tools to help improve the quality of prevention and contrast activities, thus avoiding additional possible constraints.

The combination between the criminological and the technological approach supported the process of refinement of the use-cases as well as the fine-tuning of the users' requirements.

The assessment of the phenomenon from the criminological standpoint contributed also to highlight possible scenarios and trends to be further considered also in the development of the technological tools, so to make them innovative and capable to support the LEAs in facing the possible evolution of the criminal threats/risks on the short-medium term.

Furthermore, the knowledge-generation process initiated at the very beginning of the activities also supported the identification of relevant sources where to retrieve information, data and documents, needed to develop and train the technologies used in the framework of the technological development.

The DANTE project is a good practice of reference to further encourage the **virtuous circle of knowledge-technology-enforcement**, which emerged to be fundamental to achieve effective and sustainable results when talking about innovation actions in the field of counter-terrorism.

This virtuous circle cannot avoid including also the key role of the inter-disciplinary and knowledge-based approach in contributing at enhancing operational skills within the LEAs. From this standpoint, the role of **training/capacity building** activities was as well a fundamental one. As emerged from the DANTE project, the multidisciplinary perspective should be further reflected into the LEAs capacities, through tailored actions – also to encourage the exchange of existing practices.

Due to the transnational nature of the terrorist phenomenon, the DANTE project confirmed that **inter-regional cooperation** is a crucial aspect, too. The online framework has no geographical boundaries, but the dynamics and trends of the criminal phenomena – including terrorism - are strictly related to the offline dimension, which is geographically located. Accordingly, the online-offline should be a priority, as well as the cooperation among national agencies. It could be further promoted for example through actions able to convey transnational analysis on the phenomenon, technologies covering more languages, trainings and capacity-building events involving participants from different LEAs… These examples, emerging from the DANTE experience, were further processed to elaborate the recommendations included in this report.

Finally, as a cross-cutting element, the management of **ethical, legal and privacy** issues need to be reflected and addressed in all the above-mentioned areas. The DANTE project confirmed that the virtuous circle of knowledge-technology-enforcement should be further supported by the contribution of these disciplines so to achieve sustainable, efficient and effective results.

The detailed analysis of each key-area relevant **RECOMMENDATIONS** is included in Part III of the present report.

In conclusion, terrorism online has been shifting from linearity, where technology is used as a mean to perpetrate attacks (cyber-terrorism), to complexity, where technology becomes an hybrid warfare enabler (digital terror or (cyber-)social terrorism), up to an hyper-complex form of 'onlife' terrorism, also defined as a post-truth warfare in an hybrid world.[1] In simpler terms, the online dimension is propagating in all the different aspects of the terrorist phenomenon, including several aspects of the daily life of potential terrorists and extremists, thus expanding their possibility to reach their objectives and, at the same time, requiring a substantial shift into the countering strategies to be adopted.

The right balance needs to be found, between the need to detect and remove terrorist contents online and the maintenance of more traditional investigative techniques, by keeping in mind that prevention has to be the first objective. At the same time, while LEAs need to acquire knowledge on the potentialities offered by new technological tools, they also have to be aware of their main limits. End users can play a crucial role in the co-design and co-creation of such tools, by highlighting their needs, first, and thus ensuring the sustainability of the developed tools.

While the differences characterizing the European legal and operational scenario have to be considered, the possibilities for cooperation and information exchange are probably with no precedents nowadays. Thus, projects such as DANTE represent the perfect platform for dialogue, technological and knowledge development, and their results represent a solid basis for further developments and improvements in this field.

[1] This definition was proposed by prof. Arije Antinori during the experts' workshop on CT training for LEAs held on 19 February 2019 in Rome, in the framework of the DANTE project.

1 Introduction

Based on the evaluation of the overall DANTE project, this report presents a set of recommendations, mainly addressing Law Enforcement Authorities (LEAs), on how to best improve intelligence and investigations practices in the fight against terrorism, especially on the Web. It is mostly based on the standpoint of both criminologists and LEAs. However, the experience of the DANTE project has been so diversified as to suggest a reflection also on the role of technology and on how to increase the overall results and the impact of the results achieved, thus promoting further developments.

This deliverable is part of the knowledge-generation activities, which permeated the overall project implementation. It builds upon the results of the inter-disciplinary and criminological analysis developed within *WP2 - Analysis of terrorist related activities and recommendations*, acting this work package as a connector among the rest of the DANTE project WPs.

More specifically, the knowledge base played a twofold role:

- first of all, the main findings provided guidance during the technological development process, in order to adapt, as much as possible, the main functions of the tools to the real needs of the LEAs, to the main trends of terrorist-related activities and to the modus operandi of the main actors active in the current international scenario;
- secondly, the knowledge generation/sharing and the analytical work continued also during the entire testing phase of the platform tools, and specifically during the pilot sessions with the LEAs, organized within DANTE to provide the end users with the demos of the different modules. Inputs, suggestions for improvements, feedback on the practical applications of the tools in the enforcement environment were elaborated in close cooperation with the end users to support in the finalization of the platform development. At the same time, the pilots' evaluation and assessment were used to come up with the set of recommendations herewith described.

Part I of this report sets the framework of the analysis, by outlining the main challenges and opportunities in the monitoring of online activities operated by the LEAs. In fact, the online surveillance and monitoring have become part of our daily lives in the last decades and increased the opportunities for ensuring citizens' risks in many fields, including counter-terrorism. Based on the evolution of the terrorist activities and entities, the overall approach to the information flow has changed: in line with the focus of DANTE, the Islamist or Jihadist terrorism phenomenon has been the focus of this report. The cooperation between public and private is also investigated, as modern security companies and technological providers more and more play a crucial role in the intelligence and counter-terrorism field – and the exchange with governments and public institution is crucial. The possible concerns related to fundamental rights are also included in the picture, as the right to privacy, to non-discrimination and to data-protection can be heavily affected by the technological evolutions in the surveillance and security field: the right balance needs therefore to be found.

Part II then goes further in details into some of the main functionalities of the DANTE platform. The following approach was adopted: keeping into consideration the main target readers of this report (LEAs, policy makers, academics and other potential relevant (also non-technical) stakeholders), some services/modules of the platform was selected. The selection was done in close cooperation with the technical developers, by keeping in mind the final aim of this analysis. In fact, this report is not aimed at describing the entire DANTE system from a technical point of view, but it rather aims at looking from a cross-cutting perspective at the technical potentialities of the platform and the concrete application in the investigation and intelligence field, with the intent to further suggest possible application in the enforcement environment and improvements.
Accordingly, the elaborations presented in Part II of the report are the results of a joint assessment, carried out in cooperation with the end users and the technological developers. Most of the inputs from the LEAs

have been collected during the pilot sessions[2] of the DANTE project held in Rome (23-25 October 2018), Lisbon (20-22 November 2018), Madrid (16-17 January 2019) and Rome (18 February 2019), mainly through an external observation of the demonstration of the modules to the end users as well as through direct interviews. The elaboration of the collected feedback has been then shared with the technological partners of DANTE, and thus reflected/incorporated into the final release of the DANTE system.

On this basis, a set of practical recommendations and guidelines for LEAs to improve counter-terrorism intelligence on the web have been developed and are presented in the last section of this report – Part III.

Based on the activities and related findings of the DANTE project, implemented as a response to some of the main challenges in the CT-scenario identified by policy makers at EC level, the recommendations are organized in five key areas, identified as strategic to work towards an improvement of LEAs capabilities in counter-terrorism. The five areas include:

- knowledge,
- technology,
- training and operation capacity building,
- transnational cooperation, and
- ethics, privacy, fundamental rights.

[2] The pilot sessions consisted in practical demonstrations to the LEAs of the different modules of the DANTE platform, to collect practical inputs for revisions and improvements. For each DANTE use case, 2 pilot sessions have been organized with representatives of the LEAs participating in the consortium.

2 Part I - Challenges and opportunities of online monitoring technologies in counter-terrorism

2.1 Online monitoring activities in the field of terrorism prevention and contrast

Access to information, and the means to create and process information, have grown dramatically in the last few decades. The Internet and the new technological tools provide a platform for different purposes, including dissemination of people's work at virtually no cost, political activism, social interactions etc., and permits the instantaneous search and exchange of volumes of data. One of the consequences is that it has now become much easier for organizations and individuals to capture, process and disseminate information about individuals.

A number of entities are able, nowadays, to track and monitor online behaviours and everyday online activities by, for instance, monitoring digital networks, or by tapping into the huge amount of data that are recovered and which concern personal transactions. "The Internet is unquestionably a surveillance medium *par excellence*".[3]

Working definitions

The terminology concerning the online environment and relevant (cyber) operations is in continuous evolution. First of all, the 'cyber space', according to the Cambridge dictionary, corresponds to "the internet considered as an imaginary area without limits where you can meet people and discover information about any subject"[4]. According to the Oxford dictionary definition, it is "the notional environment in which communication over computer network occurs".[5]

Monitoring (also referred to as 'surveillance') generally includes the gathering and analysis of information in the pursuit of various finalities – in particular, preventing certain risks, orienting human behaviours and, in the event of a problem, locating the persons responsible.[6]

Privacy International[7] refers to 'communications surveillance' as the "[…] monitoring, interception, collection, preservation and retention of information that has been communicated, relayed or generated over communications networks to a group of recipients by a third party. This third party could be a law

[3] Bennett, C. J., Clement, A., and Milberry, K. (2012), "Editorial: Introduction to Cyber-Surveillance", in *Surveillance & Society* 9(4): 339-347.
[4] https://dictionary.cambridge.org/dictionary/english/cyberspace.
[5] https://en.oxforddictionaries.com/definition/cyberspace.
[6] Commission de l'étique de la science et de la technologie (2008), "Viseur un just équilibre: un regard étique sur les nouvelles technologies de surveillance et de controle à des fins de sécurité", adopted at 34th meeting of the Commission on 12 February. French version available at: http://www.ethique.gouv.qc.ca/fr/assets/documents/NTSC/Avis-NTSC-FR.pdf;
Cahen, M., "Le role de l'administrateur réseau dans la cybersurveillance", lecture notes, ENAP; Boudreau C. (2006). Article available at: http://www.netalya.com/fr/Article2.asp?CLE=162.
"Multipolarité de la surveillance et gestion des médicaments au Québec", in *Recherches sociographiques*, vol.47, n.2, pp. 299-320. Article available at: https://www.erudit.org/fr/revues/rs/2006-v47-n2-rs1449/014205ar/.
[7] Privacy International (PI) is a registered charity based in London that works at the intersection of modern technologies and rights. Further information available at: https://privacyinternational.org.

enforcement agency, intelligence agency, a private company, or a malicious actor”.[8]

According to the Encyclopedic dictionary of public administration[9], “cyber-surveillance is a mechanism for the surveillance of persons, objects or processes that is based on new technologies and that is operated from and on data networks, such as the Internet. Its purpose is to facilitate surveillance, in keeping with the quantity, rapidity or complexity of the data to be processed”.

It has also been defined as a component of the so-called Intelligence, Surveillance, Reconnaissance (ISR) process, and namely as “the continuous observation, direct or indirect, of the operations of the ‘enemy’, in the cyber space or in portions of the cyber space of the Area of Informative Interest”.[10]

For the purpose of this report the word ‘monitoring’ will be used with the general meaning of surveillance, as herewith defined.

Figure 1: Monitoring and surveillance concepts

Source: RiSSC

Terrorism prevention and contrast

The online surveillance and monitoring have now become part of our daily lives, as the spread and improvement of its components have augmented the opportunities for managing emerging risks and ensuring the security of people, places, data, infrastructures and processes across a number of sectors.

Even though the digital environment is potentially perfect for the expansion of monitoring practices, as long as there are bodies able (and willing) to exploit them, “it is [also] easy to over-estimate the ubiquity and detail of digitally mediated surveillance practices based on these overall trends”. For sure, the considerably expanded capabilities in the digital networking field have become a matter of controversy and even concern. In fact, law enforcement and intelligence agencies have tried to develop new surveillance capabilities and acquire legal powers to monitor Internet-related communications.

The international terrorist attacks, in particular since 9/11 on, led to an intensification of monitoring methods and techniques, including online, based on the need to apply measure to reduce, on one side, the potential threats to States security and, on the other, to defend the security of population and the general sense of

[8] Privacy International (PI), Communication Surveillance, available at: https://privacyinternational.org/explainer/1309/communications-surveillance.

[9] Tremblay, M. (2012). “Cyber-surveillance”, in L. Coté and J.-F. Savard (eds.), *Encyclopedic Dictionary of Public Administration*, available at: http://www.dictionnaire.enap.ca/dictionnaire/docs/definitions/definitions_anglais/cyber_surveillance.pdf.

[10] Centro Alti Studi per la Difesa – ISRI (2014-2015), L’evoluzione della capacità cyber, da cyber defence a cyber warfare. Available at: http://www.difesa.it/SMD_/CASD/IM/ISSMI/Documents/L_evoluzione_della_capacita_Cyber.pdf.

fear deriving from such threats.

While there is not an internationally agreed definition of ‘terrorism’ or ‘terrorist acts’[11], at EU level a terrorist act is intended as an act committed with the aim of “seriously intimidating a population”, “unduly compelling a government or international organization to perform or abstain from performing any act” or “seriously destabilising or destroying the fundamental political, constitutional, economic or social structures of a country or an international organization”.[12]

A number of studies and analysis have been developed on the main threats posed by terrorists using the Internet for terrorist purposes.[13] Internet is used, among others, for the following purposes: to carry out attacks to IT structures or at objects in the ‘physical’ word; to inform, threaten and attract attention also through the dissemination of propaganda material; to exchange encrypted information or obtain information about possible targets; to radicalize and train potential terrorists; to raise funds and obtain profits to fund potential attacks of the organization structure.

At the same time, some characteristics of the Internet – such as its openness and accessibility – provide the intelligence community with a variety of material for foundation intelligence and analysis. In the last decades, intelligence agencies within and outside EU have been focused most of their efforts on the threat posed by the so-called Islamist terrorism.[14]

Islamist groups like al Qaida and the Islamic State (IS) use violence against non-Muslims with the aim of establishing a political institution based on the *shariah*. In particular, Islamists have appropriated the concept of jihad to legitimise an offensive ‘holy war’ against non-Muslims.[15]

“Without the internet, the radical groups making up the global jihad’s cadre of militants would remain a widely dispersed and isolated group of cells that happened to claim the same historical roots. It is the internet which has ‘globalized’ the jihad movement. The network of global jihad is a product of the communications revolution”.[16]

Analysts worldwide have well understood, on the other side, the potential represented by the other side of the coin, and thus how much can be learnt by what the ‘enemies’ tell about themselves and among themselves. Of course, there are obvious barriers, partly related to linguist issues or to the incredibly huge number of sources available online.

In this regard, the extent to which private companies and corporations are implicated in these evolving trends is more and more obvious (as it will be explored in the next paragraph). The private sector in fact has been capable to develop research and tools to improve the data gathering and analysis in this field, in some cases even receiving funding from governmental authorities.

As a general tendency, intelligence analysis leads to the understanding of phenomena and trends. “[…]

[11] European Parliament (2015), Understanding definitions of terrorism. Available at: http://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_ATA(2015)571320.

[12] CoE (2001), Common Position 2001/931/CFSP of 27 December 2001 on the application of specific measures to combat terrorism, Official Journal of the European Communities, L344/93. Available at: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex:32001E0931.

[13] For example: ICCT; UNICRI; UNODC; Max Planck Institute; Council of Europe.

[14] The term Islamism refers to the political ideology of relatively small groups that borrow concepts like *sharia* or *jihad* from Islam and reinterpret them to gain legitimacy for their political goals.

[15] An in-depth focus, which summarises some main academic and media-based contributions is available at: http://theconversation.com/why-the-media-needs-to-be-more-responsible-for-how-it-links-islam-and-islamist-terrorism-103170.

[16] Shahar Y. (2008), “The Internet as a tool for intelligence and counter-terrorism”, in *Responses to Cyber Terrorism – Centre of Excellence Defence Against Terrorism*, IOS Press.

intelligence-gathering should lead to a fundamental understanding of what is going on in the sphere of interest, who the main players are, and who their friends and enemies are".[17] As a matter of fact, "proper foundation intelligence can lead to a better understanding not only of how things are likely to unfold in the sphere of interest, but also what effect contingencies outside of that sphere may have on events. Extrapolation from foundation intelligence is the work of professional analysts, who provide the crucial link from *foundation intelligence* to *tactical intelligence*".[18] The Internet revealed to be an important source for what is considered *foundation intelligence*, and thus regarding components such as ideology; strategy; tactics and structure.

At the same time, based on the evolution of the terrorist entities, - with a specific focus on Islamist terrorism - even the approach to the so-called information flow has changed, shifting from top-down (vertical) approach to a more horizontal one. What does this mean and what are the consequences? The initiatives for intelligence gathering and analysis are moving from the hands of the public to those of the private sector. "It is important that the 'professional' intelligence community understands this shift […]. They need to learn [how] to make proper use of the ground-up efforts of academics, internet sleuths and independent analysts".[19] At the same time, information operations are also shifting from the government to the private sector, as the public sector is more and more recognizing the potential of the privates to manage their own information and data.

"For the [terrorists], this is a two-edged sword; the greater their dependence on the Internet, the greater their reach and efficiency, but also the greater their vulnerability".[20]

2.2 The cooperation with the private sector

"Defence-related production is a branch of the economy, often an extremely profitable one, which can make major contributions to full employment and produce shared and spun-off technologies that are also useful for the civil sector".[21]

The North Atlantic Treaty Organization (NATO) has developed its own systems to manage and activate private-sector assets in the event of a crisis. The 1990s – a decade in which the fall of the Berlin wall brought to a general re-shape of borders and power balances within and outside Europe - are generally considered as a turning point, when the traditional notions of the link between the defence and the business area began to diversify. "[…] it may be argued that in the last decade of the 20th century the independence, the variety and the salience of private-sector roles in global security all increased".[22]

Even though the issue of public-private interactions in security was undergoing complex development even during the last part of the XX century, the terrorist attacks of 9/11 – and the reactions to them – gave the security agenda a massive lift. Business and state modified the ways in which they used to interact among (and help) each other, and the need for a more systematic framework to regulate these interactions emerged

[17] Ibid., p.122
[18] Ibid., p.122
[19] Ibid., p.132
[20] Ibid., p.133
[21] Bailes A., Frommelt I. (2004), Business and security: public-private sector relationship in a new security environment, SIPRI, p.2. Available at: https://www.sipri.org/publications/2004/business-and-security-public-private-sector-relationships-new-security-environment.
[22] Ibid, p.4

as a consequence.

Focusing on the role of the private sector in the security industry, to then look more closely at the role played in the intelligence and counter-terrorism field, modern security companies provide a range of services. Most of them have an analytical and intelligence arm, as well as a practical/on field one.

With 9/11, States had to question themselves on why the existing intelligence systems failed in anticipating or warning the attacks. According to the Joint Inquiry released by the Senate Committee on Intelligence: "Prior to September 11, the Intelligence Community's understanding of al-Qa'ida was hampered by insufficient analytic focus and quality, particularly in terms of strategic analysis […] there was a dearth of creative, aggressive analysis targeting Bin Ladin and a persistent inability to comprehend the collective significance of individual pieces of intelligence. These analytic deficiencies seriously undercut the ability of US policy makers to understand the full nature of the threat, and to make fully informed decisions".[23]

In modern times several other 'failures' related to crisis-linked or terrorism-linked intelligence took place, including the collapse of the Soviet Union, the Argentinian invasion of the Falkland Island etc. At present times the challenges related to the Islamist extremism are getting the picture even more complex.

The involvement of the private sector in intelligence-related tasks can bring several benefits, partly related to analytical skills, but also linked to the effects of globalization on companies, and the possibility for the employees to draw from a range of different regions and cultural backgrounds. "The best collection principle for human intelligence is to make use of everyone […] This notion lies behind the best designed anti-criminal and anti-terrorist information campaigns […] The same argument can be made for the added value to be gained from private intelligence collection."[24] Two benefits may derive from the successful inclusion of private-sector expertise into the counter-terrorism strategy, including in the threat identification and assessment: it could bring an increase in analytical skills and new policy thinking and elaboration; it could also represent an achievement in terms of transparency and restoration of public trust.

In this context, the role of technology is becoming crucial. The DANTE project has found its *raison d'être* in the urgency of enabling LEAs and intelligence officials to continuously monitor in near real-time online relevant (for the purpose of counter terrorism and under lawful warrant) communications and contents, both in the Shallow - surface or static - Web and in the Deep Web.

There are estimates that Deep Web contains more than 95% of the available data on the Internet, not always indexed by automated search engines, but potentially providing useful contents and information for detecting and fighting terrorist activities. Huge parts of the WWW 2.0 and WWW 3.0[25] content is often protected by registration, and thus not indexed by search engines. However, forums are often used by radical organisation to exchange knowledge, to spread disinformation and to organise supporter communities in the so-called Deep Web. Dark nets like Tor and I2P are used to anonymise the communication in the Deep Web, as well as to provide hidden services. In the dark net of the Deep Web, black markets are used to exchange illegal goods and to exchange knowledge.

[23] US Congress (2002), Joint Inquiry into Intelligence Community Activities before and after the Terrorist Attacks of September 11, 2001: Abridged Findings and Conclusions.

[24] Bailes A., Frommelt I. (2004), cit. 21, pp. 181-182.

[25] According to Techopedia.com, *Web 3.0* is slated to be the new paradigm in web interaction and will mark a fundamental change in how developers can create websites, but more importantly, how people interact with those websites. Instead, Web 2.0 was simply an evolution from the original Web which can be compared to a library, as Web 1.0 was simply a place where people just placed walls upon walls of text which people can read but usually cannot interact with. Web 2.0 previously changed this, by allowing user interaction with dynamic websites that acted more as applications than simply pages of information. For more information see also: https://www.techopedia.com/definition/4923/web-30.

Due to the escalating number of Internet users and the increasing speed of creation/deletion of Internet contents, it is clear that searching terrorist-related contents (i.e. generated by terrorists or individuals linked to terrorists, or linked to terrorist activities, including relevant contents generated by non-terrorists) and information by keywords and manually is highly error-prone in precision and a lot time-consuming, dramatically slow and obsolete, making impractical the examination of huge amount of resources. There are several unsolved scientific and technical issues involved in getting this data efficiently. Existing software solutions (such as Palantir) can support generic Internet analytics and analyse the massive amount of data in minimal time. However, being able to take advantage of this information, currently referred to as Big Data, is fundamental for LEAs, who are aiming at more efficient and intelligent systems for managing information specific for counter terrorism and intelligence purposes.

2.3 Fundamental rights and security concerns

As previously stated, in response to the use of the Internet for criminal purposes, including by terrorism suspects, policing and intelligence agencies have developed new capabilities and lobbied for new legal powers to put Internet users under surveillance.[26] Some of these powers include requirements for Internet Service Providers (ISPs) to facilitate wiretaps and to store information about their customers' communications and Web browsing activities.[27] As it is easy to imagine, this causes serious concerns to the society and the citizens, as they started to feel their private life overly invaded.

The so-called 'right to privacy' is an integral part of the right to respect for private life, as guaranteed by Article 8 of the European Convention on Human Rights (ECHR), which encompasses another special right, usually referred to as 'data protection'.

This right does not concern simply protecting individuals from intrusions into their privacy or private life, but also more broadly is about guarding against the improper collecting, storing, sharing and use of their data.[28]

Terrorism - which is generally seen as a not passing phenomenon - and some of the measures adopted against it, tend to pose long-term threats to the fundamental values of the States; those related to data collection, storage and sharing are of concern. Data protection is sometime seen as an obstacle to anti-terrorist measures; yet it is crucial to the safeguarding of fundamental democratic values.

A first set of technologies aims at direct surveillance (CCTV, motorway cameras, face recognition software, technologies to monitor and analyse billions of phone and mail communications etc.). Secondly, there is a massive expansion in the so-called "dataveillance", which consists in the monitoring of "data trails" left by individuals in a number of transactions, including through access to private and public-sector databases. Moreover, the police and the secret services search through such databases in order to find a possible 'match' against a predetermined 'profile': such searches are then increasingly intelligence-led and carried out as part of EU polices, and not only national.[29]

[26] Brown I. and Korff D. (2009), "Terrorism and the proportionality of Internet surveillance", in *European Journal of Criminology*, SAGE.

[27] Brown I. and Korff D. (2004), "Striking the right balance: respecting the Privacy of individuals and protecting the public from crime", Information Commissioner's Office.

[28] Council of Europe (CoE) (2008), Protecting the Right to Privacy in the Fight against Terrorism, Issue paper of the Council of Europe Commissioner for Human Rights. Available at: http://www.coe.int/t/commissioner/Activities/IPList_en.asp.

[29] For more information, see https://rm.coe.int/ref/CommDH/IssuePaper(2008)3.

Many of the technologies newly developed pose threats to potentially every citizen's privacy and freedom, and the legal framework aimed at limiting the negative effects is thus in continuous evolution. More in general, the legal context defining the right to privacy in the context of anti-terrorism is complex. The data protection regulation developed on the basis of Article 8 of the ECHR; it was also given a separate provision in the EU Charter of Fundamental Rights (art.8).[30]

As a general matter, the fight against terrorism is seen as a global problem and thus require a global response. The EU strongly encourages the use of new IT technologies in different areas, such as e-government, e-health and including in relation to law enforcement – with counter-terrorism acting as a powerful catalyst.

Data gathered by LEAs are now available to be shared across Europe under the principle of 'availability', defined in the Hague Programme of the EU[31], and thus allowing free access and data sharing without any of the obstacles contained in the previous instruments for transnational cooperation.

The profile usually developed by LEAs to target suspects, including in the field of counter-terrorism, are increasingly the result of a joint-effort of different police forces at international level, mainly supported by new technologies. Many of these technologies may pose diverse threats to privacy, as they allow the States to closely control the lives of their citizens.

The risk of discrimination is another potential infringement of fundamental rights of citizens, as there is a current focus from enforcement agencies to try to prevent individuals (predominantly 'young Muslims') from being drawn into extremism, as part of the overall fight against terrorism. "As with previous action against individuals deemed to be suspicious because they belong to a specific group, or fit a certain stereotype, this is almost certain to lead to discrimination against such minority groups. The fact that a supposedly sophisticated computer-generated algorithm replaces a coarse stereotype does little to prevent this. By being incomprehensible even to those that rely on it, and effectively unchallengeable by those that are targeted, it aggravates the risk of discrimination".[32]

States thus have the difficult task of balancing human rights interests: they have to protect the society from the terrorist threats, while safeguarding the rights of individuals. For instance, even though new technologies such as data mining and profiling can contribute to capturing terrorists, there can be also a proportion of 'false negatives' and thus the general risk is to give up freedom without actually gaining security. Finding the correct balance in this field is one of the most importance challenges in the field of monitoring technologies applied to the international fight against terrorism.

[30] The most recent data protection instruments developed within the EU include, among other: i) the CoE Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data and its Additional Protocol, ii) Directive 95/46/EC of the EU Parliament and the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data; iii) Directive 2002/58/EC of the EU Parliament and of the Council concerning the processing of personal data and the protection of privacy in the electronic communications sector (e-Privacy Directive); CoE Recommendation R(87)15 of the Committee of Ministers to MS, regulating the Use of Personal Data in the Police Sector (1987). A number of standards can also be derived from the judgments of the ECHR and are Reflected in the case-law of the European Court of Justice; Cyber Intelligence Sharing and Protection Act.

[31] Available at https://www.mvcr.cz/mvcren/file/the-hague-programme.aspx.

[32] Brown I. and Korff D. (2009), "Terrorism and the proportionality of Internet surveillance", in *European Journal of Criminology*, SAGE, p.6.

3 Part II – Potentialities of the DANTE platform in counter-terrorism intelligence and investigations on the Web

3.1 Counter-terrorism investigation and intelligence: operational tips for LEAs

Since 9/11, policymakers worldwide started to state that a new intelligence system to fight terrorism had to be developed. During what is generally referred to as 'war on terror', intelligence has played a critical role in both offensive and defensive operations (e.g. in Iraq, Afghanistan etc.). While the role of prevention against terrorism is becoming more and more challenging, partly due to the changes in the preparation and implementation of terrorist attacks (low cost attacks operated by lone wolves *vs* more structured and organized groups and plans), the means/tools used (simple weapons such as knives *vs* elaborated set of explosives ; the role of Internet/technological means) and the relevant funding schemes, the early detection of potential suspects, fund raising activities and propaganda materials can play an essential role. In this framework, a combination of intelligence and law enforcement work should be encouraged.

The IT system developed within DANTE goes in this direction, as further explained in the following sections.

Intelligence in counter-terrorism

Before delving into a more detailed description and analysis of some main functionalities developed in the DANTE system, it may be useful to highlight the differences between:

- information,
- intelligence, and
- the analytical process involving both.

While **information** refers to raw data of basically any type, **intelligence** is data which has been worked on, and thus acquire added value or significance. **Evaluation** is thus the process of considering the information with regard to the specific context, through its source and reliability, and transforming information into intelligence.

Focusing more specifically on 'criminal intelligence', the use of Information Technology (IT) for storage and retrieval of crime information has been of notable success and provides new tools to understand the data collected and communicate them to others. The way in which intelligence can be used for law enforcement purposes is normally defined by the law; legislation also defines whether the material gathered during an investigation is protected from disclosure in criminal proceedings. Thus, "[..] intelligence analysis aids investigations by helping to target available resources and identifying information gaps to focus the investigation more clearly"[33].

Figure 2 below outlines the so-called intelligence cycle.

[33] UNODC, Criminal Intelligence. Manual for Analysts, 2011. Available at: https://www.unodc.org/documents/organized-crime/Law-Enforcement/Criminal_Intelligence_for_Analysts.pdf.

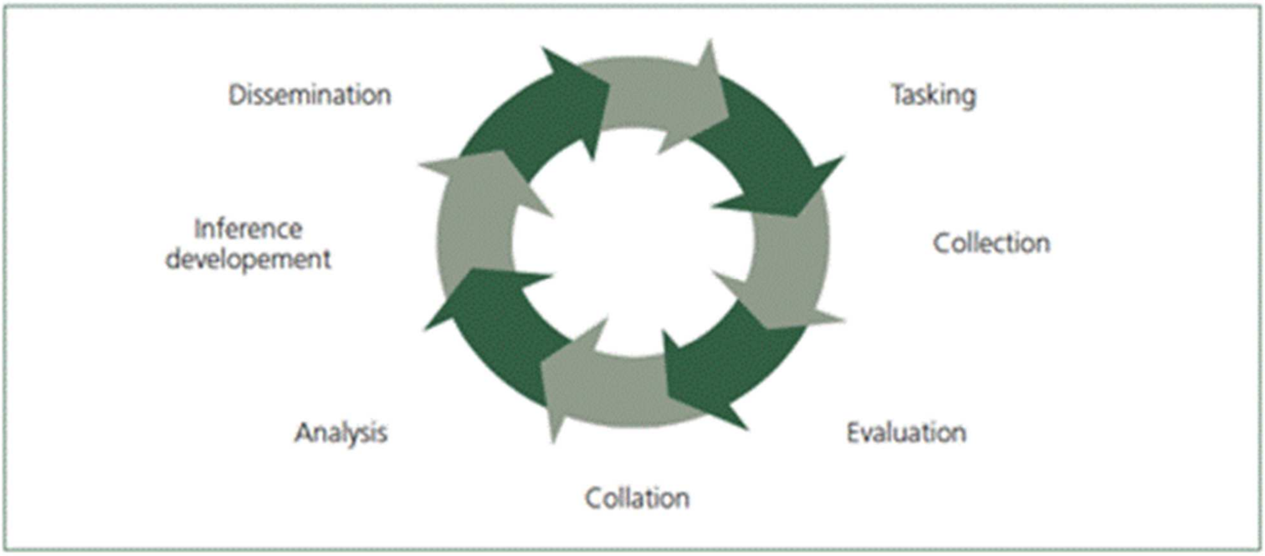


Figure 2: Intelligence cycle

Source: UNODC, 2011

Direction/Tasking: intelligence analysis is driven by the needs of the client (eg. LEAs); for tasking to work effectively, a good client/analyst relation should exist.

Collection: the gathering of data. Defining a collection plan helps in avoiding data overload and ensuring order and precision. Three main types of sources: open source-OSINT (information publicly available); closed source (information collected for a specific purpose with limited access and availability); classified (information collected by specifically tasked covert means including use of human/technical resources).

Evaluation: assessment of the source reliability and the quality of the information.

Collation: organization of data collected into a format from which it can be retrieved and analysed.

Data integration and analysis: analysis consists in the careful examination of the information collected, to discover its meaning and main features – the analytical process is aimed at the use and development of intelligence to direct law enforcement objectives. Data integration represents the first phase, of combining information from different sources to then formulate inferences. The next phase will be interpretation or logical reasoning.

Dissemination: the release of the results of analysis to the client.

When looking at intelligence applied to counter-terrorism, the analysis of the literature[34] highlights different positions. The so-called orthodox school argues that intelligence failures such as 9/11 are unpreventable; even though some improvements are possible, we should assume that surprise attacks will often succeed and thus it's rather more effective to focus on how to deal with their tragic effects. Instead, scholars who are closer to the reformist school, say that the surprise effect (such as on 9/11) has been caused by preventable analytical / exchange failures of the intelligence community and the policy makers- who lacked how to Figure out the full picture and "connect the dots". According to them a reform of intelligence could prevent such failures in the future, and efforts in this direction should be further shared/coordinated within the intelligence community. The third approach, instead, asserts that terrorism presents a dramatically new and different threat against which the current intelligence and security organizations have not been prepared. This is partly because terrorist groups, nowadays, have 'smaller signatures' than the nation-state enemies of the past, and because terrorist practice deception.

Some of the existing recommendations[35] for intelligence agencies, in the effort against terrorism, already include:

- greater attention on tracking terrorist finances and fund-raising techniques;
- increased monitoring of terrorist related Internet sites;
- more emphasis on the broader themes and messages of propaganda and training materials disseminated on the Surface and Deep Web.

This can be considered as a starting point, upon which the DANTE project built and elaborated a set of specific guidelines and recommendations for law enforcement intelligence and investigations on the Web, based on the practical application of the integrated platform developed and tested with the end users in the framework of the H2020-funded project.

The DANTE Platform for data collection and analysis

The DANTE project followed a multidisciplinary approach, where scientists from criminology, counter-terrorism, sociology, law and ethics, as well as technologists and LEAs from different EU countries have been working in synergy to analyse and improve intelligence processes. Ethical, legal, privacy and security issues have been all taken into considerations, with the final aim of developing socio-technical solutions in support of those processes which foster the cooperation and collaboration within and among LEAs. The project outcomes have been thus designed for the main counter-terrorism stakeholders, including analysts, with the aim of supporting the most advanced intelligence processes through big data collection and analysis.

One of the key points of the proposed technological solutions is the knowledge of facts and events in advance, to prevent potential terrorist threats and limit the financing of terrorist activities, through the detection of relevant online contents over the Internet. The proposed solutions, embedded in the DANTE platform, are thus mainly aimed at supporting the automatic detection and analysis of relevant sources and contents in the Surface and in the Deep Web and Dark Net. In fact, one of the important elements of the platform is to innovate and improve the intelligence processes in the dark part of the Internet, which is usually accessible

[34] The main references to the literature for this paragraph can be found in Dahl E.J. and Viola D. (updated 2017), Intelligence and Terrorism, in International Studies. Available at: http://oxfordre.com/internationalstudies/view/10.1093/acrefore/9780190846626.001.0001/acrefore-9780190846626-e-91.

[35] See, among others, Rudner (2005); Williams (2005); Winston (2007); Renfer & Haas (2008); Bouzis (2015). Available at: https://scholarcommons.usf.edu/cgi/viewcontent.cgi?referer=https://www.google.it/&httpsredir=1&article=1020&context=jss.

only with specific clients, such as hidden services in TOR[36] and I2P.[37]

On one side, the detected multimedia contents to be analysed through the DANTE platform, should contain information about activities and events: in fact, one of the goals is to automatically identify and cluster such activities and reconstruct the ‘criminal’ chain. On the other side, through the platform, people, groups and relationships among them will be detected and examined, through the identities’ identification and the understanding of the capabilities and intentions of the individuals – with a special focus on online fund-raising activities, propaganda, training and disinformation.

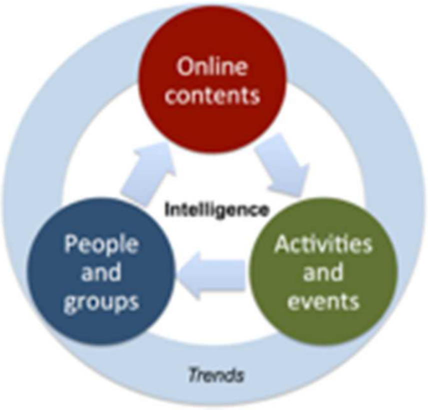


Figure 3: DANTE project rationale

Source: DANTE DoA

As mentioned in the introductory part of this report, due to the transnational nature of the terrorist phenomenon, the *collaboration*-element becomes one of the main factors to be introduced in the so-called ‘intelligence cycle’: a global strategic perspective, exceeding the boundaries of each organization and country, is a necessary element.

In order to provide concrete support to the LEAs across Europe, the project DANTE has designed and developed a set of services and an *integrated platform for automatically detecting, analysing and monitoring a huge amount of terrorist-related online contents and activities*. More specifically, it provides capabilities for source identification and management, content acquisition, pre-filtering and normalisation, textual and multimedia analysis and mining, relevant content storage including for forensic purposes, information fusion and analytics, smart visualization and trend analysis. The DANTE project thus combined techniques, technologies and innovative models for audio analysis, text mining, natural language processing, image analysis, video analysis, social network analysis, human computer interaction, information fusion.

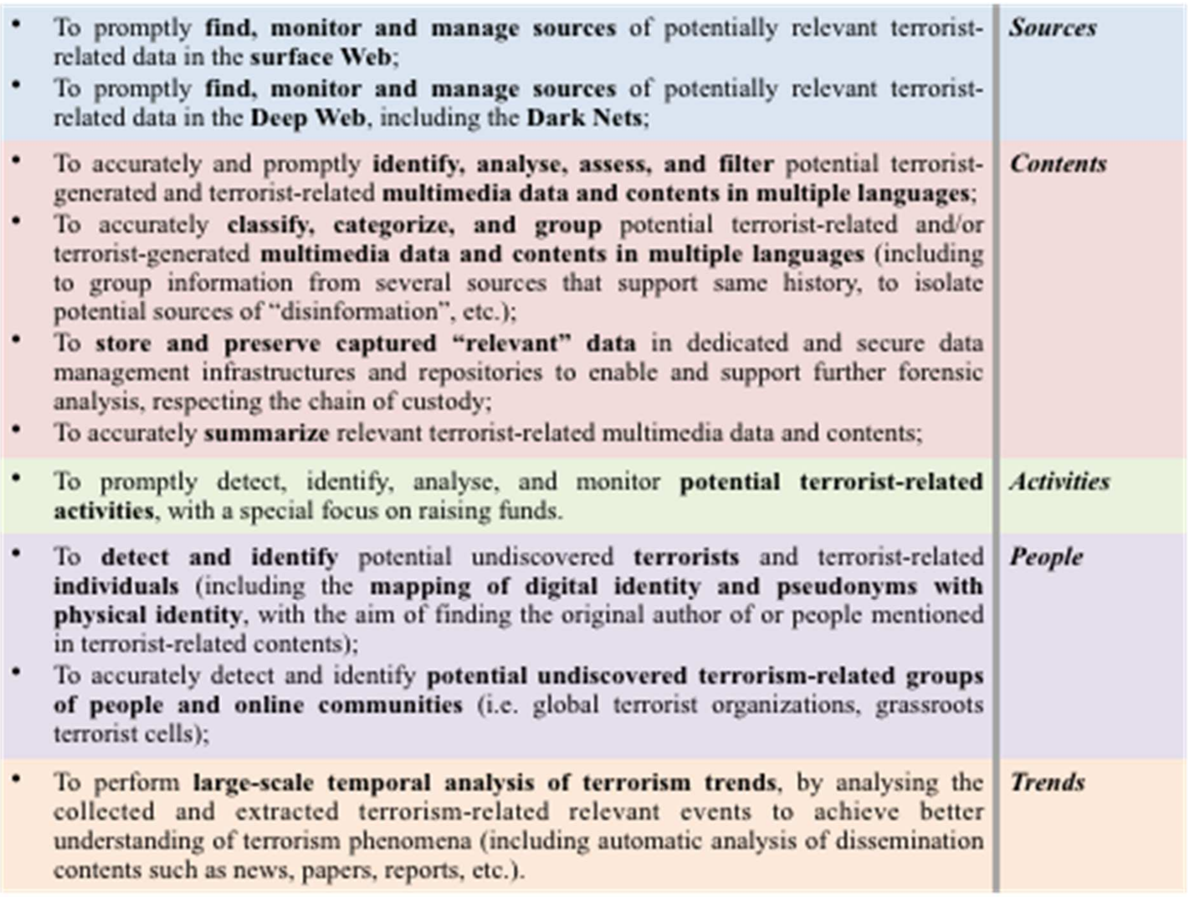
Another important aspect is that the integrated system prototype does actually fulfil the requirements as

[36] According to techopedia.com, The Onion Router (Tor) is an open source software program, that allows users to protect their privacy and security against a common form of Internet surveillance known as traffic analysis. The main idea behind designing Tor was to protect personal privacy of network users and allow them to conduct confidential business. Tor is also widely used in location hidden services to provide anonymity to servers.

[37] According to techopedia.com, I2P is an open source project attempting to create an anonymous network for communication over the Internet. Communication in I2P is encrypted to provide protection and security against attackers and hackers. I2P was originally the short term for Invisible Internet project.

defined by the end users of the system, namely the LEAs both internal and external to DANTE.

The Figure below (Figure 4) summarizes the automated functionalities embedded in the DANTE platform. Moreover, Figure 5 graphically presents the overall logical architecture of the DANTE system.



• To promptly **find, monitor and manage sources** of potentially relevant terrorist-related data in the **surface Web**; • To promptly **find, monitor and manage sources** of potentially relevant terrorist-related data in the **Deep Web**, including the **Dark Nets**;	***Sources***
• To accurately and promptly **identify, analyse, assess, and filter** potential terrorist-generated and terrorist-related **multimedia data and contents in multiple languages**; • To accurately **classify, categorize, and group** potential terrorist-related and/or terrorist-generated **multimedia data and contents in multiple languages** (including to group information from several sources that support same history, to isolate potential sources of "disinformation", etc.); • To **store and preserve captured "relevant" data** in dedicated and secure data management infrastructures and repositories to enable and support further forensic analysis, respecting the chain of custody; • To accurately **summarize** relevant terrorist-related multimedia data and contents;	***Contents***
• To promptly detect, identify, analyse, and monitor **potential terrorist-related activities**, with a special focus on raising funds.	***Activities***
• To **detect and identify** potential undiscovered **terrorists** and terrorist-related **individuals** (including the **mapping of digital identity and pseudonyms with physical identity**, with the aim of finding the original author of or people mentioned in terrorist-related contents); • To accurately detect and identify **potential undiscovered terrorism-related groups of people and online communities** (i.e. global terrorist organizations, grassroots terrorist cells);	***People***
• To perform **large-scale temporal analysis of terrorism trends**, by analysing the collected and extracted terrorism-related relevant events to achieve better understanding of terrorism phenomena (including automatic analysis of dissemination contents such as news, papers, reports, etc.).	***Trends***

Figure 4: Automated functionalities of DANTE system

Source: DANTE DoA

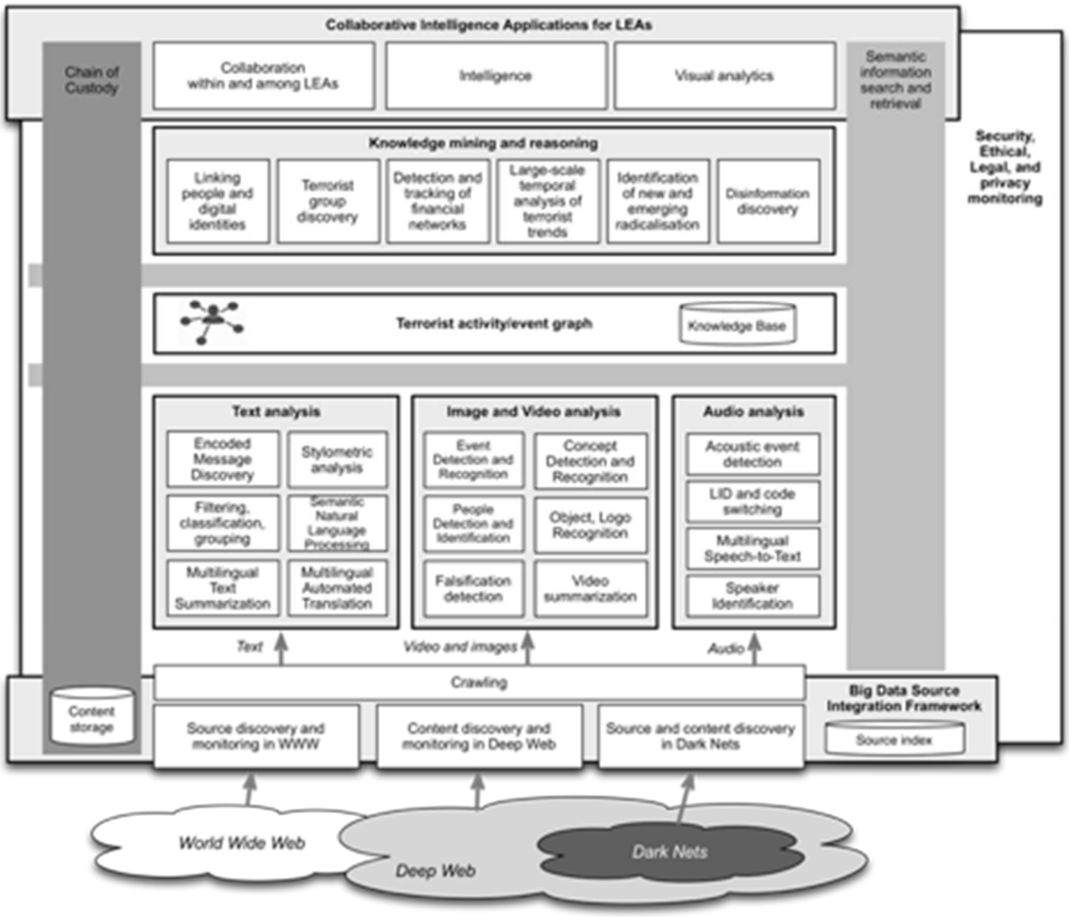

Collaborative Intelligence Applications for LEAs
Chain of Custody
Collaboration within and among LEAs
Intelligence
Visual analytics
Semantic information search and retrieval
Security, Ethical, Legal, and privacy monitoring
Knowledge mining and reasoning
Linking people and digital identities
Terrorist group discovery
Detection and tracking of financial networks
Large-scale temporal analysis of terrorist trends
Identification of new and emerging radicalisation
Disinformation discovery
Terrorist activity/event graph
Knowledge Base
Text analysis
Encoded Message Discovery
Stylometric analysis
Filtering, classification, grouping
Semantic Natural Language Processing
Multilingual Text Summarization
Multilingual Automated Translation
Image and Video analysis
Event Detection and Recognition
Concept Detection and Recognition
People Detection and Identification
Object, Logo Recognition
Falsification detection
Video summarization
Audio analysis
Acoustic event detection
LID and code switching
Multilingual Speech-to-Text
Speaker Identification
Text
Video and images
Audio
Crawling
Content storage
Source discovery and monitoring in WWW
Content discovery and monitoring in Deep Web
Source and content discovery in Dark Nets
Big Data Source Integration Framework
Source index
World Wide Web
Deep Web
Dark Nets

Figure 5: DANTE system, logical architecture

Source: DANTE DoA

The DANTE system is actually based on a service-oriented architecture, in which most of the functionalities are provided as services, with specific Application Programming Interfaces (APIs).[38] In the following paragraphs of the report the functionalities will be briefly described with specific regards to some of the services/modules of the platform, with the objective of outlining the possible concrete applications in the daily activities of counter-terrorism LEAs.

At this stage, it is important to mention that, keeping into consideration the main target readers of this report - *LEAs, policy makers, academics and other potential relevant (also non-technical) stakeholders* - only a number of services/modules of the platform were selected, to be included in the next sections. The selection of the relevant modules was done in close cooperation with the technical developers, by keeping in mind the final aim of this analysis. In fact, this report is not aimed at describing the entire DANTE system from a

[38] According to techopedia.com, an Application Programming Interface (API) is a set of protocols, routines, functions and/or commands that programmers use to develop software or facilitate interaction between distinct systems. APIs are available for both desktop and mobile use, and are typically useful for programming graphic user interface (GUI) components, as well as allowing a software program to request and accommodate services from another program.

technical point of view, but it rather aims at looking from a cross-cutting perspective at the technical potentialities of the platform and the concrete application in the investigation and intelligence field. With the final aim of developing a set of recommendations for improving intelligence and investigation practices in the counter-terrorism field.

Thus, the elaborations herewith presented are the results of a joint assessment, carried out in cooperation with the end users (LEAs) and the technological developers. Most of the inputs from the LEAs have been collected during the pilot[39] sessions- of the DANTE project held in Rome (23-25 October 2018), Lisbon (20-22 November 2018), Madrid (16-17 January 2019) and Rome (18 February 2019), mainly through an external observation of the demonstration of the modules to the end users as well as through interviews carried out by the authors.

The elaboration of the collected feedback has been then shared with the technological partners of DANTE, and thus reflected/incorporated into the final release of the DANTE system. On this basis, a set of practical recommendations and guidelines for LEAs to improve counter-terrorism intelligence on the web have been developed and are presented in the last section of this report (Part III).

Starting the investigation: Crawling the Surface/Deep Web and the Dark Net

Deep web databases are not indexed by any search engines, not densely dispersed and they keep changing as it is dynamic data. Hence, it is challenging to locate these database contents. In order to acquire data from the online environment, as a possible first step of an investigation, the DANTE platform includes a set of crawling mechanisms - both addressing the surface and the deep web, and partly tailored on the social networks.

Developing a professional searching strategy represents a preliminary and crucial step to ensure the quality and relevance of information gathered.

In order to search the Internet professionally, one needs to define:

- where to search (search tools, search engines etc.);
- how to search and how to do it effectively;
- how to do the analytics.

Figure 6, based on AIT elaborations, is a graphic representation of the complexity characterising professional searching over the Internet.

[39] As already explained in the Introduction, the pilot sessions consisted in practical demonstrations to the LEAs of the different modules of the platform, to collect practical inputs for revisions and improvements. For each DANTE use case, 2 pilot sessions have been organized with representatives of LEAs participating in the consortium.

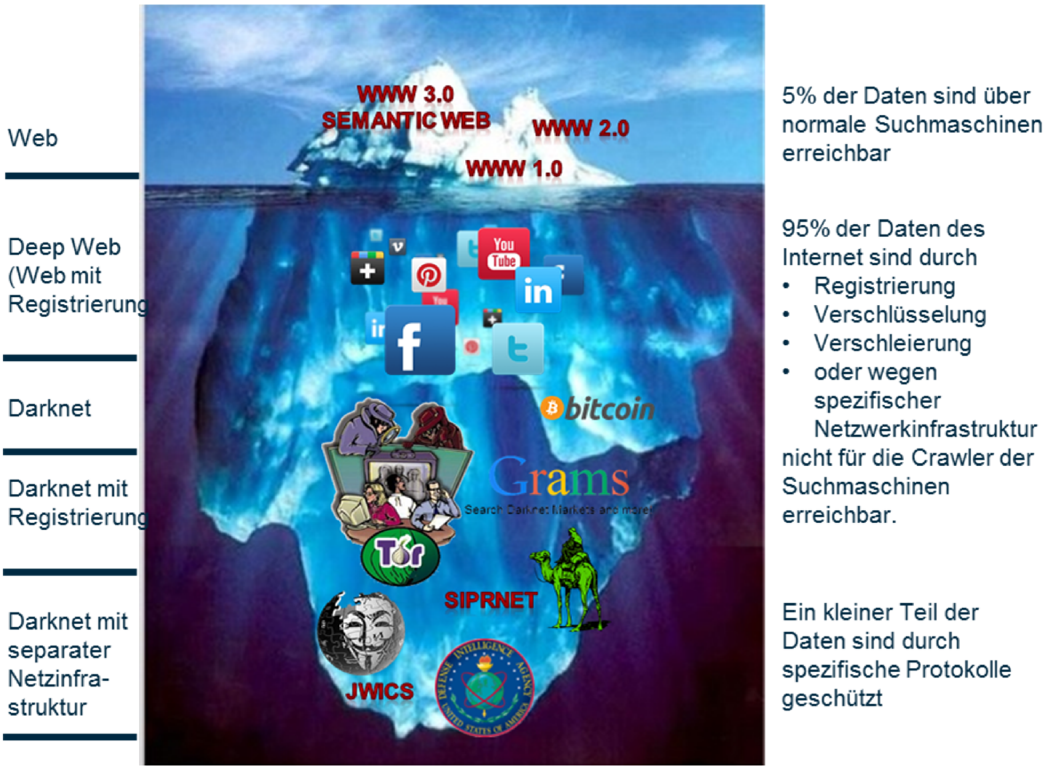

Web
Deep Web (Web mit Registrierung
Darknet
Darknet mit Registrierung
Darknet mit separater Netzinfra-struktur
WWW 3.0
SEMANTIC WEB
WWW 2.0
WWW 1.0
bitcoin
Grams
Search Darknet Markets and more
Tor
SIPRNET
JWICS
5% der Daten sind über normale Suchmaschinen erreichbar
95% der Daten des Internet sind durch
• Registrierung
• Verschlüsselung
• Verschleierung
• oder wegen spezifischer Netzwerkinfrastruktur
nicht für die Crawler der Suchmaschinen erreichbar.
Ein kleiner Teil der Daten sind durch spezifische Protokolle geschützt

Figure 6: Searching the Internet (professionally)

Source: AIT

Using existing search engines can be the starting point to develop crawling strategies.

Based on the social network analysis (SNA) theory, Internet is a network of networks with hubs and authorities. A relevant data set is essential for the identification of terrorist threats: in order to refine the searching strategies, and identify what is relevant, it can be useful to search for terms statistics - as reported in the example below.

The most famous search engine, Google, while indexing the surface web sites also offers search statistics on its results list. Usually, the first 100 results from a Google search offer clues on the quality of the results: if false positives are in this list, it can be expected to have wrong positives also in the results of the crawling activity. In addition to the indexing statistics, Google provides basic data about the quantity of searches, for words and words combination. This information can assist in finding out whether the operation is done with the appropriate terms.

The guiding questions should be: which semantic fields must be considered for the purpose of our crawling? Who uses which terms, where and when?

One example showing the challenges related to the identification of relevant threats, and their combinations, can be the word 'mujahideen' - a term which is used differently according to the regions, and is often used for propaganda purposes within the networks of Islamist radicals and extremists.[40]

[40] The traditional Islamic jurisprudence defines Jihad as struggle or resistance: for many, it can be an internal struggle, to resist temptation or achieve state of spiritual progression. One who engages in this struggle towards self-improvement is referred to as a

Within the framework of DANTE, AIT performed an analysis based on the numbers of queries (indexed) with the term ‘mujahideen’, from 2004 to 2018, for Google search worldwide.

First of all, when searching the word in Google, it appears with a number of different ways of spelling: eg. the term ‘mujahedin’ gives 0.5 million results; when spelled ‘mujahideen’ gives, instead, 5 million results. This can be considered already as an indication of the spelling to be used when launching the crawler. Another way to check if the context of the search was correct is by checking the set of images found by the search engine: on this basis the more appropriate or relevant spelling for the selected word can be identified.

The Google trends tool can help, by inserting the same word with the different spelling, to identify picks: the pick can, for instance, correspond to a set of big terrorist attacks, and so also to huge online discussions on the topic. Figure 7 represents it graphically.

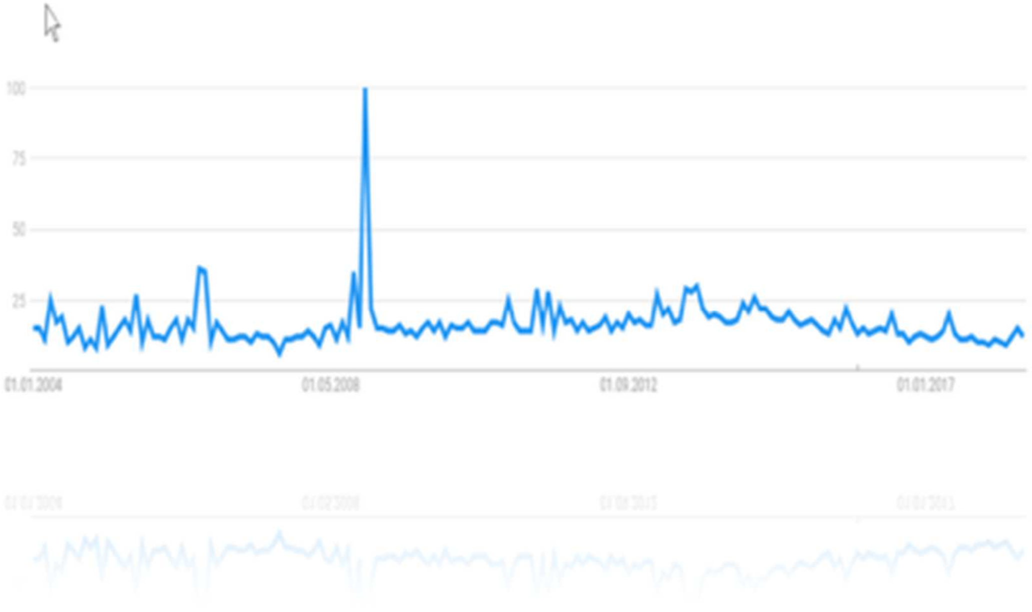


Figure 7: Quality of searches, with the term ‘mujahideen’

Source: AIT – Google

Moreover, the distribution by region can be analysed, to obtain information on the use of languages in different countries (e.g. word ‘mujahideen’ mostly used in Afghanistan and Algeria).

mujahid (plural: mujahideen). The term Mujahid did not possess its current connotation until the 1980s, when thousands of guerrillas fighters knowns as Afghan Mujahideen united forces to resist to the invasion from the Soviet Union forces.

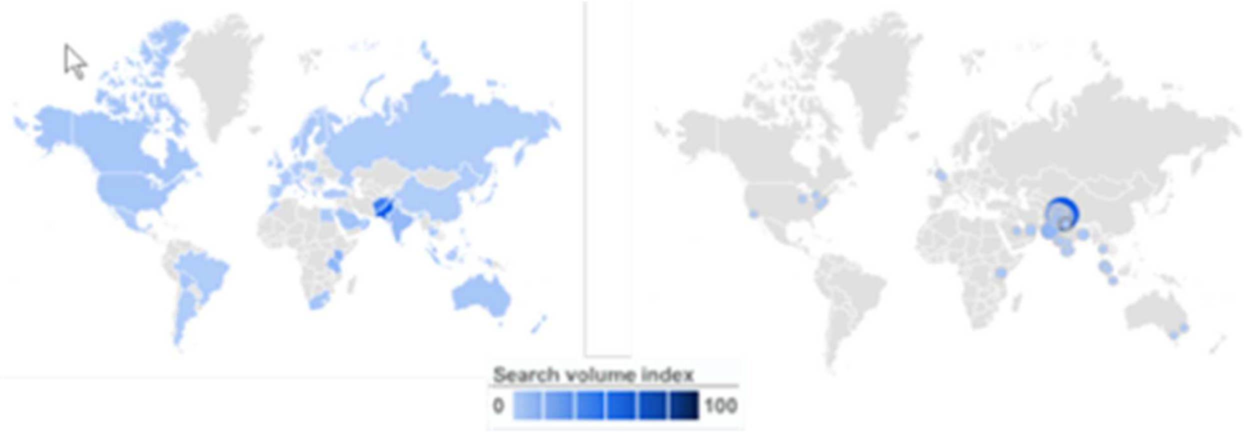


Figure 8: Geographic distribution: searches containing the term 'mujahideen'

Source: AIT – Google

A combination of words, such as mujahideen + wahabism, can give even better results than a search with a single word. In general, by doing this exercise and check the websites when the searched words are actually mentioned, the initial list of websites to be included in the crawler is obtained, which allows to perform a more accurate crawling activity, with (ideally) better results.

The crawler even helps in developing and visualizing networks, through visualization software integrated in the platform. One of the comments collected from the LEAs is that existing software, such as Maltego or Analyst notebook etc., may generate challenges - such as loosing previous information.

Neutrality was explicitly considered in the development of the DANTE crawler. One of the most useful functions highlighted by the DANTE's end users is the possibility, through the crawlers integrated in the platform, to continue the monitoring activities and finding additional results. One can even filter information by domain, including by specific country-domains, or other filters e.g. social media domain, checking links from surface web to Facebook etc.

Some results can even be related to hidden financial networks: among the different religious and diplomatic networks, one can find for instance specific names of mosques often mentioned on terrorist-related content web pages. Further analysis could then start, focusing for examples on why these mosques are specifically receiving funds?

The crawler can even detect bitcoin addresses in the analysed websites; this information will be further used by additional DANTE tools (e.g. Tag sharing and Graph sense modules).

Box 1 – Main crawlers' functions

A crawler is a program that visits websites and reads their pages, and other information, to retrieve that information in other store systems. The major search engines on the Web have this kind of program, which is also known as a "spider" or a "bot".

A web site shows everything to its users; consequently, a crawler can also read all its contents store them. The extent and quality of the code libraries and the open source frameworks exploded over the past decade. This especially concerns the technologies covering the fundamental principle of the open source movement: free and easily accessible data. This development affects all elements of a crawler.

There are analytics tools that allow requesting a search result, without even loading the entire page. The downloader can decide, while reading the data on a website, if it is worth storing through advanced parsers filter the stored content almost in real-time, and turn unreadable into structured re-usable data such as excel (XML) or JavaScript Object Notation (JSON) formats.

The process that automatically recollects data or information from the WWW is called Web Scraping. It’s a field particularly developed in these moments, with similar goals to the semantic web application, an ambitious initiative that still requires breakthroughs in text processing, semantic understanding, artificial intelligence and human-computer interactions.

All the web scraping process could be done by one of the next approaches that are explained next.

The Human interaction is based on the interaction by humans with the web page. In this case, an individual must copy all the content of the information manually and save it in other storage systems. This task is very time-consuming. Also, in many cases, this is the only way to recollect as some webs have barriers to forbid the automation of data extraction in the context of the web page. Text pattern recognition could also extract information from a web page, this approach is very helpful to extract all the information from webs that contains specific words but doesn’t differentiate the context of one word, so one could be extracting information that does not want.

3.3 First level input - basic mono modal technologies

Audio analysis

The Figure below (Figure 9) provides an overview on the different steps, considered in the DANTE platform, for the audio analysis: it presents a high-level use of these technologies in the context of a tool to help humans analysing huge quantities of audio-visual data. The process is a continuous one, as it builds upon the knowledge-based intelligence, to end up with producing additional intelligence after a set of analysis. The audio (and, in some cases, video) material is obtained through the crawling of both the surface and the deep web. Focusing on the audio material, the following functions will be described in the next sections: speech-to-text transcription, speaker identification and audio event detection.

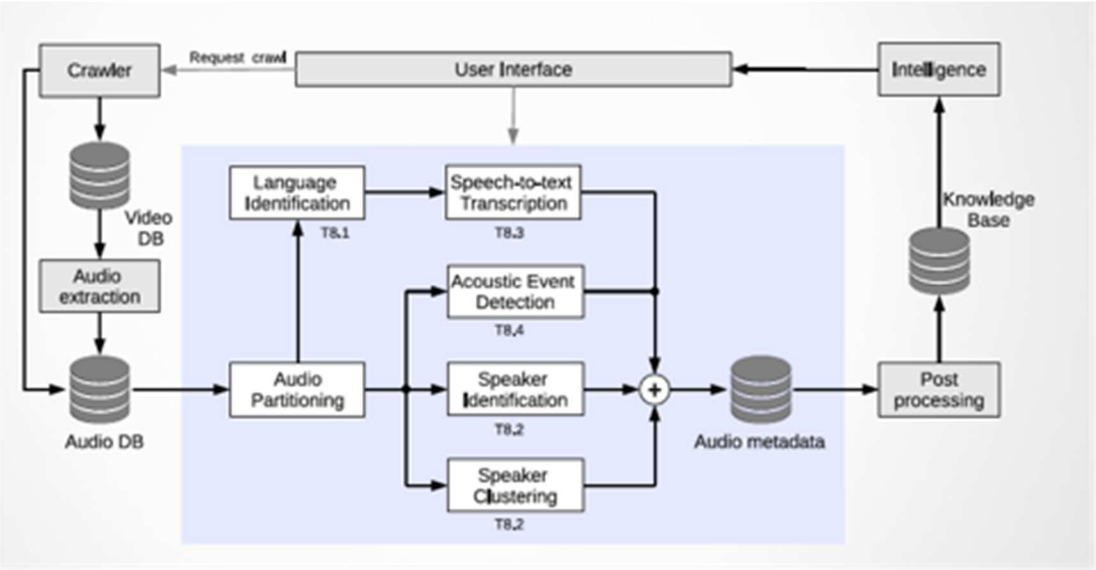


Figure 9: Audio-video analysis process

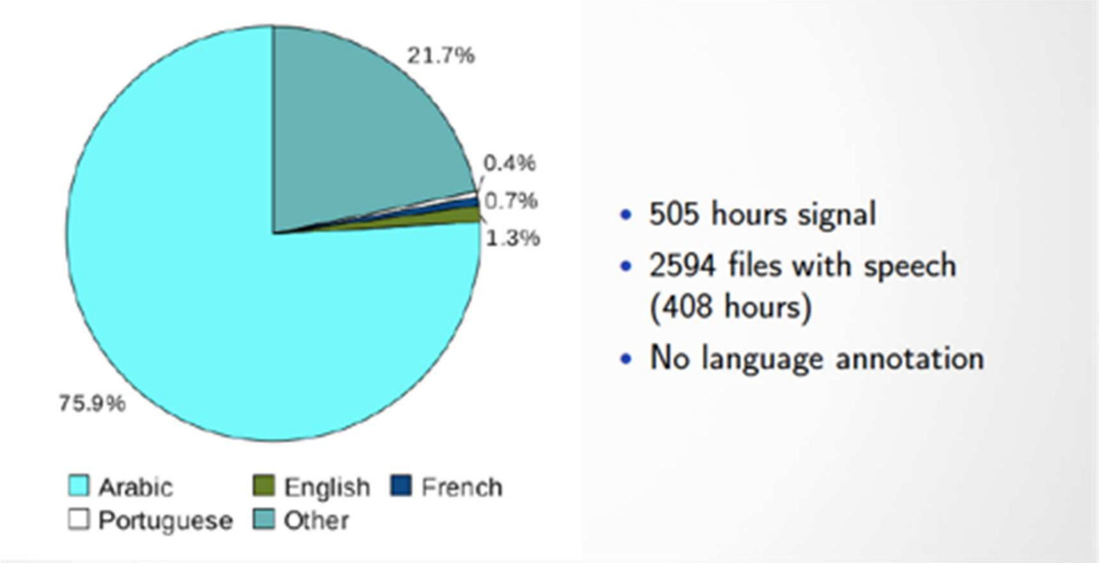
Source: Vocapia, DANTE WP8

Language identification - Multilingual speech to text

A preliminary step to multilingual speech to text analysis is the automatic language identification (LID), which can be speech or text based. LID systems perform automatic detection of the spoken language(s), using the characteristics of the speech signal. Some of the sources for application of LID include, for speech-based LID, automated dialogue systems, call routing, call centres, household devices, consumer products, wiretapping etc.

As outlined in the previous module, language recognition and identification are different than language verification: identification systems generally seek to identify an unknown language, while verification systems usually compare the analysed language to a reference one.

A preliminary test for language identification has been developed by Vocapia, based on a set of audio files provided by one of the end users of the project (Guardia Civil). The composition of the data set is represented below:



- 505 hours signal
- 2594 files with speech (408 hours)
- No language annotation

Figure 10: LID on dataset provided by DANTE end users

Source: Vocapia

There are of course a number of challenges in the process of language identification, as in the same sentence the speaker can use two or more different languages. The tool is able to apply the so-called code-switching, which refers to the process of switching from one language to another in the same written or oral conversation. The switching can be inter-sentential, when the switch from one language to another takes place at a sentence boundary; or intra-sentential, when it takes place within a sentence boundary.

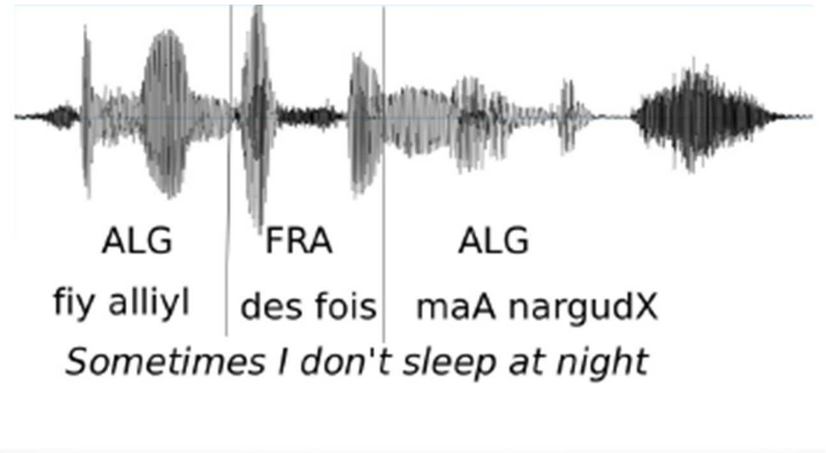


Figure 11: Example of intra-sentential switching

Source: Vocapia

Thanks to the DANTE platform, it will be possible to find audio files also through the crawling functions.[41] It will be possible to obtain: transcription; automatic translation; playing the file and visualize the words found in the speech at the same time. At the moment, the tool can work on six languages (English, Spanish, Portuguese, Italian, French and Arabic), but more languages could be added.[42]

The end users, during the pilot demonstrations of the module, highlighted the key element of the tool working with multi-languages: for the topic of terrorism, in particular, more languages would be useful, such as Russian or German. Transcription in Arabics are of great importance, as nowadays most LEAs face challenges in understanding and recognizing Arabic for instance from Morocco, or from Egypt etc.

The materials analysed within Dante, being mostly related to propaganda, were in standard Arabic – as the final aim is to make the messages understandable by the highest number of people. In terms of analysis of interpersonal communications or exchanges, the possibility to expand the tool to the analysis of dialects would be recommended.

One of the tricky aspects highlighted by the end users is the importance of the context of the speech. Some groups may create their own code language: end users from Portugal provided the example of radicalized foreign people in their country, who actually learnt Portuguese from the comics. In this regard, another interesting feature would be to instruct the tool with these specific 'coded-languages'[43], to be then automatically detected and analysed.

In the context of Dante, the module has been so far applied to propaganda material (as propaganda was one of the use cases identified, together with training and terrorist financing) – which is usually less complex than wiretapping for instance, an aspect to be possibly expanded. Telephone conversations are indeed more complex for ASR tasks, including for privacy issues when sharing the wiretappings to be analysed.

One feature of the tool is that all its functionalities can be used both online and offline. Moreover, the users have the possibility to manually modify the transcriptions, if any error is noticed. As a general remark, the analysis of open sources materials (such as propaganda speeches detected online) is generally more useful for trends analysis or strategic analysis of the phenomena and the groups spreading messages online. While investigators dealing with specific cases usually work with specific information and closed-data collected during the investigation phases. At the same time, the detection of languages per se is useful for all types of crime, and not only for terrorism: when facing with a case of robbery, for instance, the automated detection of the intercepted exchanges among the suspects involved would turn out to be helpful in understanding their possible country of origin.

Speaker identification

Speaker identification, or recognition, consists in the identification of a person from the characteristics of his/her voice.[44] In general terms, identification is the task of determining an unknown speakers' identity: it is

[41] With regard to propaganda material, it was highlighted that Europol already collect propaganda speech, and detains a dataset of all these speeches already transcribed.

[42] These languages were selected in the initial phase of the project, in agreement with the end users, and based on their countries of origin, and in correspondence with the text analysis tools. However, the partner responsible of this module - Vocapia - offers transcription systems for open-source data in nearly 30 languages, which could be added to the platform.

[43] This is also similar to what is called 'key word spotting' or 'spoken term detection', which can be used to flag documents that may be pertinent to the user, and without necessarily needing a full transcript.

[44] Poddar A., Sahidullah Md, Saha G. (2018), Speaker verification with short utterances: a review of challenges, trends and opportunities". Available at:https://ieeexplore.ieee.org/document/8302747.

a 1:N match where the voice is compared against a certain amount of templates.

"In the criminal field, the experts for speaker identification and audio analysis are usually called at a very early stage of the crime […] In many cases the analysis of acoustic material is extremely important in order to gather the decisive clues needed to solve a criminal offence and to prevent follow-up offences".[45]

The recordings can be found on items of evidence seized (e.g. cassettes from answering/dictating machines etc.), as well as on digital sound storage media containing under-cover recording or recording from telecommunication surveillance. The 'lawfully' intercepted sounds, including ambient conversation, can thus be matched with voices gleaned from phone or social media and against a 'blacklist' database –managed by police forces. Samples could come, for instance, from mobile, landline or voice-over Internet protocol (VOIP) recordings, or even from snatches of audio captured from recruitment or propaganda video posted on social media.[46] Moreover, the captured voice clip, may include some descriptive metadata – sometime added by LEAs.

Within the DANTE WP8 on audio-video analysis, a dedicated module on speaker recognition has been developed. Below, the results of the analysis developed by AGNI are summarized. Moreover, a set of general comments on the module's features collected among the end users (LEAs) at the pilot session held in Lisbon in November 2018 are presented: the final aim is to elaborate on the practical applications and further developments of this module, in relation to the overall DANTE platform.

In the identification scenario (1:N)-, one audio has been selected with the aim of finding matches in a set of *N audios*. The test was divided into two scenarios, one referring to a closed group case – in which there are audios from the target speakers included in the list; and another one related to an open group, in which there might be no audios linked to the target speakers. Two sizes of N have been considered, to explore the different results: N=10 and N=50.

The graphics below show the comparison of final results when considering open group and closed group. A threshold value can then be found, for which the size of the open group is zero and for closed group is one.

In more practical terms, establishing a threshold means defining a list of candidates. The results shown in the graphics below has to be considered as a way of filtering: the tool is then filtering, from a number of audio, which ones have the highest percentage to contain the voice of one of the models (eg. suspects), thus facilitating and speeding up the analysis process for the LEAs.

[45] Bundeskriminalat, The extortionist's voice, article available at: https://www.bka.de/EN/OurTasks/SupportOfInvestigationAndPrevention/ForensicScience/PhysicalEvidence/Extortion/SpeakerIdentification/speakeridentification_node.html.
[46] Dumiak M. (2018), Interpol's new software will recognize criminals by their voices", available at: https://spectrum.ieee.org/tech-talk/consumer-electronics/audiovideo/interpols-new-automated-platform-will-recognize-criminals-by-their-voice.

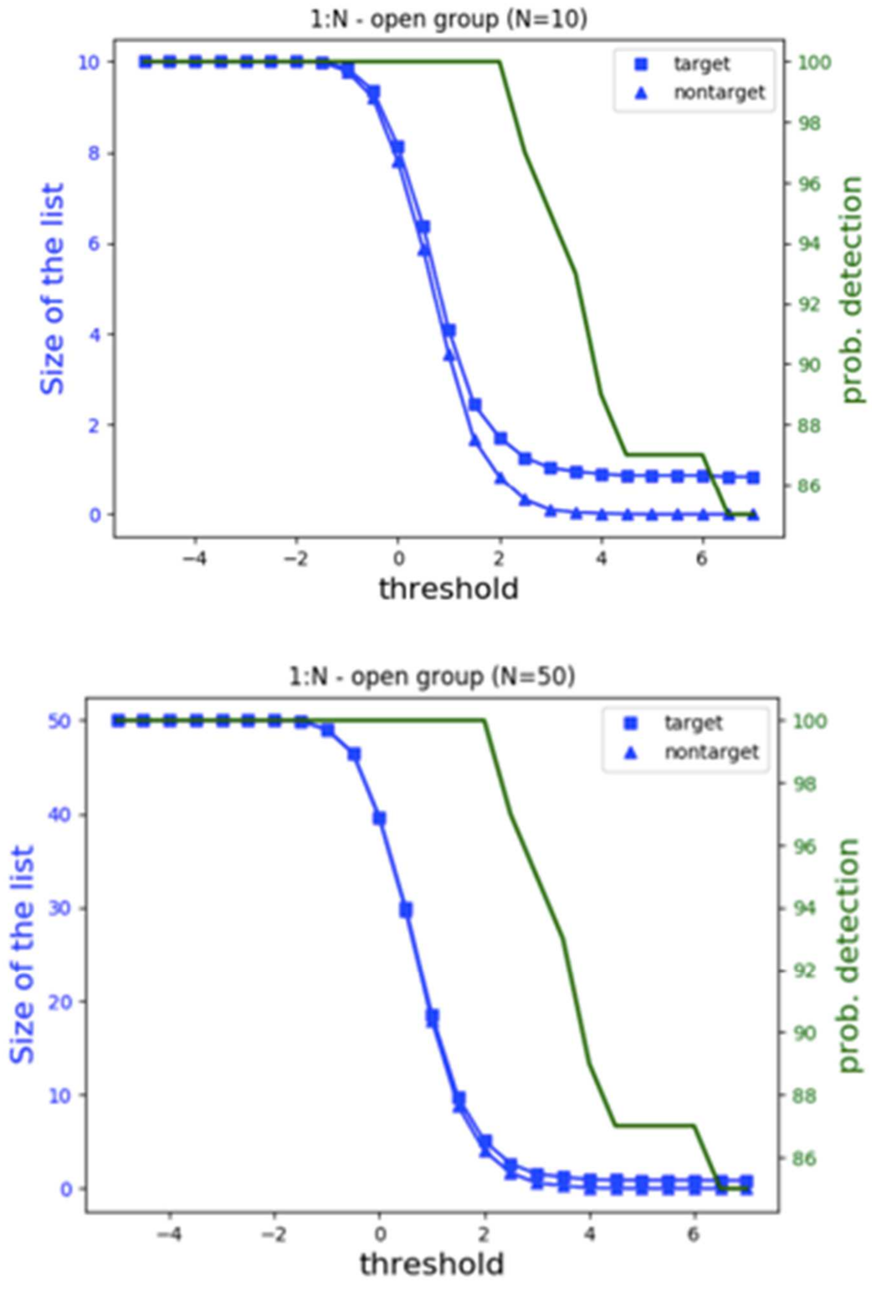


Figure 12: Threshold values

Source: AGNITIO

One of the useful elements highlighted by the LEAs during the pilot sessions is that the tool can work on any audio source and with basically every format: the system will then extract the characteristics of the audio. Depending on the quality of the file, the tool can get rid of the most part of the background noise – improving its quality.

The tool works independently from the type of channel, gender or language of the speaker. The system identifies the voice print regardless of the channel effect.

As outlined in the previous paragraphs, the tool provides a score, on the basis of a threshold and according to the database analysed. Thresholds do vary according to such datasets. The results do not represent a percentage, but rather a score above or below a certain threshold, indicating the probability of a correct identification of the speaker. Scores are normalized, so that it should not be difficult to set a reasonable threshold; however, testing the accuracy of the system within domain data would help finding a valid threshold.

So far, the main functionality of the system is to filter the audio files included in the analysed dataset, and does not involve any following steps with regard to the forensic use of the files.

An interesting element to be considered, when all the modules here described are finally integrated in the DANTE platform, is the possible link between this module and the stylometric analysis: based on this initial filtering, further analysis can be developed on the wiretappings contents through stylometry, with the final aim of authorship identification or even profiling (see additional elaboration in the Recommendation section).

Acoustic event detection

"Acoustic event detection (AED) is the task of automatically recognizing different types of sounds (whether impulsive, continuous or intermittent) that can be of interest in an audio signal. AED can be used on its own, or in association with other technologies, bringing complementary information to automatic video analysis, for example".[47]

The level of accuracy of AED, nowadays, is not so high, and further research in this field needs to be done. While showing to the end users an example of propaganda video with a song and a gunshot as a background noises, the analysts highlighted that an ideal function (not possible at this stage) would be for the tool to detect the noises based, first, on the clearest ones, and then to the less clear ones. For example, it would be useful for the LEAs to be able to differentiate, in a propaganda video, when the voice - for instance in Arabic - is only a song with no music background, rather than a voice reciting the Quran. As this is done differently from one country to another, it could be helpful for the end users to detect the differences - also in combination with the multilingual speech transcription function. Another interesting function (not yet available) could be detecting, in an audio file with TV sounds in the background, which program is being projected, in order to collect relevant metadata.

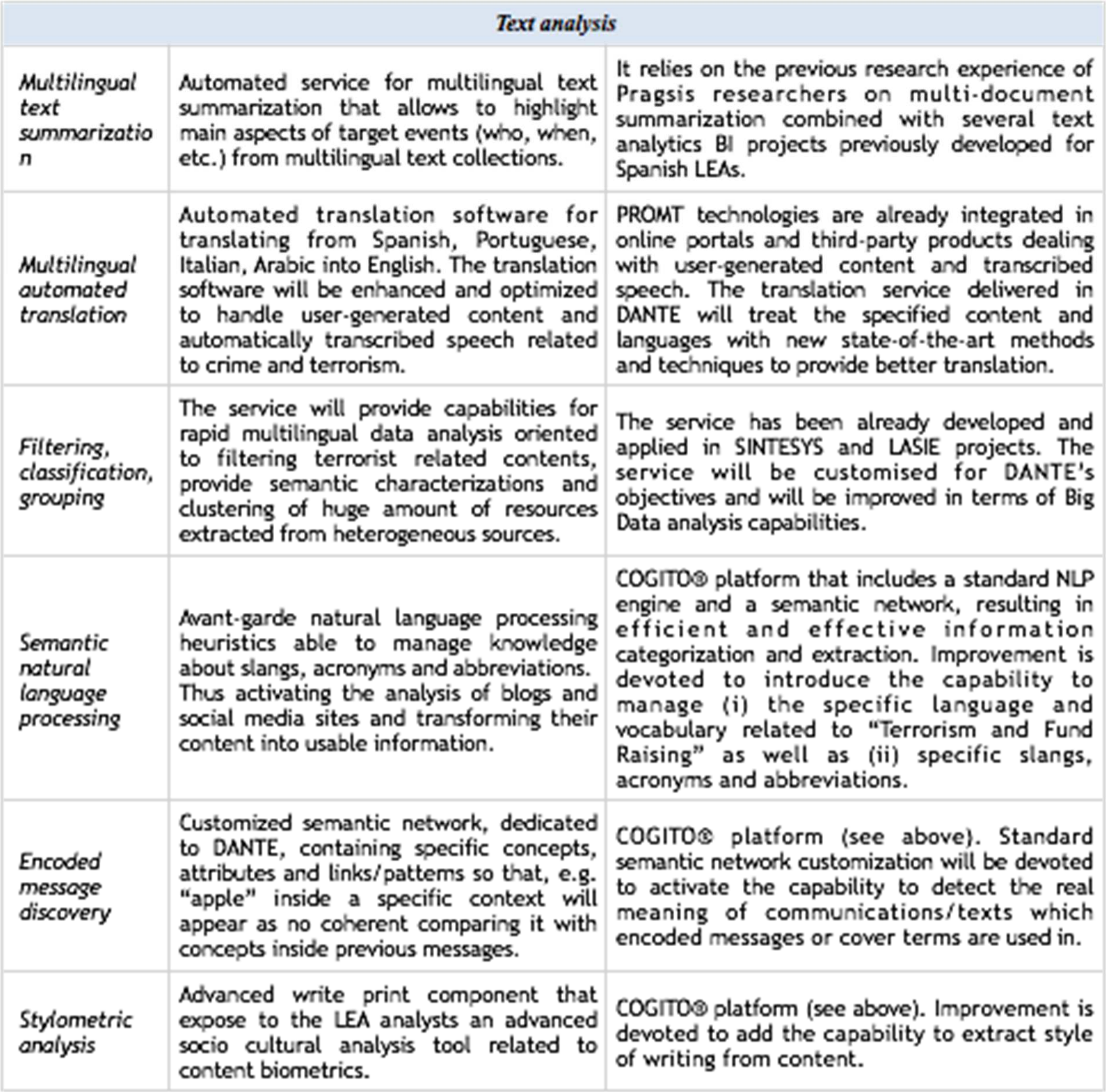
One of the main issues highlighted by the end users, which is actually recurring, is that in the propaganda videos realized by foreign terrorist fighters (FTF), they don't speak the classic Arabic, and therefore it becomes difficult to understand all the details recorded.

Text analysis

The following set of modules and services relate to text analysis.

Within the DANTE system, the following modules related to text analysis have been developed.

[47] Gauvain J., Lamel L., Bac Le V., Despres J., Gauvain J-L., Messaoudi A., Vieru B., Kheder W.B. (2018), Challenges in audio processing of terrorist-related data. Available at: https://link.springer.com/chapter/10.1007%2F978-3-030-05716-9_7.



Text analysis		
Multilingual text summarization	Automated service for multilingual text summarization that allows to highlight main aspects of target events (who, when, etc.) from multilingual text collections.	It relies on the previous research experience of Pragsis researchers on multi-document summarization combined with several text analytics BI projects previously developed for Spanish LEAs.
Multilingual automated translation	Automated translation software for translating from Spanish, Portuguese, Italian, Arabic into English. The translation software will be enhanced and optimized to handle user-generated content and automatically transcribed speech related to crime and terrorism.	PROMT technologies are already integrated in online portals and third-party products dealing with user-generated content and transcribed speech. The translation service delivered in DANTE will treat the specified content and languages with new state-of-the-art methods and techniques to provide better translation.
Filtering, classification, grouping	The service will provide capabilities for rapid multilingual data analysis oriented to filtering terrorist related contents, provide semantic characterizations and clustering of huge amount of resources extracted from heterogeneous sources.	The service has been already developed and applied in SINTESYS and LASIE projects. The service will be customised for DANTE's objectives and will be improved in terms of Big Data analysis capabilities.
Semantic natural language processing	Avant-garde natural language processing heuristics able to manage knowledge about slangs, acronyms and abbreviations. Thus activating the analysis of blogs and social media sites and transforming their content into usable information.	COGITO® platform that includes a standard NLP engine and a semantic network, resulting in efficient and effective information categorization and extraction. Improvement is devoted to introduce the capability to manage (i) the specific language and vocabulary related to "Terrorism and Fund Raising" as well as (ii) specific slangs, acronyms and abbreviations.
Encoded message discovery	Customized semantic network, dedicated to DANTE, containing specific concepts, attributes and links/patterns so that, e.g. "apple" inside a specific context will appear as no coherent comparing it with concepts inside previous messages.	COGITO® platform (see above). Standard semantic network customization will be devoted to activate the capability to detect the real meaning of communications/texts which encoded messages or cover terms are used in.
Stylometric analysis	Advanced write print component that expose to the LEA analysts an advanced socio cultural analysis tool related to content biometrics.	COGITO® platform (see above). Improvement is devoted to add the capability to extract style of writing from content.

Figure 13: Modules on Text analysis

Source: DANTE DoA

"Text analysis is about parsing texts, in order to extract machine-readable facts from them. The purpose […] is to create structured data out of free text content. The process can be thought of as dicing heaps of unstructured, heterogeneous documents into easy-to-manage and interpret data pieces".[48] One of the challenges related to text analysis is the ambiguity of human languages; others can be related to the original language of the analysts and that of the text to be analysed (e.g. whether or not the analysts read Arabic, Chinese, Dutch, Spanish, French etc.). Similar expressions to text analysis can be text mining, text analytics,

[48] For additional information, see for instance: https://www.ontotext.com/knowledgehub/fundamentals/text-analysis/.

or information extraction.[49]

In order to explore some concrete applications of text analysis to the online investigation and intelligence process, a more in-depth assessment of one of the most interesting modules is presented below, namely ‘stylometryic analysis’. The assessment includes, after an introductory theoretical framework, some of the inputs collected from the LEAs on the main challenges and opportunities of exploiting this tool in counter-terrorism activities - as well as the reference to related modules of the platform.

Stylometric analysis

According to the Oxford dictionary, ‘stylometry’ is the statistical analysis of variations in literary style between one writer or genre and another. A stylometry analysis service is, therefore, a system capable of predicting the personal traits of age and gender belonging to the author of a written text, whose true identity is unknown. This is performed by means of extracting linguistic parameters (called ‘stylemes’) from the text itself and deemed relevant for the authorship identification.

For the purpose of the DANTE project, the main objective of this module is to obtain a web service for submitting a text and receiving a prediction on the above-mentioned information and concerning the author.

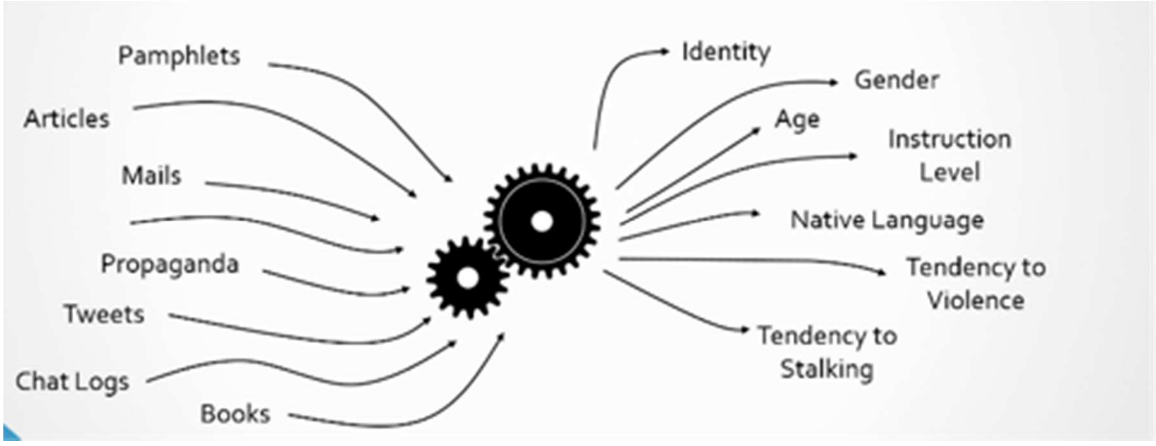


Figure 14: Possible areas of application for stylometry

Source: ESI

In principle, in order to recognize the author (person), the system needs to be trained with written texts, whose authors are the candidate suspected ones. At the current state of the art, the system only allows identifying the age and gender of the authors. However, some additional features, to be possibly explored in the future, include age, instruction level, native language, tendency to violence etc. (Figure 14).

The workflow of the stylometric analysis consists in two separate packages (Figure 15). The first one is based

[49] For additional information, see for instance: https://www.ontotext.com/knowledgehub/fundamentals/information-extraction/

on linguistic, as illustrated on the left side of the Figure below.

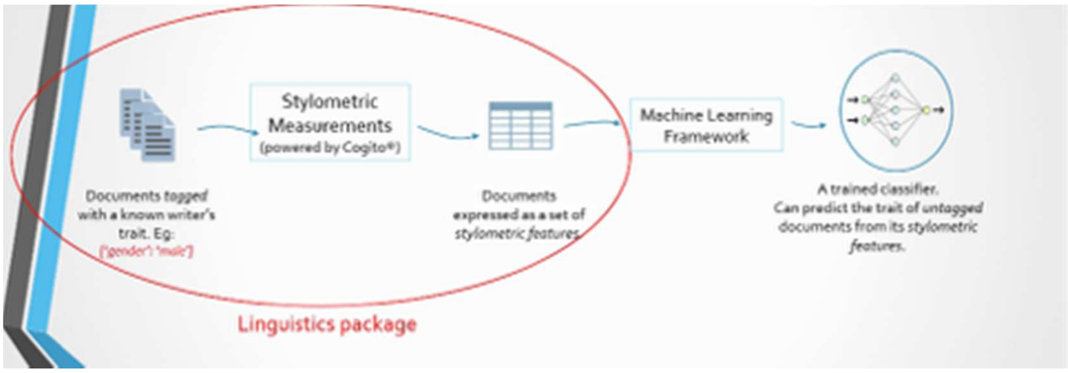

Stylometric Measurements
(powered by Cogito®)
Documents tagged with a known writer's trait. Eg:
{'gender': 'male'}
Documents expressed as a set of stylometric features
Machine Learning Framework
A trained classifier.
Can predict the trait of untagged documents from its stylometric features.
Linguistics package

Figure 15: Linguistic and machine learning hybrid approach

Source: ESI

At this stage, COGITO® provides an exceptional stylometric picture embedded in the text. The result is a statistic representation of the stylemes as authorial DNA (Fig 16).

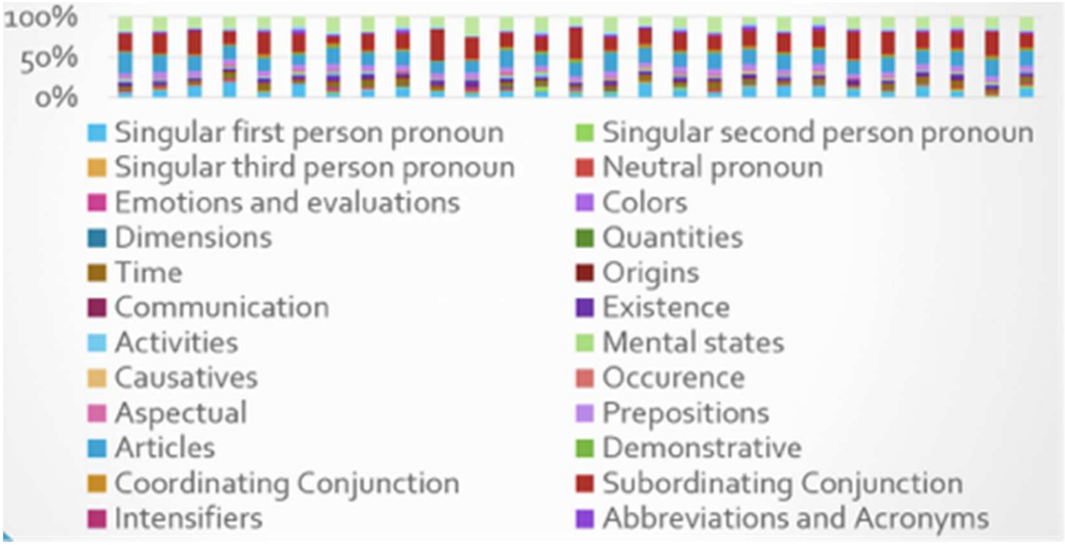
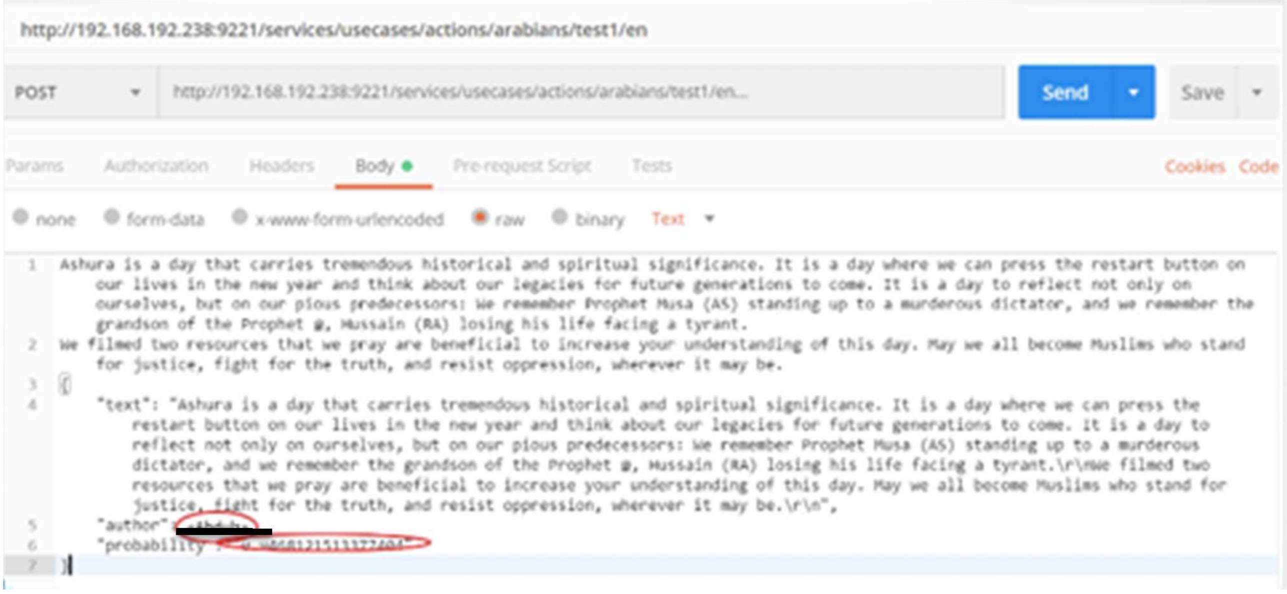


Figure 16: Stylemes as authorial DNA

Source: ESI

We can distinguish between two types of stylemes: quantitative (n° words in the sentence, length of the text etc.) and qualitative (abbreviations, metaphors, euphemism etc.). Even the absence of specific traits can be distinctive for detecting an author in comparison to another.

Within DANTE some initial tests have been conducted on authorship identification, based on speech transcription, starting from a number of sources provided by the DANTE LEAs, together with relevant information concerning the suspect terrorists (authors). The corpus of transcriptions contains 117 documents, divided into files which correspond to 7 authors; the information provided have been used to train a classifier, with the aim of recognizing the different styles and being able to make predictions on the authorship of the texts. The algorithm used was multilayer perceptron. Below, the preliminary results of one use-case on Arabic authorship identification is provided: the analysis of transcription allows to predict which author, among the 7 previously filed, is the original author of the text – with a certain % of certainty.


http://192.168.192.238:9221/services/usecases/actions/arabians/test1/en
POST
http://192.168.192.238:9221/services/usecases/actions/arabians/test1/en...
Send
Save
Params
Authorization
Headers
Body
Pre-request Script
Tests
Cookies
Code
○ none
○ form-data
○ x-www-form-urlencoded
● raw
○ binary
Text

```
Ashura is a day that carries tremendous historical and spiritual significance. It is a day where we can press the restart button on
    our lives in the new year and think about our legacies for future generations to come. It is a day to reflect not only on
    ourselves, but on our pious predecessors: We remember Prophet Musa (AS) standing up to a murderous dictator, and we remember the
    grandson of the Prophet ﷺ, Hussain (RA) losing his life facing a tyrant.
We filmed two resources that we pray are beneficial to increase your understanding of this day. May we all become Muslims who stand
    for justice, fight for the truth, and resist oppression, wherever it may be.
{
    "text": "Ashura is a day that carries tremendous historical and spiritual significance. It is a day where we can press the
        restart button on our lives in the new year and think about our legacies for future generations to come. It is a day to
        reflect not only on ourselves, but on our pious predecessors: We remember Prophet Musa (AS) standing up to a murderous
        dictator, and we remember the grandson of the Prophet ﷺ, Hussain (RA) losing his life facing a tyrant.\r\nWe filmed two
        resources that we pray are beneficial to increase your understanding of this day. May we all become Muslims who stand for
        justice, fight for the truth, and resist oppression, wherever it may be.\r\n",
    "author": [redacted]
    "probability": 0.9868121511327404"
}
```

Figure 17: Use case: authorship identification (Arabic language)

Source: ESI

Looking at the possible applications of this module to the police activity, two main areas have been highlighted: authorship identification and profiling.

The authorship identification has been tested both on written texts and on transcriptions. In order to work and recognize the authorship of a text, the machine needs to be instructed with previous texts or speeches (transcribed) of the author himself. This has to be done separately, as the style of writing and speaking is quite different for an individual. The application on transcriptions (from oral conversations) is relevant if applied to the analysis of wiretappings: in this context, the opportunities of linking stylometry with the DANTE platform's modules on speaker identification and multilingual speech to text will be highlighted in the next sections.

Moreover, when previous materials from the authors are not available, and thus the machine cannot receive the relevant instructions, a profiling activity can be still performed. The possibility to identify, at least, the gender and age of the suspects can work as an additional filter among a wider group of suspects.

In more general terms, one of the main opportunities provided by stylometry in the field of crime prevention and contrast (not only terrorist-related), is the possibility to counter anonymity: applying stylometric analysis technique could help identifying individuals, based on textual traces.

Even though a significant amount of literature has been produced, showing high accuracy rates for long documents, it is still challenging to identify accurately authors of short unstructured documents, in particular when dealing with large authors populations.

LEAs, governments and security agencies, financial institutions etc. are generally interested in tracing the identities of suspects, for instance, through their online conversations. When moving to the Dark Web, it's generally known that participants usually have multiple accounts on various platforms: "being able to validate that multiple accounts on different Dark Web forums belong to the same person with high enough confidence allows us to combine various scattering pieces of information into a more concrete and advanced form of knowledge". And thus, to be used for both investigation and intelligence purposes.

There are as well critics addressing the use of stylometry in detecting suspects or criminals: according to some authors[50], the demerits associated with the technique is more compared to the advantages and the rating made in terms of accuracy would not satisfy the required range of credibility.

More research needs to be done to explore both challenges and opportunities related to this tool, including through the DANTE platform.

Two other modules in support of text analysis are briefly presented in the boxes below

Box 2 - Multilingual Textual Content Summarization

The text summarization service provides an extractive summarizer, able to significantly reduce the length of a given text without missing the key points of the overall meaning, by means of choosing the most relevant sentences from the original document. This tool allows to extract a concise summary from any textual terrorism-related document in different languages covered by DANTE (English, Spanish, Portuguese, Italian), generating an abstract of multilingual collections of documents related to a concrete event or topic. The final goal of this service is to help LEAs to save time providing a concise and informative summary of long documents in order to reduce the information overload and quickly determine which documents are worth reading

Box 3 - NLP information extraction

This module is partly related to the stylometric analysis, previously described. In fact, through the tool integrated in the DANTE platform, the end user can open a text file (e.g. selecting an article from a magazine related to ISIS propaganda) and immediately get metadata on when the text was written. If the document is interesting for the investigation, the user can trigger the analysis, by applying the entity recognition (on everything that is named and you can detect in the text), the text classification (on the main topics included in the text), the stylometric analysis (as described above), the automatic summarization (Box 2), the multilingual automated translation, in case the doc was uploaded in different languages.

Based on the initial testing, as well as on the quality of the text and of the crawling activity previously launched, the level of accuracy can be up to 80-90%. Some risks of 'counter-action' by terrorists can be related to this tool, however: if the terrorist group knows, for instance, which are the key words automatically detected by the program, they could replace these words considered 'dangerous' or 'at risk' with something else, in order to avoid detection. A related tool trying to detect if the text has been encoded is also under development.

[50] Maurice D. (2015), New threats and countermeasures in digital crime and cyber terrorism, IGI Global.

Based on the feedback collected from the LEAs, some additional features could be implemented - including in a follow up project: as DANTE is mainly focusing on Islamist terrorism, one of the main languages detected and translated is (classic) Arabic, and is only translated into English. However, in regions that might be affected by local forms of terrorism, local languages such as Basque or Catalan should be covered, or even dialects, and translations into Castellan would be preferred in these specific cases.

Another useful feature is that the interesting words automatically highlighted in the text, are then linked to the semantic network already identified - also useful to analyse the documents. The application should be also used to instruct the crawler, and automatize the filtering action: by doing this, one user could analyse more than 4,000 documents in one day, and all the categories of interest could be used as filters.

Video-images analysis

In this specific section of the platform, the aim of the DANTE project is to provide a set of tools for analyzing large volumes of online video contents regarding terrorist-related activities on the Surface/Deep Web and Dark Nets. Due to the nature of the analyzed material, partly deriving from the crawling activity or from datasets provided by the DANTE end users, the content investigated is diverse, and a number of challenging aspects are also related to the low quality or the presence of artefacts in the obtained videos.

The ultimate goal of the visual analysis modules is to perform automatic analysis of the vast amounts of collected data, that are made available to the analyst, and hence significantly boosting the investigation procedure.

Below, a description of some of the developed video/images analysis modules developed and tested within DANTE is reported.

Concept detection and recognition (image/video)

The goal of this module is to detect a wide range of high-level semantic entities that may be present in the visual medium and which are of interest to the human user. These entities may correspond to many different levels of semantic granularity or abstraction, e.g. ranging from specific object types to individual scene categories; hence, making their robust detection a challenging task.

More concretely, this tool can provide an automatic categorization of large-scale terrorist-related content. For instance, in a video representing an explosion, the texture of the ‘smoke’ is the prevalent element / pattern, based on which the algorithm will find similar photos, as presented in the sample images below.

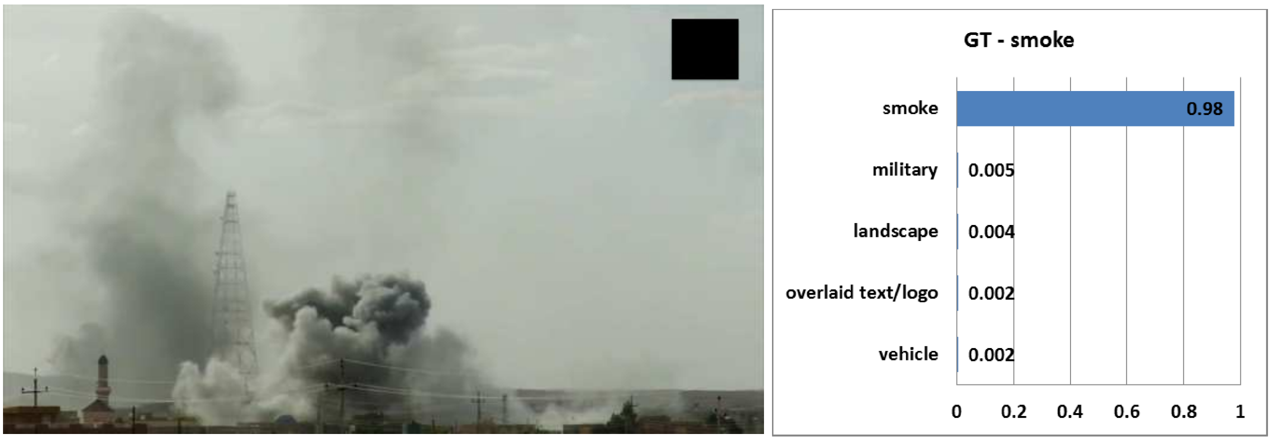

[redacted]

Figure 18: Concept detection of smoke/element

Source: CERTH

Differently than in the object detection (see paragraph below), the tool provides an overview on the overall image, by highlighting the main elements in it and thus instructing the tool for additional searches.

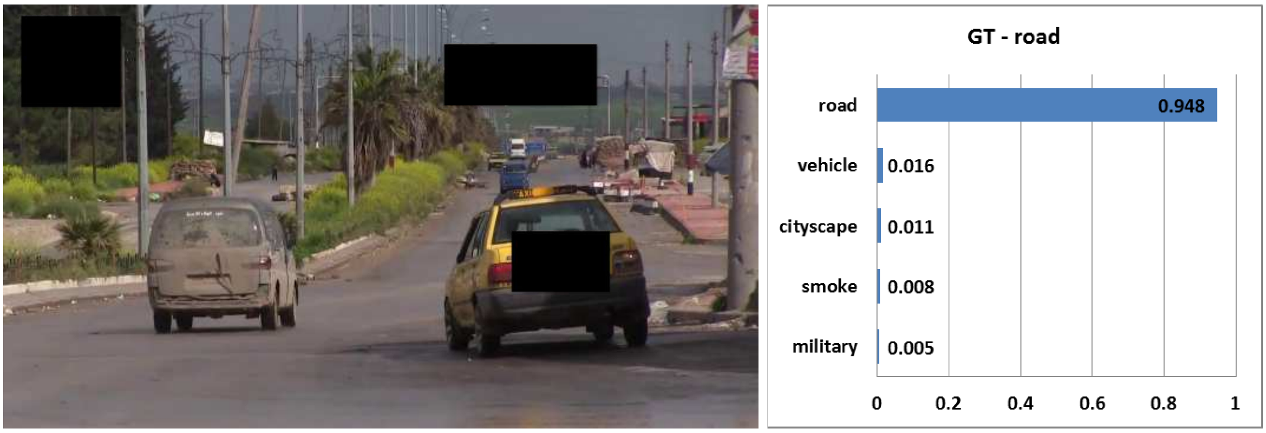

[redacted]
[redacted]
[redacted]

Figure 19: Concept detection of road / element

Source: CERTH

One of the issues addressed by the end users is related to the possibility for the final user to train the program in order to recognize any kind of object. In order to this, the algorithm needs to be re-trained, and the databased of training data updated and then re-adapted to update the repository.

For example, in order to find a specific logo (e.g. not the general ISIS logo, but for example the logo of a local organization such as the terrorist group ETA), the user needs to specifically train the tool or link to another module of the platform – e.g. the semantic one.

Object detection (image/video)

The goal of this module is to locate and identify the real-world objects or distinctive artefacts (e.g. logos, tattoos) that are present in the examined visual content. Achieving the latter will enable the efficient implementation of common content manipulation tasks, such as indexing, search and retrieval. One of the main contributions provided by this tool, for instance in the analysis of propaganda-related materials, is the possibility to filter out a vast number of unnecessary video scenes, by only keeping those including terrorist-related objects. One example is provided in Figure 20, below.

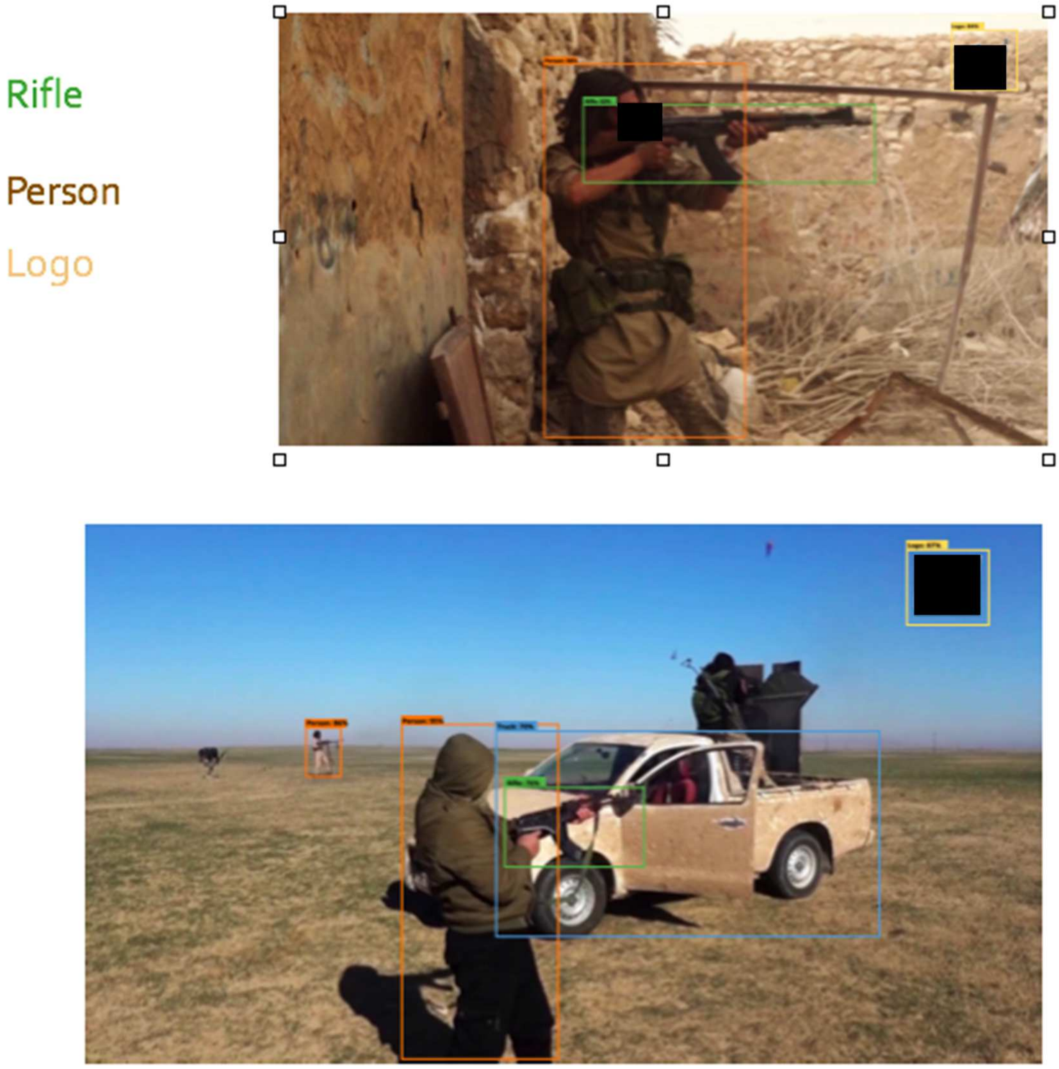


Figure 20: Object detection

Source: CERTH

In practical terms, when the image is received as an input in the platform, the user launches the analysis and the results obtained are, for instance, the persons, the rifle and the logo - which are automatically detected in the image. According to the end user, this tool is of key importance as, most of the time, the object is indeed the most important part to detect. If launching through the whole DANTE dataset a search with the key word 'flag', the user will be able to find all the images containing the ISIS logo with a relevant % of accuracy.

People detection and identification (image)

This module aims at the detection of the individuals that are present in the visual medium to be analysed and the eventual unravelling of their true identities. For achieving the latter, the visual appearance of the depicted individuals will be analysed and compared with similar descriptions from other visual sources. Thanks to this tool, the user can find in which particular part of the image or in which frame of the same, or even different, video(s) the same persons can possibly re-appear.

The recognition of individuals in the images or videos is generally easier, as the body is more recognizable than the faces.

The spatial-temporal constraints can be a limit, when analysing material from different places or periods in time (e.g. videos from Syria in 2014 different than most recent ones). However, a number of tests were made - for instance to search for 'Jihadi John'. When launching the search, a number of screenshots were extracted, all representing individuals with a certain probability of matching the real person.

One of the aspects addressed by the DANTE LEAs testing the platform was if, behind finding similar profiles, the tool is able to detect who is in reality the guy in the images. However, in order to do so, the tool should be linked to an additional module performing these kinds of linking.

Box 4 - Video summarization

The module related to video summarization aims at producing a short summary of the analyzed videos, including only the content that is supposed to be relevant for the needs of the end user. The overall goal is to significantly reduce the video footage that needs to be manually inspected by the human user, thus supporting the analysis to analyze and detect the relevant content in the videos' segments.

While showing the demo of this module to the DANTE LEAs, the analysts highlighted that those parts which are more interesting to them were not highlighted by the tool. In fact, for the analysts, the most important parts of the video are related to the introduction of the video - where the signature and logo of the author is usually reported - and the final section - where they can read the date of production, in the bottom-right part of the screen.

This comment, even if very specific, is of particular importance because underlines how crucial is the involvement of the domain's experts in the annotation phase. This phase is in fact crucial in training the module, in order to ensure obtaining useful information through the tool itself.

3.4 Second level – platform enabling technology

Keeping the chain of custody and evidences' gathering

Chain of custody is an essential first-step in cyber forensics investigations, as it is basically documenting the way that the items, acquired during the investigation, are secured, transported and verified in an appropriate manner. Through the chain of custody, it should be possible to define by whom/when a certain piece of evidence was accessed or modified. Digital evidences, as presented in the previous paragraphs, can consist in digital videos, audios or even texts. The 4-steps path followed by the process for collecting digital forensics is reported in Figure 21, below:

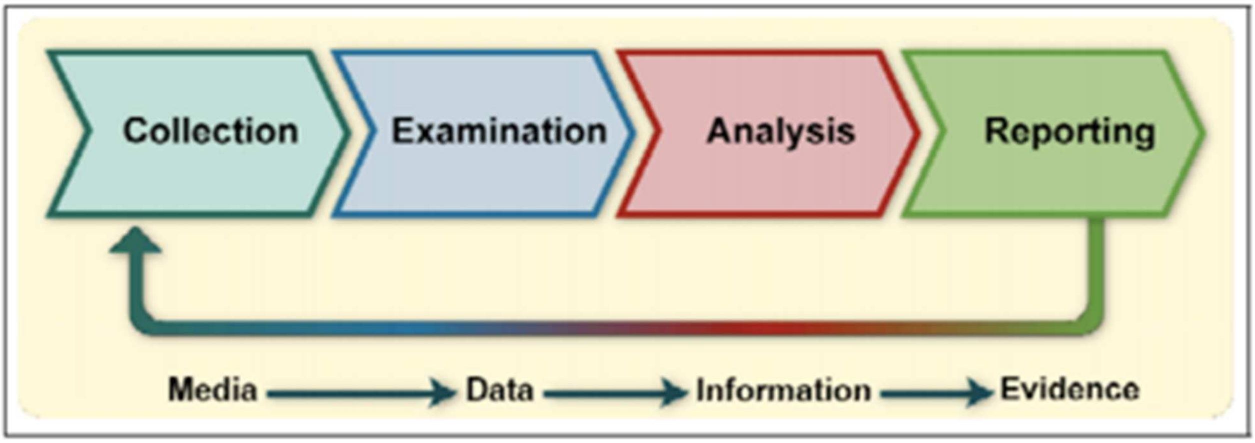


Figure 21: Chain of custody

Source: Kent (2006)[51]

The need of implementing a chain of custody and a chain of evidence mechanisms in the DANTE architecture originates from the danger that stored information could be accessed and manipulated in unauthorized and malicious way in order to compromise current investigations.

In simple terms, thanks to the chain of custody mechanism developed within the DANTE platform, it is possible to track different accesses to the resource, and thus define the story of all the different users that interacted with the resource. In fact, based on the different national legislation, when bringing the case in front the court, in order to present the digital evidences, this storyline should be provided. The final aim is to be able to discover who, eventually, modified or tampered with the evidences [chain of custody].

Moreover, once the platform obtains a new resource, the platform calculates an ‘hash’ (an alpha-numeric string) that identifies the resource. Any possible change or manipulation of the resource develops a different ‘hash’ - therefore the originality and integrity of the resource is guaranteed [chain of evidence].

Another relevant aspect may be related to the protection of personal data as part of the resources collected, and thus of the integrity and confidentiality of data. It has to be ensured that the DANTE system implements certain control principles, to prevent unauthorized access, processing, alteration, deletion and disclosure, specifically of personal data, and organizational measures are taken to ensure the confidentiality.

Any access to DANTE resources has to be authorized, and protected against accidental or unlawful destruction, accidental loss or processing, by means of policies and technical solutions.

In terms of data storage, the authorities can only retain what is ‘strictly’ necessary, and must adjust the retention so that only personal data and information on private lives is kept if it is useful for the purposes and legitimate aims it was collected[52]. The information must be vital to intelligence operations seeking to

[51] Kent, K., Chevalier, S., & Grance, T. (2006). Guide to Integrating Forensic Techniques into Incident.

[52] See also DANTE D3.1, Privacy Ethical and Legal Constraints.

protect democratic institutions.

However, based on a set of consultations with relevant stakeholders and the European Commission, at present, the resources collected and analyzed through the DANTE platform are not to be considered as digital evidences to be brought in front of the court. Further developments should be and will be explored in follow up actions, also building upon the DANTE results.[53]

Counter-terrorism intelligence: operational tips for LEAs

3.5 Third level - connecting the dots

Visual analytics

The module on visual analytics is included in the System Integration and Applications package, developed to provide interactive Visual Analytics services related to terrorist activities, allowing to explore geo-spatial and context awareness information retrieved by DANTE. It presents the outcomes of the services developed on WP6, 7 and 8 in order to collect and summarize the multilingual relevant information and knowledge, and present it by mean of suitable visual metaphors, that take into account the time variable in relation with other elements of interest.

It aims to provide an analytic toolset with innovative dashboards for an effective collaborative decision-making environment. It is based on a user interface that allows data representations and transformations, visual representations by using different kind of charts and interaction and techniques to support production, presentation and dissemination of the results of an analysis.

In order to give a complete description of the three use cases of DANTE, it is disposed in different dashboards, which should report information from the platform about diverse features about the terrorist attacks, such as location or targets, characteristics about suspicious activities in Social Media or economic transactions, as well as data retrieved from the modules of Video and Text Analysis (as above described).

The dashboards have specific filters, as time and geographical filters, that will be useful to aid LEAs to focus their investigations and achieve more concrete results.

The main overviews provided by the platform include:

a. General info

b. Social media

c. Fundraising

d. Video info

e. Text info extraction

f. Trend analysis

As this module was one of the last ones to be integrated in the platform, some tests have been performed

[53] See the developments of the project ANITA, available at: https://www.anita-project.eu/.

by analyzing data contained in the Global Terrorism Database (GTD), an open-source database including information on terrorist events around the world from 1970 through 2017 (with annual updates planned for the future).

By using the GTD data, for instance, the types of data visually showed by the module include:

- overview on terrorist attacks worldwide / linked to countries, types of events (attacks, kidnappings etc.) - all graphs are interactive, the user can select what he/she wants to visualize
- types of terrorist groups
- time evolution
- types of weapons used in terrorist attacks
- objectives: where are the attacks directed?
- analysis of posts from social networks
- social networks' users: age, gender, place. moreover, by including the name of the social networks' user, the user can check the followers, the number of posts, access the network of people linked to him/her (eg. on FB, Instagram). These data, in particular, are coming from the DANTE crawler of social networks. It's important to highlight that the crawler can only access public data. In order to use the crawler, ad hoc users on the networks of interest have been created
- Fundraising: check network and transactions between people. Specifically, on bitcoin, the categorization is based on tags or ID of transactions. Usually, investigation is carried out on ID of transactions, thus the user can only see transactions from one point to the other
- Summary of information of video analysis: categories, languages, number of videos visualized at the moment etc.
- Temporal analysis of previous information. This kind of analysis of particular importance, as it allows elaborating predictions. For instance, by comparing the different sources related to the types of weapons used in the terrorist attacks, and looking at them through a time-frame, it is possible to anticipate the trends and understand the main trends.

In general terms, the main goal of this toolset is to help LEAs to visualize a huge amount of information in a visual and intuitive way. Furthermore, it includes a series of filters to allow investigators to focus their research. However, as underlined by a number of LEAs during the pilot-demos, this tool is mainly useful for the heads of the units or departments, to elaborate strategical analysis. What is missing, for instance, for a police officer who is working on a particular case, is the possibility to focus on the information of a specific case. The developers then specified the utility in terms of starting an investigation, as through the analysis and visualization of aggregated data a user may identify another related case that he/she was not actually following yet. All graphs and visual elaboration can be exported, together with their related data: one significant added value is the presence of an interactive map, showing the evolution in time - which at present is not available for other similar tools.

Some examples of visualizations are the following ones.

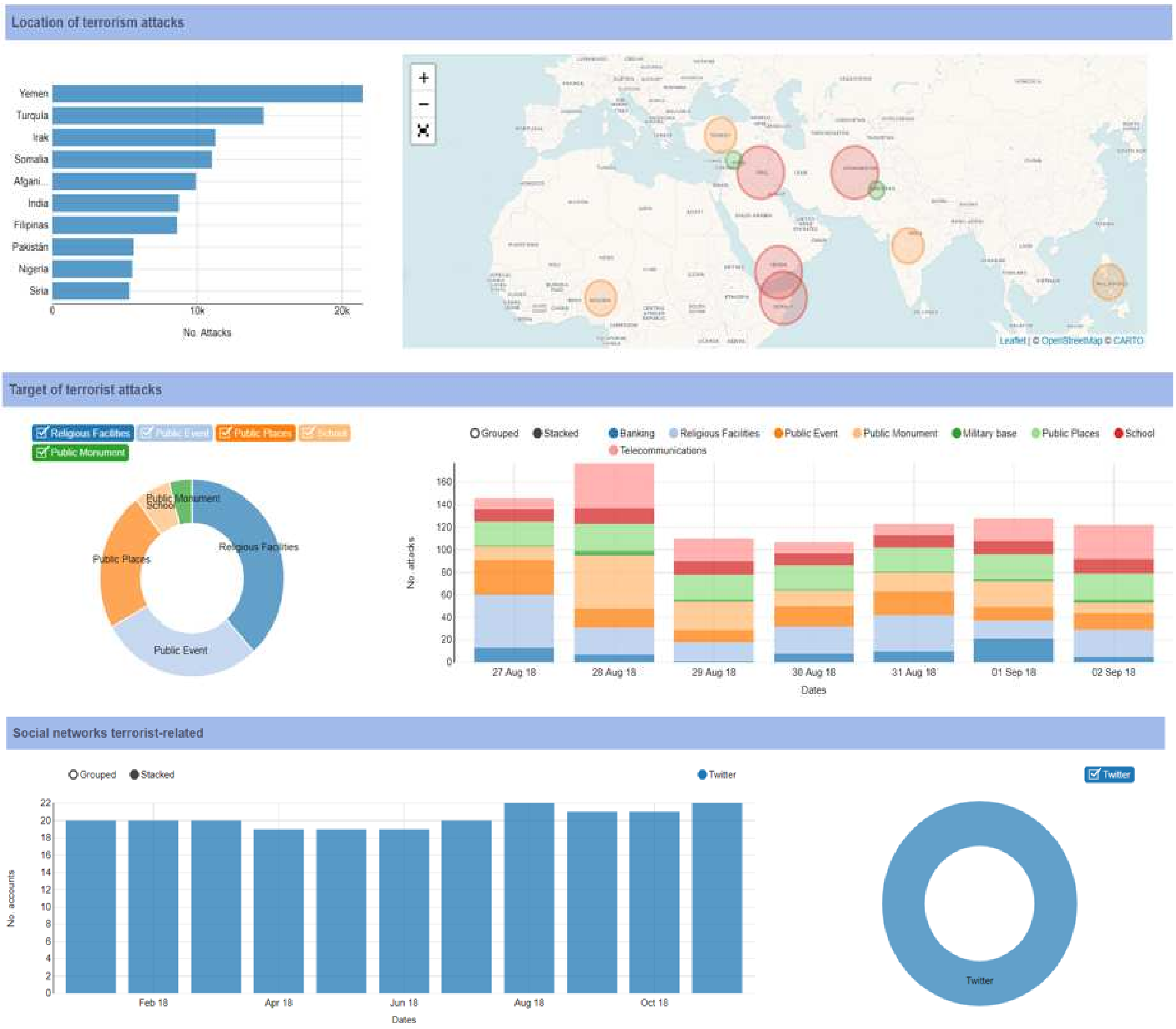


Figure 22: Visualizations of analysis performed through the Visual Analytics tools

Source: PRG

Trend Analysis

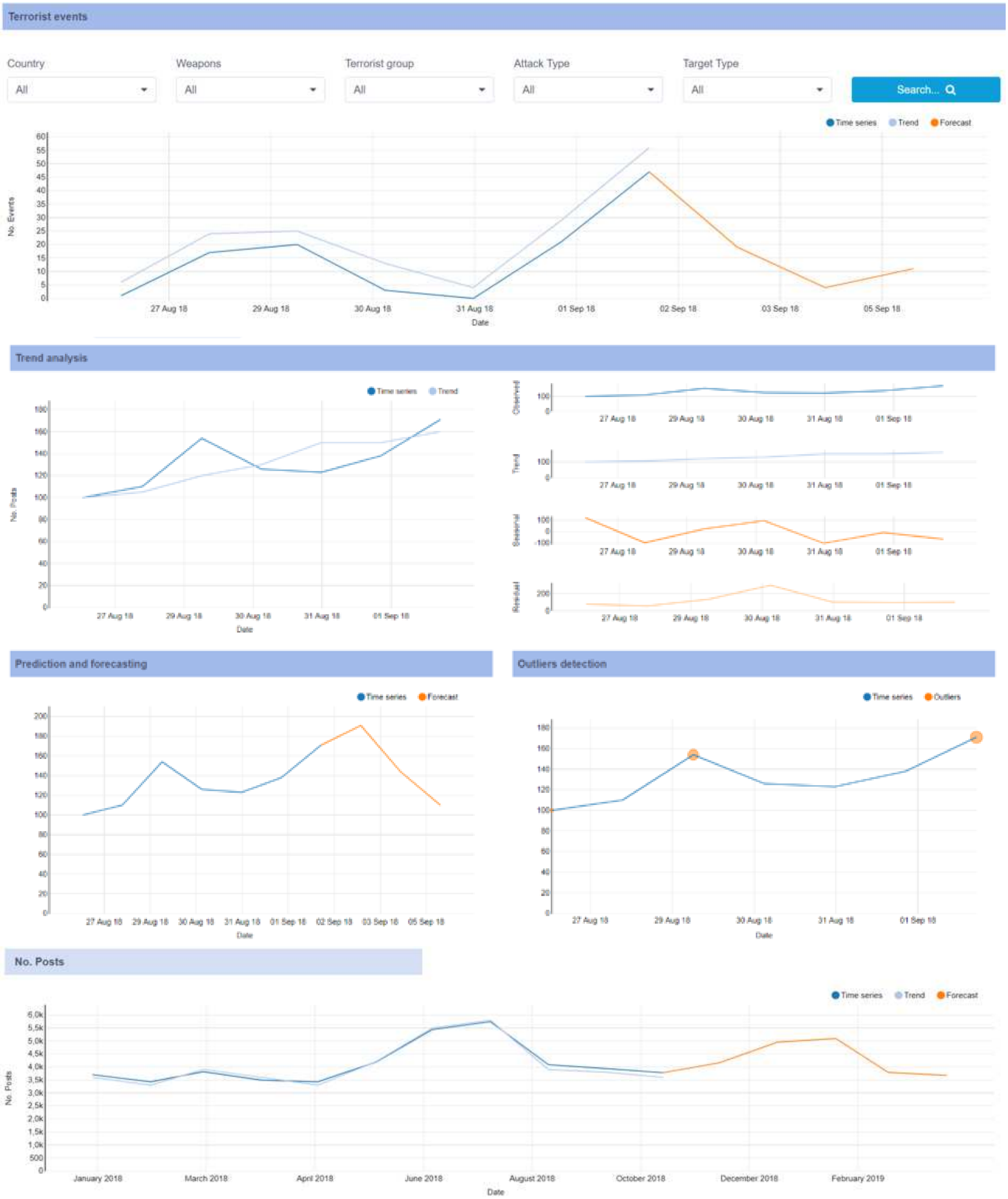
Trend Analysis consists in a large-scale temporal analysis of terrorism trends. The module is part of the Knowledge reasoning and Fusion package, which is designed to visualize browsed information, using time in relation with other characteristic elements, such as the features that describes a terrorist event, or the tracking on the activity in Social Media.

This module aims at providing a visual representation of predictive analysis and comparative analysis from the outcomes from DANTE and other sources based on Time Series.

This task focuses on the modelling of terrorist related activity, in order to find similar behaviours during the crawling of unknown/unidentified sources. The possibility to identify those actions should facilitate the LEAs investigations by allowing to explore different events, being able to improve the early detection of such events and, therefore, take a real advantage to anticipate to these attacks.

Similarly, to the visual analytics, two tests were carried out:

- On one hand, the information retrieved by DANTE and related to terrorist attacks will have been examined, with the aim of showing predictions according to the main trends of the last years. The same analysis was carried on a set of social media-related data.
- On the other, information from the Global Terrorism Database were analysed. In this context, different characteristics of the attacks were studied, including for instance the number of casualties, the types of attack or the terrorist groups involved, in order to display trends, predictions and outliers.



Terrorist events

Country: All

Weapons: All

Terrorist group: All

Attack Type: All

Target Type: All

Search...

Figure 23: Typologies of trend analysis

Source: PRG

Social graph analysis: group discovery

DANTE's Group Discovery module mainly aims at detecting terrorist presence, in terms of potential suspects and networks, on the Surface and Deep Web. The module combines information extraction methods, tools and natural language processing techniques, together with semantic information derived from social network analysis, in order to automatically process online content coming from disparate sources, and identify people and relationships that may be linked to terrorist activities.

The main functionality of the module on Group Discovery is to build a set of social relationship graphs, starting from the information retrieved by a Twitter account. More specifically, it allows retrieving information on the accounts' relationships among each other. For all the accounts, it is possible to see all the information related to what that person has discussed, to other (related) social network accounts etc. The version developed within Dante specifically operates on Twitter, but the developers are working on further integration with Facebook and Google Plus.

In practical terms, a specific Twitter account is detected - based on investigative inputs and needs. The user can choose how many tweets as of a date need to be retrieved, in order to reconstruct the social relationships' graph.

The graphic below shows an example of a suspicious Twitter account.[54] The thickness of the edges reflects their importance in terms of the number of times the corresponding action has been performed by the user (# times commented, # times mentioned, etc.). The person under scrutiny is represented as the grey central node of the graph, whereas users who commented on his/her Tweets are represented as purple ellipses; users who mentioned him/her are depicted as cyan rhombi; and users who have been mentioned by him/her are depicted as red rectangles.

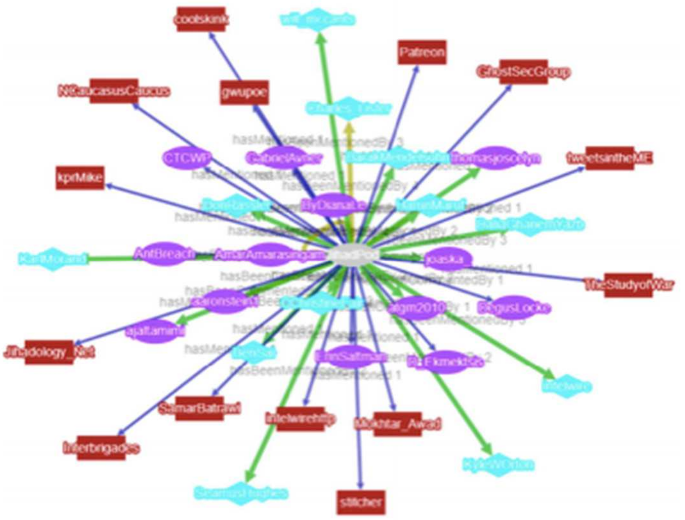


Figure 24: Social relationships' graph

[54] Please note that the screenshots herewith included correspond to the demo version of the group discovery module. The visualization on the integrated platform may appear differently.

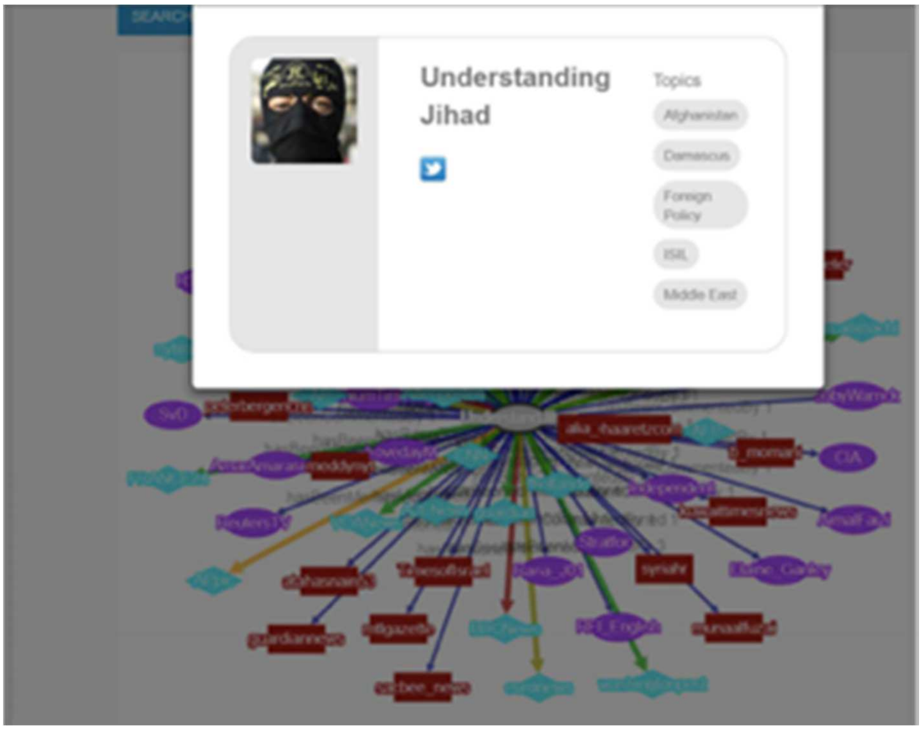


Figure 25: Example on Understanding Jihad's account (public)

The card, represented in Figure above, shows the details of a given user in the analyzed network, in this case named 'understanding Jihad', including his social profiles and related topics.[55]

Additional and potential enhancements of this module include the possibility of using further natural language processing techniques, in order to conceptualize textual messages and documents exchanged online, and use the corresponding results to detect non-explicit, hidden relationships among online users.

Moreover, the approach implemented in this module could be in principle applied to other application scenarios, including those related to the identification of hate-speech and hate spreaders over the Web.

During the pilot demonstrations, the LEAs explored the possibility, through the same tool, to get to see the number of visualizations of public Facebook pages: in fact, knowing the number of visualizations on FB public pages could be important to allow police officers to detect who is actually visualizing public propaganda material online, without the need to know whose friends on FB these people are, who are they in contact with etc.

This matter is also related to the privacy issue, in fact - as specified also for other modules - only public content can be accessed, other this would violate the rules on piracy and data protection.

When looking at the relationships among different users, police officers need to know what specific social media each account is using; in this view, what could help is some sort of filtering option, allowing the user

[55] The expansion to FB and Instagram could be considered as a useful enhancement for future developments.

to restrict the user's search.

LEAs also interrogated on whether, through the same tool, it is possible to retrieve the phone number: again, it's possible only if the number is publicly available. The same is valid for the user/account location.

One thing to be particularly stressed is that there is an external library, which allows getting the public information. No matter where the retrieved info is coming from, DANTE needs to retrieve a huge number of information (both through open source or closed/internal sources), otherwise the tool cannot be exploited at its maximum potential. The module thus receives what has been retrieved by the DANTE social media crawler, and then the relationships are built from publicly available material

In general terms, the tools itself seems useful to the DANTE LEAs: in fact, if a suspect is trading explosives on FB, the tool could help narrowing the focus down, to whom is actually selling or buying from more easily, by overcoming the main obstacles e.g. related to not having a FB account, or not being among the user's friend on FB etc.

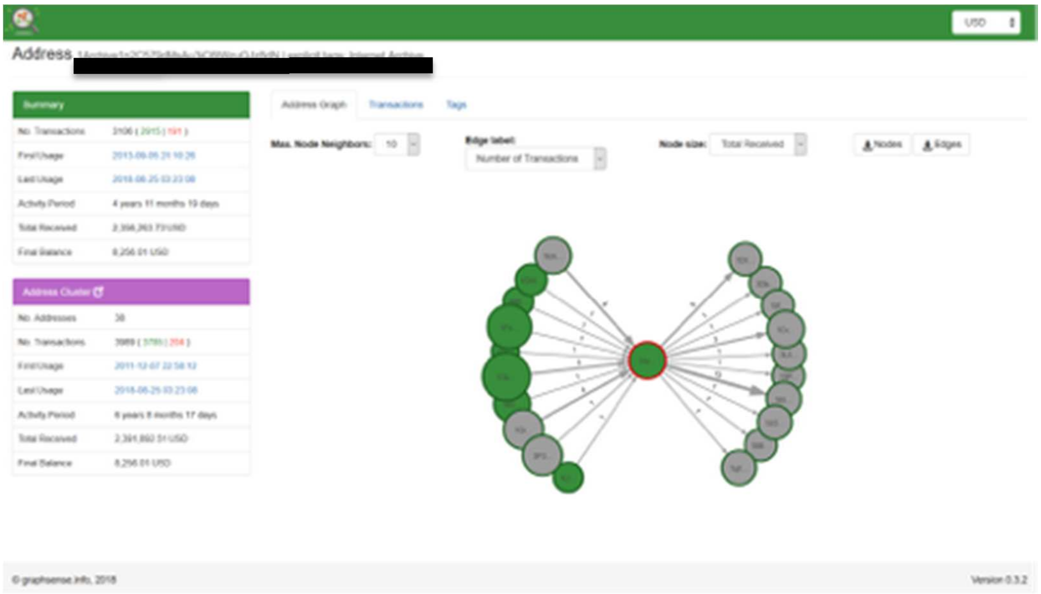
With regard to the identification issues, probably the combination of this tool with other tools, eg. able to detect images, would allow the users to ensure that they are monitoring the right person of interest, in case this person for examples have a face already known by the authorities.

Financial networks transactions

The Internet's appeal with regards to raise and transfer funds to support terrorist activities is offering a broad reach, timely efficiency, and a certain degree of anonymity and security for both donors and recipients. While this threat is widely recognized, many countries still lack the technical capabilities necessary to investigate online terrorist activity, including financial transactions.

With particular regard to virtual currencies, within DANTE the basic tool for analysis of cryptocurrency is the open source project GraphSense. It applies a graph-centric perspective on digital currency transactions. It allows users to explore transaction and follow the money flow, facilitates analytics by semantically enriching the transaction graph, support path and graph pattern search, and guides analysis to anomalous data points.

The graph below shows the dashboard of Graph Sense, which at the moment includes the analysis of different virtual currencies including Bitcoin, Litecoin, Bitcoin Cash and ZCash.



USD

Address [redacted]

Summary	
No. Transactions	3106 (2915 | 191)
First Usage	2013-09-06 21:10:26
Last Usage	2018-06-25 03:23:08
Activity Period	4 years 11 months 19 days
Total Received	2,356,263.73 USD
Final Balance	8,256.01 USD

Address Cluster	
No. Addresses	38
No. Transactions	3989 (3785 | 204)
First Usage	2011-12-07 22:58:12
Last Usage	2018-06-25 03:23:08
Activity Period	6 years 8 months 17 days
Total Received	2,391,892.51 USD
Final Balance	8,256.01 USD

Max. Node Neighbors: 10

Edge label: Number of Transactions

Node size: Total Received

© graphsense.info, 2018

Version 0.3.2

Figure 26: Graph Sense Dashboard

Source: AIT

As DANTE focus is to obtain and share information on potential terrorist financing activities, through the use of virtual currencies, an extension to the previous program was developed - which is called Tag Sense. This is a stand-alone web application, interacting with Graph Sense, and mostly focused on linking people behind e-wallets and their digital identities.

“The legal status of virtual currencies in the context of EU GDPR remains unclear in most legal jurisdictions. On one hand, the nature of Blockchains means that the distributed register associated with a given virtual currency are publicly available data. The question is, do these data constitute personal data? Under earlier data protection regimes, the answer would have been no, as these data consist only of anonymous addresses associated with transactions. However, under GDPR, data can be considered personal data if it is theoretically possible to associate an anonymous ID with a real person. These addresses are derived from asymmetric encryption keys, with the private key under the control of a user. Thus, it is clearly the case that one could theoretically match a private key in the possession of a user with a publicly known address”.[56]

When looking at the linkages among electronic wallets, the main correlated inputs and outputs are bitcoin (or other currencies-related) addresses, which are pseudonyms hiding real identities.

[56] D9.3 DANTE Knowledge mining services_V0.6.

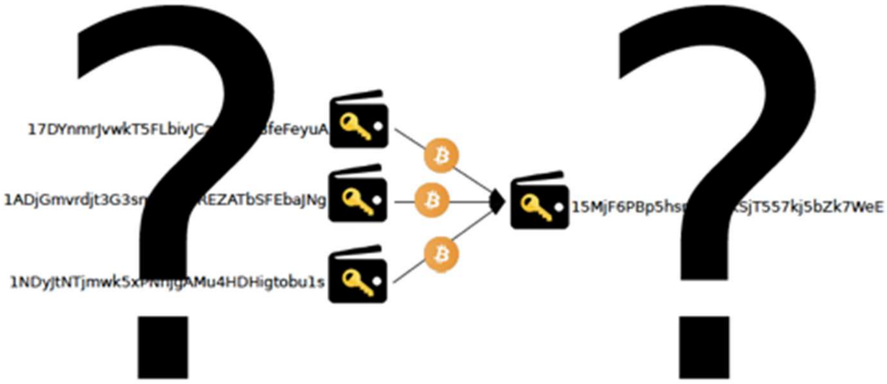


Figure 27: Bitcoin addresses as pseudonyms

Source: AIT

There are some websites where one can explore e-wallets (walletexplorer.com; poloniex.com), but the relevant information are not always available.

Through the GraphSense platform, previously described, it is possible to explore cryptocurrencies transactions, both outgoing and incoming. Other information is provided, such as the date of the first usage, the total amount of cryptocurrencies received, the activity period etc. While the analytical process is automatic, the first set of information (tag + bitcoin address) needs to be manually inserted - and thus the human effort by the police officer, or the website manager etc. is required.

The information to be collected through Tag Sense, instead, are not publicly available: at the moment they are only accessible through the developer's Intranet (and once integrated, also through the DANTE platform). The added value here is the collaboration among officers: the quality of the analysis is based on the type of information uploaded and on the collaboration element among officers uploading the information (eg. to reveal encrypted tags, as illustrated in the Figure below).

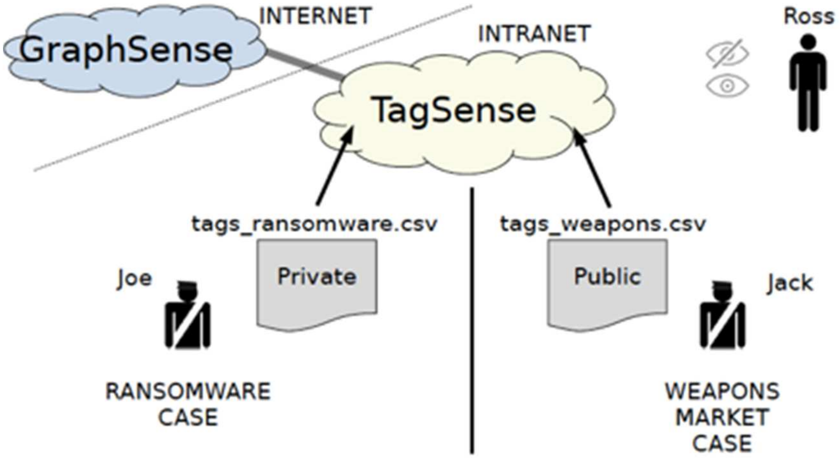


Figure 28: Overview on Tag sharing

Source: AIT

As a general outcome from the pilot sessions with the DANTE end users, there are not so many evidences of terrorist or potential terrorists using bitcoin at present, in current cases or investigations. Even though similar applications still do not see concrete or daily applications, further developments are encouraged in this field - as the migration of terrorist network to the Internet for financing purposes, including the Deep and Dark web, is more and more a reality.

4 Part III – Guidelines and Recommendations for LEAs: improving intelligence and investigation practices in counter-terrorism

In the previous sections of this report, an analysis of the main functions developed within the DANTE integrated platform has been performed, with a specific focus on its operational use and possible improvements to support the end users in intelligence and investigation activities. On this basis, a set of guidelines and recommendations on how to enhance intelligence and investigations in the counter-terrorism field is herewith presented.

At this stage, it is important to reiterate the starting point of the DANTE project, which was developed to address some relevant challenges in the actual counter-terrorism scenario, as highlighted also by the European Commission[57], and including among others:

1. The need for more efficient and effective automated techniques, to cope with the dangers involved in the use of the Internet by global terrorist organizations and grassroots terrorist cells;
2. The huge amount of data [on both surface and darkweb sites], which represents a major obstacle to the reliable and fast analysis of their contents;
3. The need of performing large-scale temporal analysis of terrorism trends;
4. The importance of involving law enforcement bodies from the design phase to the prototyping and test phase of the proposed tool;
5. The need to address the management of personal data, and related ethical and legal issues.

Based on the activities implemented and the findings achieved by the DANTE project, the followings have been identified as key areas to strategically work towards an improvement of LEAs capabilities in countering-terrorism, including on the Internet:

- Knowledge
- Technology
- Training and operational capacity building
- Transnational and Inter-regional Cooperation
- Ethics, privacy and fundamental rights

The reference framework settles within the EU counter-terrorism strategy[58] and its four main strands, which are: *prevent*, *protect*, *pursue* and *respond*.
In particular, the third strand - *pursue* - is at the core of DANTE action. In fact, the project is strictly related to the EU's focus, which is on:

- improving cooperation and information exchange between police and judicial authorities;
- depriving terrorists of their means of support and communication, and
- tackling terrorist financing.

[57] For more information, *the H2020-EU.3.7. - Secure societies - Protecting freedom and security of Europe and its citizens* Programme is available at: https://cordis.europa.eu/programme/rcn/664463/en. The *FCT-06-2015 - Law Enforcement capabilities 2: Detection and analysis of terrorist-related content on the Internet* Topic is available at: https://cordis.europa.eu/programme/rcn/665101/en.

[58] In 2005, the Council adopted the EU counter-terrorism strategy to fight terrorism globally and make Europe safer while respecting human rights based on four strategic strands of work, namely prevent, protect, pursue and respond. Combating terrorism was defined as a primary responsibility of the Member States, supported by the added value of the EU role. The strategy was further revised over the years, but the four main strands mentioned above confirmed as relevant.

Before delving more in details into the set of specific recommendations, a preliminary overview on the inter-linkages among the areas specified above is outlined, also in relation to the challenges previously highlighted.

In general terms, the overall experience of DANTE confirmed the theorical framework initially designed around the project concept. In fact, it clearly emerged that the role of **technology** was – and still is - crucial because LEAs need to be provided with automated tools for improving detection and monitoring of terrorist related contents and activities, especially online. Moreover, tools are needed by the LEAs to further enhance the handling of huge amount of heterogeneous (big) data, and information about the terrorist threats and the possible interdependencies with other criminal activities. In this view, the project implementation substantiated that a continuous exchange between the end users (LEAs) and the technology providers is of fundamental importance to develop truly relevant technological tools for preventing and countering terrorism.

This stated, the element which emerged as the silver thread throughout the project implementation is the importance of an **inter-disciplinary approach** and the involvement of professionals with different and complementary characteristics and skills. In fact, this was the key to facilitate the dialogue and the mutual understanding between the LEAs and the technology providers, which presented some challenges due to the complexities of both domains.

Alongside the interdisciplinary approach, also the knowledge-based approach has proved to be decisive. In fact, the DANTE project confirmed that the development of technological tools, in order to be effective, must be based not only on the LEAs needs, but also on the characteristics of the phenomenon to be faced. The creation and sharing of knowledge, based on the inter-disciplinary approach mentioned above, was crucial under many respects. Some of them are hereby listed as a matter of example.

- It definitely supported the dialogue and cooperation among the end users and the IT developers. Based on the criminological analysis, along with the relevant inputs from the legal, privacy and ethical assessment, the LEAs were facilitated in explaining the procedures and activities they usually implemented to investigate it, as well as to monitor the online environment. On the other side, the technological partners were facilitated in understanding the different constraints affecting the enforcement environment as well as the in proposing tools to help improve the quality of prevention and contrast activities, thus avoiding additional possible constraints.
- The combination between the criminological and the technological approach supported the process of refinement of the use-cases as well as the fine-tuning of the users' requirements.
- The assessment of the phenomenon from the criminological standpoint contributed also to highlight possible scenarios and trends to be further considered also in the development of the technological tools, so to make them innovative and capable to support the LEAs in facing the possible evolution of the criminal threats/risks on the short-medium term.
- Furthermore, the knowledge-generation process initiated at the very beginning of the activities, also supported the identification of relevant sources where to retrieve information, data and documents, needed to develop and train the technologies used in the framework of the technological development.

The DANTE project is a good practice of reference to further encourage the **virtuous circle of knowledge-technology-enforcement**, which emerged to be fundamental to achieve effective and sustainable results when talking about innovation actions in the field of counter-terrorism.

This virtuous circle cannot avoid including also the key role of the criminological analysis and the knowledge-based approach in contributed at enhancing operational skills within the LEAs. From this standpoint, the role of **training/capacity building** activities was as well a fundamental one. As emerged from the DANTE project, the multidisciplinary perspective should be reflected into the LEAs capacities, through tailored actions – also to encourage the exchange of existing practices.

Due to the transnational nature of the terrorist phenomenon, the DANTE project confirmed that **inter-regional cooperation** is a crucial aspect, too. The online framework has no geographical boundaries, but the dynamics and trends of the criminal phenomena – including terrorism - are strictly related to the offline dimension, which is geographically located. Accordingly, the online-offline should be a priority, as well as the cooperation among national agencies. It could be further promoted for example through actions able to convey transnational analysis on the phenomenon, technologies covering more languages, trainings and capacity-building events involving participants from different LEAs… These examples, emerging from the DANTE experience, were further processed to elaborate the recommendations included in this report.

Finally, as a cross-cutting element, the management of **ethical, legal and privacy** issues need to be reflected and addressed in all the above-mentioned areas. The DANTE project confirmed that the virtuous circle of knowledge-technology-enforcement should be further supported by the contribution of these disciplines so to achieve sustainable, efficient and effective results.

Based on this framework, the recommendations are herewith described according to the five key areas mentioned above.

KNOWLEDGE

The approach adopted by the DANTE project, as well as its main findings in terms of inter-disciplinary and criminological research on the one side and technological development on the other, emphasized the need to keep on implementing a **knowledge-driven perspective** in the fight against terrorism.

Different disciplines - such as for example criminology, social sciences, laws and ethics, along with engineering and IT – need to further improve their capacity to interact and cooperate on a common basis, thus contributing to achieve common goals. Although it is technology-centred, the DANTE experience confirmed that innovation actions do need the support of other disciplines, and a greater willingness to interact is necessary.

It is recommended that the incoming call for proposals clearly recall not only the importance of collaboration between different disciplines, but also the complementary and equal role they play in achieving the objectives. In some cases, the technology-centred approach has proven to be discordant with the police environment because it fails to grasp the real needs and constraints. On the other hand, the knowledge-centred approach can easily assess the LEAs requirements, but it needs to be further enhanced by technological tools to respond to the operational needs and the ever-changing nature of criminal threats.

On their part, future actions should be structured around an initial pivotal phase where these domains could have the possibility to agree on the inter-disciplinary approach needed to implement the activities, to discuss and settle around a clear and agreed description of the contribution to be provided. For example, generating and sharing inter-disciplinary knowledge on a criminal phenomenon and its main trends is a complex task since crime is a fluid entity which can be observed from many perspectives. Accordingly, to be effective, the analysis needs to be targeted on the specific issues, factors of interest and/or needs of the project, especially in case of innovation actions involving technological development.

In fact, as confirmed by the DANTE experience, the strategic criminology-based knowledge generation can be organised so to contributing, among other purposes, at orientating the technological tools towards the LEAs needs and expectations by highlighting the criminal schemes and trends they have to face in their daily activities. Within DANTE, the criminological analysis focused on the use-cases of specific interest for the project, thus including the assessment of the role of the Internet in disseminating propaganda and training activities, preparing terrorist attacks and collecting funding to support the terrorist activities at different levels.

What is recommended is a closer relationship between the research team and the LEAs in order to integrate the findings from the literature analysis, the desk research and the other research methods with the information deriving from the direct expertise of the officials. This emerged to be a crucial element in the knowledge-generation process along with the involvement of specialised units within the LEAs participating in the consortia, which should be considered of utmost importance. In fact, both the cooperation between the researchers and the LEAs and the possibility to interact with skilled officials facilitate the data/information collection process, the assessment of the knowledge base relevant for the technological innovation, as well as the analysis of the LEAs needs, requirements and possible related constraints. Based on this close relationship with the LEAs, in many cases researchers are able to 'mediate' between the LEAs on the one hand, and the technological partners on the other, thus filling some gaps (in terms of communication, legal and ethical restrictions, assessment of needs, improvements needed etc.) which often affect the initial phases of the projects.

Additionally, when considering the knowledge generation process, the involvement of the national judicial authorities should be recommended, too. In fact, in compliance with current legislation, the analysis of judicial cases can further enhance not only the criminological analysis of the phenomena of interest and its main trends, but also the assessment of key pieces of information which could be relevant for the technological development. As confirmed by DANTE, the use of multiple and complementary sources should be considered an asset when detailing the methodological approach to generate knowledge.

In this context and considering the continuous changes and evolutions in the structure and modus operandi of terrorist groups (and individuals), another crucial element is to keep the knowledge on the current trend of terrorist phenomena always up to date throughout the action. The use of real cases studies, as those selected by DANTE [1. Online terrorist financing 2. Online propaganda 3. Online training and networking], is very useful from a methodological point of view. For this reason, additional case studies of interest should be explored in follow up actions, including for instance the links with other serious crimes as source of profits (trafficking of NPS, drugs, firearms etc.), the role of foreign terrorist fighters, the connections with other extremist movements etc.

Furthermore, what is recommended is to support actions contributing at developing scalable and sustainable methodological approach, which could be easily replicated also to generate knowledge on other criminal phenomena. Structured, time series and cross-sectional data play a key role in generating operational knowledge for the LEAs.

For example, the DANTE knowledge generation process could be easily used also to expand the focus from Islamist terrorism to other forms, including local (for example far-right groups active in many EU Member States). Further insights on other typologies of terrorism could support improvements of the technological tools developed by DANTE, thus providing the LEAs with advanced functionalities. In this way, the projects could be as much effective as possible in contributing at responding to emerging terrorist threats at national, regional and international level.

TECHNOLOGY

As previously mentioned, the role of technology is crucial to provide LEAs with automated tools for detection and monitoring of terrorist related contents and activities, and they have to be designed to support LEAs in handling the huge amount of (big) data to prevent and counter terrorism actions, including online. In fact, the DANTE project experience confirms that while “technology itself cannot guarantee security, [but] security without the support of technology is impossible”.[59]

First, to make an innovation action achieve effective results, it is recommended that a virtuous circle is always established and composed by:

- The assessment of the strategic role of the proposed technological solutions, with specific regards to the possibility to get knowledge of facts and events in advance, thus preventing potential terrorist threats and limiting the financing of terrorist activities;
- The ex-ante evaluation and continuous monitoring of the impact and possible integration of the technological tools in the LEAs operational environment. It is therefore of utmost importance the active involvement of the end users in the whole phases of the actions – such as for example needs assessment, pilot test of modules, revision of results etc;
- The continuous interaction and the mutual understanding between the technological development on the one side and the knowledge generation process and the LEAs input on the other. The inter-disciplinary approach – with a leading role of criminology, which is strictly focused on the criminal phenomena – should be implemented throughout the overall technological development so to limit the difficulties which often emerge when there is a technology-oriented approach.

The inter-disciplinary perspective adopted within DANTE, and already highlighted as fundamental approach in the knowledge development phases, confirmed the need to avoid technology-centred actions, with limited integration with other domains. For example, based on DANTE, the involvement of experts in different disciplines during the pilots was an added value to facilitate the evaluation process, the impact assessment and to further promote refinements addressing the LEAs needs.

Furthermore, the aspect of security in the intelligence gathering process has also to be ensured by the technological development, to guarantee the relevance of the analysed information and to avoid risks of disinformation. It is recommended that experts in the field flank the technology providers with specific input to help them assess the risks and find the best mitigation strategies.

Entering into details and considering the specific technological aspects, it is important to highlight that one of the main added value of the DANTE platform is the fact that it combines a set of different tools, thus providing the end users with a unique environment to process data and perform analysis. Being an integrated platform, the DANTE functionalities have several benefits for the LEAs and this approach should be further replicated also in future actions. On this basis, a set of recommendations focused on specific modules of the platform are presented below.

Tools for investigative purposes

Crawling tools: providing the LEAs with advanced tools to develop a professional searching strategy should be considered a priority since it represents the preliminary and crucial step to ensure quality and relevance of the information gathered. Crawling tools should be included in a dynamic and continuous assessment and refinement process because the results of the monitoring process, implemented through the crawlers, could help in refining the searching strategy, and thus obtaining even better and more tailored results.

[59] Research for a Secure Europe Report of the Group of Personalities in the field of Security Research, European Communities 2004

Audio analysis

- It is important to differentiate between the analysis of audio materials related to i) propaganda material ii) inter-personal communication. The context in which the audio file is placed represents a key factor, as it can be related – for instance – to the different type of language used (classic Arabic for dissemination propaganda material as much as possible; local dialects or slangs for communication within groups etc.)

- The use of tools detecting noises in an audio file can be also strategic. As a practical example, recognizing a specific TV program as a background sound in an audio file can help getting additional 'metadata'. Moreover, understanding if a voice in Arabic is singing without music or if it is reciting the Quran, can provide with useful information on the countries or groups involved.

- It is important to take advantage of the integrated platform and the possible links between different modules: the speaker detection can help to filter a set of files based on possible authors; the text analysis, in particular the stylometry, can help to identify authors and define profiles.

Text analysis

- The stylometric analysis may respond to the need of countering anonymity of users both on surface and on dark-web/net. It would be useful to link this functionality with other modules (eg. NLP), to obtain information extraction on texts' authors or even link with relevant semantic networks already identified.

- A strategic use of multilingual automated translation is also very relevant: so far, the main language detected and translated is classic Arabic into English. However, for countries or regions affected by local forms of terrorism, there is a need to integrate also local languages both as detected and translated languages (e.g. Basque, Catalan etc.).

Video analysis

- Regarding the concept detection function, the possibility to provide an automatic categorization of large-scale terrorist related content is highlighted as relevant. Ideally, the user should be able to train the program in order to recognize any kind of object: in order to do this, the relevant algorithm needs to be re-trained and possible links with the 'semantic analysis' function would be of support. As stated for some of the previous functions, focusing on specific needs related to the context is of key importance.

- Regarding the object detection function, the possibility to locate and identify real-world objectives (e.g. logos, tattoos) is highlighted as relevant. This tool is of fundamental importance for the end users, as the objects are among the most important parts to detect in the video files.

- The people detection function allows detecting individuals in the visual medium: in this case, the next step for the LEAs would be, rather than only finding similar profiles, to further detect the real identity besides the guys in the video. It's recommendable to include a linking function, to create links between detected persons and profiles of existing suspects / individuals / terrorist group members etc.

- Concerning the video summarization tool, it was highlighted that the most important parts of propaganda videos, for instance, are the introduction - where signature and logo of author is usually reported - and the final section - where the date of production is reported. Thus, these segments should be always selected when analysing propaganda videos.

One general recommendation, which concerns all the different modules above described, concerns the crucial role to be played by the domain's experts in the annotation phase of the audio-video materials. This phase is crucial in training the module, to ensure the obtainment of useful and relevant information through the tool itself.

Finally, it is important to stress the need of integrating platform enabling technology for keeping the so-called 'chain of custody' and 'chain of evidence', with the final aim of facing the challenge of access or manipulation of stored information in an unauthorized or malicious way. However, at present, the resources collected and analysed through the DANTE integrated platform are not to be considered as digital evidences to be brought in front of the court. Further developments should be and will be explored in follow up actions, also building upon the DANTE results.

Tools for intelligence purposes

As a general guideline, the analysis of open source materials (e.g. propaganda or training materials available on surface and deep-web) is more useful for the development of trend or strategic analysis. On the contrary, the investigators should be more focused on the analysis of closed-data, mainly collected during the investigation phase. Below, an overview on specific recommendations elaborated on the DANTE tools more focused on intelligence purposes are reported:

Visual analytics consists in an analytic tool-set, which main function is to help the LEAs in visualizing the huge amount of information collected:

- this module was specifically recommended by the end users as a useful tool to elaborate strategical analysis in CT-related and other relevant areas. For a police officer, however, it would be useful to use the tool to focus on information related to a specific case: for example, when starting an investigation, the analysis of aggregated data to identify related cases of interest could be of help.
- The significant added value is the interactive map, which shows the evolution in time of phenomena, and can be used also for prediction purposes.

Trend analysis consists in the visual representation of the information and should support in elaborating predictive analysis and for early detection purposes.

Group discovery

- This tool allows building a social relation graph: at this stage, the information elaborated mainly comes from Twitter accounts' information, but further integration is envisaged by linking to additional social networks such as Facebook and Google Plus. These developments would be relevant, in order to use the available corresponding results to detect non-explicit or hidden relationships among the online users.
- There is a need to combine this tool with others, such as images detection, for specific identification purposes.

Financial networks transactions

- The tag sharing tool should help to put identities and identify people who are behind the detected e-wallets. It is recommended to look for this kind of information through external sources, by means of the DANTE platform itself. This would give the possibility to obtain information related to bitcoin address (e.g. first usage, total amount received, activity period) and obtain the visualization of the relevant networks.

- Closer interactions among officers dealing with different cases should be further promoted in order to find connections because up to now there is are very few cases of potential terrorists using bitcoin at the moment. This issue should be further explored.

TRAINING / OPERATIONAL CAPACITY

One of the emerging aspects of the DANTE project is the need to address the technological gaps in the LEAs' capacities, including skills on how to use the integrated platform in standard activities. It clearly emerged that there is a growing demand of trainings where the LEAs can:

- collect insights on the criminal phenomena from a wider standpoint,
- get access to analysis and materials elaborated by adopting an inter-disciplinary approach, and
- can exchange experiences and views with LEAs from other Member States.

Based on DANTE, it is recommended that similar technical trainings are developed involving also judges and prosecutors, with the final aim of strengthening the enforcement chain. Joint training actions would also help in encouraging the dialogue and information exchange between investigators, analysists, lawyers, judges and prosecutors, thus limiting the difficulties posed by the lack of mutual understanding.

In terms of training contents, existing gaps were detected specifically with regard to the knowledge of the dark web, the dark net and the use of crypto currencies. Moreover, it would be useful to develop focused material, aimed at understanding the money flows, as well as the flows of preparatory activities for preparing terrorist attacks – which are more and more often even low-budget attacks.

In general terms, the multidisciplinary perspective highlighted in the previous sections needs to be brought into the LEAs approach, by stressing the importance of using different perspectives to the analytical processes, such as the use of biographical approach or social network analysis etc. – with the final aim of developing targeted investigative strategies.

The training contents and programmes should be then adapted and tailored to specific targets [*training for local city units – watchmen; basic CT training; advanced CT training; training of trainers*], with the aim of creating a virtuous circular model, namely a system where the knowledge and the related technology are constantly updated.

The adopted training techniques should be mixed: the use of e-learning tools is key to ensure a constant update on the evolution of the CT phenomena. The face-to-face sessions are also of key importance, as they should facilitate the exchange between different police units and departments (eg. CT units, cyber-crime units, financial investigation units etc.). Moreover, the need for simulations in virtual environments is also recommended.

In this regard, the role of the pilot sessions adopted within DANTE needs to be highlighted: they played a key role in the development of this set of guidelines, also looking at future actions. One crucial point was the focus on specific uses cases [terrorist financing, online propaganda and training], allowing specific tests and the collection of ad hoc feedback from the end users.

Last but not least, there is a need for strengthening synergies among the relevant agencies at EU level involved training modules, to ensure a coordinate approach, and to build upon existing expertise and share good practices.

TRANSNATIONAL and INTER-REGIONAL COOPERATION

Due to transnational nature of the terrorist phenomenon, the collaboration element both inter- and intra-police forces should play a crucial role in the intelligence cycle. Also based on the results of the analysis developed within DANTE, the different opportunities and obstacles of using joint investigation teams should be further explored.

The online framework has no geographical boundaries, but the dynamics and trends of the criminal phenomena – including terrorism - are strictly related to the offline dimension, which is geographically located. Accordingly, the online-offline should be a priority, as well as the cooperation among national agencies. It could be further promoted for example through actions able to convey transnational analysis on the phenomenon, technologies covering more languages, trainings and capacity-building events involving participants from different LEAs…

From a wider perspective, the cooperation and the exchanges with LEAs in strategic countries (eg. MENA region) should be encouraged, by keeping into consideration the existing challenges related to the different legal and security approached.

It is also of fundamental importance to assess the existing EU policies and strategies and their effective impact on counter terrorism activities, by means of ad hoc monitoring instruments and possible (shared) indicators. In fact, while acting at the regional level in the European scenario, the coordination with the United Nations policies and other proposed responses at international level is needed.

ETHICS / PRIVACY ISSUES / FUNDAMENTAL RIGHT

In general terms, and specifically looking at the different aspects involved in the technological development and in the effective use of the technical modules for investigation and intelligence purposes, there is the need to find a balance among the need of carrying out LEAs activities at different security levels, with the need for protection of personal data, right to privacy, non-discrimination etc.

The concept of *privacy by design* and *by default* in the design of DANTE-like platforms needs to be embedded. Moreover, there is a need for validating and evaluating the impact of DANTE platform, against EU parameters and ethical standards.

It is also important to include relevant ethical and legal aspects in CT training activities.

Moreover, additional analysis specifically related to the risks for discrimination in using big data analysis tools, and more in general on the social dimension of big data analysis tools should be encouraged.

5 References

Bailes A., Frommelt I. (2004), Business and security: public-private sector relationship in a new security environment, SIPRI. Available at: https://www.sipri.org/publications/2004/business-and-security-public-private-sector-relationships-new-security-environment.

Bennett, C. J., Clement, A., and Milberry, K. (2012), "Editorial: Introduction to Cyber-Surveillance", in Surveillance & Society.

Boudreau C. (2006), "Multipolarité de la surveillance et gestion des médicaments au Québec", in Recherches sociographiques. Article available at: http://www.netalya.com/fr/Article2.asp?CLE=162.

Brown I. and Korff D. (2004), "Striking the right balance: respecting the Privacy of individuals and protecting the public from crime", Information Commissioner's Office.

Brown I. and Korff D. (2009), "Terrorism and the proportionality of Internet surveillance", in European Journal of Criminology, SAGE.

Bundeskriminalat, The extortionist's voice. Available at:
https://www.bka.de/EN/OurTasks/SupportOfInvestigationAndPrevention/ForensicScience/PhysicalEvidence/Extortion/SpeakerIdentification/speakeridentification_node.html

Cahen, M. (n.d.), "Le role de l'administrateur réseau dans la cybersurveillance", lecture notes, ENAP. Available at: http://www.netalya.com/fr/Article2.asp?CLE=162.

Centro Alti Studi per la Difesa – ISRI (2014-2015), L'evoluzione della capacità cyber, da cyber defence a cyber warfare. Available at:
http://www.difesa.it/SMD_/CASD/IM/ISSMI/Documents/L_evoluzione_della_capacita_Cyber.pdf.

Council of Europe - CoE (2001), Common Position 2001/931/CFSP of 27 December 2001 on the application of specific measures to combat terrorism, Official Journal of the European Communities, L344/93.

Council of Europe - CoE (2008), Protecting the Right to Privacy in the Fight against Terrorism, Issue paper of the Council of Europe Commissioner for Human Rights, available at:
http://www.coe.int/t/commissioner/Activities/IPList_en.asp.

Commission de l'étique de la science et de la technologie (2008), "Viseur un just équilibre: un regard étique sur les nouvelles technologies de surveillance et de controle à des fins de sécurité", adopted at 34th meeting of the Commission on 12 February. French version available at:
http://www.ethique.gouv.qc.ca/fr/assets/documents/NTSC/Avis-NTSC-FR.pdf.

Dahl E.J. and Viola D. (updated 2017), Intelligence and Terrorism, in International Studies. Available at:
http://oxfordre.com/internationalstudies/view/10.1093/acrefore/9780190846626.001.0001/acrefore-9780190846626-e-91.

Dumiak M. (2018), Interpol's new software will recognize criminals by their voices", available at:
https://spectrum.ieee.org/tech-talk/consumer-electronics/audiovideo/interpols-new-automated-platform-will-recognize-criminals-by-their-voice

European Parliament (2015), Understanding definitions of terrorism. Available at:
http://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_ATA(2015)571320.

Gauvain J., Lamel L., Bac Le V., Despres J., Gauvain J-L., Messaoudi A., Vieru B., Kheder W.B. (2018), Challenges in audio processing of terrorist-related data. Available at: https://link.springer.com/chapter/10.1007%2F978-3-030-05716-9_7.

Kent, K., Chevalier, S., & Grance, T. (2006). Guide to Integrating Forensic Techniques into Incident.

Maurice D. (2015), New threats and countermeasures in digital crime and cyber terrorism, IGI Global.

Poddar A., Sahidullah Md, Saha G. (2018), Speaker verification with short utterances: a review of challenges, trends and opportunities". Available at: https://ieeexplore.ieee.org/document/8302747.

Privacy International (PI), Communication Surveillance. Available at: https://privacyinternational.org/explainer/1309/communications-surveillance.

Shahar Y. (2008), "The Internet as a tool for intelligence and counter-terrorism", in Responses to Cyber Terrorism – Centre of Excellence Defence Against Terrorism, IOS Press.

Tremblay, M. (2012). "Cyber-surveillance", in L. Coté and J.-F. Savard (eds.), Encyclopedic Dictionary of Public Administration, available online at: http://www.dictionnaire.enap.ca/dictionnaire/docs/definitions/definitions_anglais/cyber_surveillance.pdf

UNODC (2011), Criminal Intelligence. Manual for Analysts. Available at: https://www.unodc.org/documents/organized-crime/Law-Enforcement/Criminal_Intelligence_for_Analysts.pdf.

US Congress (2002), Joint Inquiry into Intelligence Community Activities before and after the Terrorist Attacks of September 11, 2001: Abridged Findings and Conclusions.

Websites:

www.techopedia.com

www.researchgate.net

www.privacyinternational.org

www.ontotext.com

www.expertsystem.com

https://www.start.umd.edu/gtd/

www.walletexplorer.com

www.poloniex.com