****

**Advanced Tools for fighting Online illegal trafficking**

**D4.1 – Requirements, use cases and user scenarios**

|  |  |
| --- | --- |
| WP number and title | WP4 – Use cases, Requirements and System Architecture |
| Lead Beneficiary | DSTL CAST |
| Contributor(s) | All partners |
| Deliverable type | Report |
| Planned delivery date | 31/08/2018 |
| Last Update | 10/08/2018 |
| Dissemination level | Confidential |

**---**

Disclaimer

This document contains material, which is the copyright of certain ANITA contractors, and may not be reproduced or copied without permission. All ANITA 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 ANITA Consortium consists of the following partners:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Participant No | Participant organisation name | Short Name | Type | Country |
| 1 | Engineering Ingegneria Informatica | ENG | IND | IT |
| 2 | Centre for Research and Technology Hellas CERTH - ETHNIKO KENTRO EREVNAS KAI TECHNOLOGIKIS ANAPTYXIS | CERTH | RTO | GR |
| 3 | Centro Ricerche e Studi su Sicurezza e Criminalità | RISSC | RTO | IT |
| 4 | Expert System S.p.A. | EXPSYS | SME | IT |
| 5 | AIT Austrian Institute of Technology GMBH | AIT | RTO | AT |
| 6 | Fundacio Institut de BioEnginyeria de Catalunya | IBEC | RTO | ES |
| 7 | Istituto Italiano per la Privacy | IIP | NPO | IT |
| 8 | SYSTRAN SA | SYSTRAN | SME | FR |
| 9 | Stichting Katholieke Universiteit Brabant | TIU-JADS | RTO | NL |
| 10 | Dutch Institute for Technology, Safety & Security | DITSS | NPO | NL |
| 11 | Belgian Road Safety Institute | ISBR | RTO | BE |
| *Law Enforcement Agencies (LEAs)* | | | | |
| 12 | Provincial Police Headquarters in Gdansk | KWPG | USER | PL |
| 13 | Academy of Criminalistic and Police Studies – Kriminalisticko-Policijska Akademija | AoC | USER | RS |
| 14 | Home Office CAST | CAST | USER | UK |
| 15 | National Police of the Netherlands | NPN | USER | NL |
| 16 | General Directorate Combating Organized Crime, Ministry of Interior | GDCOC | USER | BG |
| 17 | Local Police Voorkempen | LPV | USER | BE |

Document History

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **VERSION** | **DATE** | **STATUS** | **AUTHORS, REVIEWER** | **DESCRIPTION** |
| V0.1 | 10/09/2018 | Draft | CAST DSTL | First draft |
| V0.2 | 14/09/2018 | Revision first draft | CERTH | suggestions are presented |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Definitions, Acronyms and Abbreviations

|  |  |
| --- | --- |
| **ACRONYMS / ABBREVIATIONS** | **DESCRIPTION** |
| **ANITA** | Advanced Tools for fighting Online illegal trafficking |
| **UC** | Use Case |
| **DoW** | Description of Work |
| **WP** | Work Package |
| **LEAs** | Law Enforcement Agencies (LEAs) |
|  |  |

Table of Contents

Executive Summary 11

1 Introduction 12

1.1 User-centred design approach 12

2 ANITA Use Cases 16

2.1 Overview 16

2.2 Use Case 1 – Counterfeit/falsified medicine, drugs and NPS 16

2.2.1 Scenario 1: Cyberspace Drug Trafficking (responsible partner: AoC) 17

2.2.2 Scenario 2: Drug Dealer (responsible partner: LPV) 18

2.2.3 Scenario 3: Delivery of Drugs at an Address (responsible partner: LPV) 19

2.2.4 Scenario 4: Counterfeit/falsified medicine, drugs and NPS (responsible partner: GDCOC) 20

2.2.5 Scenario 5: Counterfeit/falsified medicine, drugs and NPS (responsible partner: NPN) 21

2.2.6 Scenario 6: Counterfeit/falsified medicine, drugs and NPS (responsible partner: KWPG) 22

2.2.7 Scenario 7: Counterfeit/falsified medicine, drugs and NPS (responsible partner: KWPG) 23

2.3 Use Case 2 – Weapons trafficking 24

2.3.1 Scenario 1: Cyberspace Weapons Trafficking (responsible partner: AoC) 25

2.3.2 Scenario 2: Weapon 1 (responsible partner: KWPG) 25

2.3.3 Scenario 3: Weapon 2 (responsible partner: KWPG) 26

2.3.4 Scenario 4: Weapon 3 (responsible partner: KWPG) 27

2.3.5 Scenario 5: Weapon 4 (responsible partner: KWPG) 27

2.4 Use Case 3 – Trafficking related to terrorism funding 28

2.4.1 Scenario 1: Terrorism Funding 1 (responsible partner: LPV) 29

2.4.2 Scenario 2: Terrorism Funding 2 (responsible partner: LPV) 29

2.4.3 Scenario 3: Terrorism Funding 3 (responsible partner: LPV) 30

~~2.4.4~~ ~~Scenario 4: Terrorism Funding (responsible partner: GDCOC)~~ 31

2.4.5 Scenario 4: Terrorism Funding (responsible partner: NPN) 32

2.4.6 Scenario 5: Terrorism Funding (responsible partner: CAST/DSTL) 33

2.4.7 Scenario 6: Mikser (responsible partner: Aoc) 35

3 User Requirements 37

3.1 User Functional Requirements 37

3.2 Functional Requirements in Use Cases 41

4 Available Datasets 47

4.1 Introduction 47

4.2 Use case 1 data collection 47

4.2.1 Data requirements in Use Cases 48

4.2.2 Procedural aspects of data gathering 49

4.2.3 Analysis of available data in end-users repositories 49

4.2.4 Legal and ethical aspects 49

4.3 Use case 2 data collection 50

4.3.1 Data requirements in Use Case 2 51

4.3.2 Procedural aspects of data gathering 51

4.3.3 Analysis of available data in end-users repositories 51

4.3.4 Legal and ethical aspects 51

4.4 Use case 3 data collection 52

4.4.1 Data requirements in Use Case 3 52

4.4.2 Procedural aspects of data gathering 53

4.4.3 Analysis of available data in end-users repositories 53

4.4.4 Legal and ethical aspects 53

4.5 Data collected from on-line sources 54

4.6 Publicly available data 54

5 Conclusion 56

6 References 57

7 Annex 58

7.1 Use case 1 Pharmaceuticals 58

7.1.1 Keywords 58

7.1.2 data to generate connected to a persona 58

7.1.3 Further sources 58

7.2 Use case 2 Weapons Trafficking 59

7.2.1 Keywords 59

7.2.2 data to generate connected to a persona 59

7.2.3 Further sources 59

7.3 Use case 3 Terrorism Funding 59

7.3.1 Keywords 59

7.3.2 data to generate connected to a persona 59

7.3.3 Further sources 59

List of Figures

Figure 1: User-centred design life cycle 10

Figure 2: Human-Machine interaction in the ANITA project 12

Figure 3: Counterfeit pharmaceuticals 14

Figure 4: Illegal drugs 16

Figure 5: Weapons and Firearms 19

Figure 6: Glock 9, disassembled 20

Figure 7: Illegal firearms 21

Figure 8: Terrorism prospectus 23

Figure 9: Terrorist call for funding 25

Figure 10: Fundamentalist website 26

Figure 11: Propaganda page 27

Figure 12: ISIS doctrine 29

Figure 13: Considerations for ANITA datasets 40

List of Tables

Table 1: ANITA functional requirements 30

Table 2: Functional requirements per Use Case and Scenario 34

Table 3: Attribute frequency 37

Table 4: Attributes by use case 38

Table 5: Datasets - Use Case 1 42

Table 6: Datasets - Use Case 2 43

Table 7: Datasets - Use Case 2 45

Table 8: List of publicly available datasets 46

# Executive Summary

This document is a report on the work done within WP4, regarding the relevant use cases that will eventually sustain the specifications of the ANITA project, and its technical direction. Project implementation will be based on the descriptions given in the use cases, using them as a guideline with respect to the expected output from the end-users point of view. This document comprises all user requirements of the project, as they are depicted through the use cases. The use case analysis and specific scenarios, taken together, are intended to provide the basis of the requirements for the proposed infrastructure of the project.

This report contains a comprehensive description of the use cases provided, along with an analysis of the required functionalities that are required for addressing each of them. Specifically, the use cases were designed with the collaboration of the LEAs in the consortium, after in-depth discussion of their definition. ANITA end-users constitute law enforcement agencies that have extensive operational experience in the use cases described; therefore the scenarios can be safely considered as representing real events. Furthermore, the respective functionalities are in turn translated to project requirements that stem from the LEAs’ real needs, as they have been thoroughly discussed.

The use cases and scenarios that form the bulk of this document are non-technical high level representations. The intent is to illustrate the use cases through narrative stories and therefore show the significance of the project and how it relates to scenarios in real life. Besides the use case definition, a first draft of targeted modular functionalities is contained as well as mapping of the user requirements into technical ones. The most significant goal of this work is to define a primary set of user requirements. These will then help steer the technical developments that will be used in all other work-packages, including the implementation of services in the envisaged system, according to the defined functionalities.

This work should be considered as the guide document of the project, as all the essential requirements are specified here. It contains the final use case descriptions and the corresponding functional requirements that will be addressed by ANITA. Operational needs have been translated into functionalities and they are mapped to their respective services in order to ensure the usability of the final system in real investigative scenarios.

Outlining the document, a brief introduction is presented in Section 1, which explains the design approach that has been adopted in ANITA and aims at having the end-user as the reference point of the development cycle. This is a user-centric approach, whereby the users are actively involved in all the phases of the project, from setting the initial requirements to the expected final evaluation. Section 2 gives the use cases descriptions, where real scenarios are presented and highlight the required functionality of the ANITA system. This is the core of the document since it describes the use cases and the requirements for the ANITA services. Section 3 includes a detailed description of the identified user requirements. An initial outline of the ANITA datasets that will be utilized for experiments and to evaluate the final system is given in Section 4. Conclusions are presented in Section 5.

# Introduction

## User-centred design approach

ANITA’s primary goal is twofold: a) To boost the LEA’s investigation process and to significantly increase their operational capabilities, by introducing a set of innovative tools for efficiently addressing online trafficking challenges (namely online data source analysis, blockchain analysis, Big Data analytics, knowledge modelling, incorporation of human cognitive function in the analysis pipelines, user-oriented intelligence applications), and b) To significantly facilitate the novice officers training process and to optimize the learning curve (by collecting, integrating and re-using knowledge from multiple expert officers and through the development of a recommendation functionality to transfer the acquired ‘know-how’ to the new officers).This will increase the efficiency of current investigation practices that are performed by LEAs and counter-terrorism agencies. The ANITA system will provide a variety of analytical tools that will integrate human cognitive principles and allow investigators to remain ‘in-the-loop’ of the overall analysis process. This will allow LEA operatives to more quickly analyse and extract relevant information from the vast amounts of data available. ANITA will provide a set of tools that will facilitate the analysis of this data, allowing users to identify patterns for spatial, temporal and causal correlations, and supporting decision-making processes for countermeasures to undertake. With the development of ANITA the completion of these tasks is expected to be significantly expedited, therefore, during the development and evaluation process, users are expected to have a key role in the loop, providing feedback from real investigations. The ANITA consortium will follow a user-centred design approach during the development. An overview of this approach is illustrated in Fig. 1.



Figure 1: User-centred design life cycle

According to the flow depicted in Figure 1, the use cases for ANITA are at the top of the development cycle. Therefore, the ANITA project development revolves around the use cases, which have been designed and refined after comprehensive discussions with the end users in the consortium. Specifically, in the relevant discussions about the definition of the use cases, Home Office Centre for Applied Science and Technology (CAST now DSTL), Academy of Criminalistic and Police Studies (AoC), General Directorate Combating Organised Crime (GDCOC), Provincial Police Headquarters in Gdansk (KWPG), Local Police Voorkempen (LPV) and National Police of the Netherlands (NPN) were actively involved in finalising the use case description and requirements. Therefore, the described scenarios in this document originate in real events and depict challenges that law enforcement authorities face in everyday investigations. As a consequence, the scenarios that are presented in the next section are not artificially shaped, but carefully drafted in close collaboration with the project users to ensure a practical and useful functionality for the ANITA system. Following in depth discussions with the end users, the following issues and challenges were identified, regarding development of the ANITA services:

* Usability
* Complexity of operation
* Task Match
* “Analytical research”
* Implications
* Privacy and security
* Usability in context
* Response time
* Efficiency
* Interoperability

These issues, along with other aspects of the use case definition were carefully considered during the elaboration of the use cases. Privacy and security are key aspects of ANITA that are established in this document and thus are core values for the whole consortium. A deep analysis of privacy and security is presented in a different work package and therefore will be separately reported in a different deliverable. In order for ANITA to be effective, the interaction between human and machine needs to be as efficient as possible. The tools developed need to be intuitive to use and provide results that are easy to understand and integrate with other systems



Figure 2: Human-Machine interaction in the ANITA project

From discussions between the consortium members, it was clear that in all use cases the functionality must centre on a human operator – typically an officer investigator, as such the system must allow for active human involvement. Whilst the level of autonomy in systems has developed to a very high level, we are still (from a technical standpoint) a long way from a truly autonomous system in this arena. Also ANITA needs to encompass the context and experience provided by law enforcement officers that is hard to quantify.

Finally, the use cases where analysed in accordance with conventional usability schemas involving a user centred review of all aspects under consideration including:

* Review use case description with respect to the end-user’s needs
* Initial user requirements analysis
* Define acceptability and usability criteria
* Utilizable engineering
* Design functionality and usability that lives up to the user criteria, whilst respecting constraints as well as ethical and privacy requirements

Following the usability engineering approach, the next steps in the development of ANITA will be:

* Usability and acceptability evaluation
* Evaluation and testing of the final prototype system against real data from investigations
* Experimental results of the system performance
* Loop through requirements analysis and usability engineering

Before the ANITA project moves onto the next phase, some additional steps must be considered in order to close the user-centred design loop. This will take place during trial and testing of the final version of the project. These additional steps are:

* Reviewing the prospective progress in each phase
* Implementing rest of the phase requirements
* Strong involvement of end-users in all stages of the project (user requirements, use cases, data, documents collection, testing and validation, training etc.)
* Trial implementations (usability, acceptability tests)
* Tuning up and integrating the system
* Supporting system management
* Data collection on usage, usability and acceptability
* Providing feedback to development teams for the next release

Some of the steps mentioned above are beyond the scope of this document and will be reported in other phases of the project. These outcomes will be reported on via the deliverables associated with their respective phases.

# ANITA Use Cases

## Overview

The ANITA use cases reflect how the Internet is often utilized to promote and support online trafficking acts. Three use cases have been identified from the LEAs in the project, based on their experience and taking into account the current trends. Use case descriptions are generic enough so as to allow leveraging of the envisaged functionality in other related scenarios or events and specific enough in order to illustrate actual events. The use cases will be used for the design, evaluation and user testing of the overall ANITA system. The use cases address all the aspects of technical developments stated in the DoW.

## Use Case 1 – Counterfeit/falsified medicine, drugs and NPS

The market for medicine and drugs constitutes a remarkably large share of world trade, both those purporting to be for pharmaceutical purposes or as narcotics. Apart from, the increasingly worldwide interest for New Psychoactive Substances (NPS) which have become a global phenomenon – over 100 countries have reported one or more NPS substances. There is also a huge trade in medicine-related substances. This coupled with the growth of Internet services provides a profusion of opportunities for anonymous and quick online illegal transactions. The result is that medicine trafficking costs millions to the European Nations, and also has an adverse impact on the public health of European citizens.



Figure 3: Counterfeit pharmaceuticals

There are many levels of criminal activity encompassed by this use case: from simple fraud, fake goods / non-supply; through harvesting of payment details; to supply of dangerous and / or illegal substances. This complexity is compounded by the existence of a legitimate online pharmacy trade operating entirely within the law. However, it must also be noted that legal access to certain products varies from country to country.

Central to investigating this illegal pharmaceutical trade is being able to follow the money trail both through e-payments and cryptocurrency transactions and then being able to link that with social media profiles, etc., in order to identify real people and groups. This will then allow arrests to be made and these illegal activities to be disrupted. ANITA will provide tools for automated detecting of financial markets through such means as blockchain analysis through the Public Web, Deep Web and Dark Nets. This is a task currently performed manually which is resource intensive and prone to error. In addition ANITA will also be able to automatically detect, monitor and analyse heterogeneous multimedia large-scale data from online retrieved sources, but also increase its accuracy by learning from experienced investigators. The system will also be able to transfer that learned response to less experienced operators and therefore assist knowledge transfer.

### Scenario 1: Cyberspace Drug Trafficking (responsible partner: AoC)

**Narrative:**

Catch-all scenario for several lower-level crimes May provide starting point or additional information when investigating larger cartels or serious crime groups.

I’ll work this one up nearer the end in order to fill in gaps.

* + Users knowingly click links and enter payment details:
    - One time fraud, goods not delivered
    - Multiple frauds through harvesting and selling payment details
    - No fraud but fake / falsified drugs delivered
    - Drugs supplied not licensed in this region
    - Drugs supplied cause illness or death

**Available forensic data for investigation:**

* Most data such as emails belong to private individuals so not available
* Examples of phishing emails downloaded from the deep web.

**Conventional practices followed by LEA:**

Specialist units to examine metadata and use of specialist TOR browsers etc. to investigate information from victims. Various techniques deployed depending on the exact circumstances, but likely to include some or all of the following:

* Manual review of emails, including all associated media and data. This can include headers, links and embedded objects such as images, videos, logos, etc.
* Investigation of links etc. to determine owner of fraudulent site.

Following the trail from the victim and attempting to build a map of the links from the perpetrator’s website in order to find links, similarities etc. More detail once I have built up the scenario.

### Scenario 2: Drug Dealer (responsible partner: LPV)

**Narrative:**

After a routine vehicle stop, the occupants were found to be drug users and during interview they provided the address from which they had bought the drugs they were arrested with. Investigation of the address given showed that observation of the premises would be difficult from both logistical and operational reasons. The location was difficult to observe covertly without being detected and long term surveillance would have required extensive permissions from within the judiciary. In Belgium the observation of a person for more than 5 consecutive days or five days in a month requires an added layer of authorisation. This led by necessity to gathering information on the drug dealer from open source information on the internet and data held on police databases. After the drug dealer was arrested the house was searched and it was found that the drug dealer had some knowledge of cryptocurrencies, specifically Ethereum. Evidence of cryptocurrency transactions was found but it was not possible to trace the destination of the transactions. This information could have been extremely important as other evidence discovered during the search showed that at least three more dealers were involved. The analysis tools in ANITA could have created a map of who knew who and what the levels of interaction were between the different parties.



Figure 4: Illegal drugs

**Available forensic data for investigation:**

* A great number of videos that are hosted on YouTube
* A variety of keywords are provided and can be used in order to collect visual content such as videos and images from online sources (google images, dark web, YouTube)
* Two Facebook accounts
* Transaction data
* Two phone numbers
* Cryptocurrency wallets
* Visual content associated with social media accounts

**Conventional practices followed by LEA:**

Primary investigation as per standard police work – information gathering and observation, if sufficient evidence gathered then this can lead to arrest, house search etc. In the primary phase formal investigation is limited to information held on police databases. Online research is at the discretion of the investigator and there is no formal procedure associated with this.

* Trace the cryptocurrency transactions
* Establish or disprove links between persons identified in the case
* Establish the level of interaction between persons identified as being linked.
* Identify other persons organisations linked to the case

### Scenario 3: Delivery of Drugs at an Address (responsible partner: LPV)

**Narrative:**

Narcotics delivered through thepost to a specific PO Box but with different recipients. Once the owner of the PO Box was identified it was found that his name corresponded with one of the recipients. He was arrested and during interview he admitted buying narcotics on a Dark Web site using Bitcoin. Though he was arrested and charged for narcotics offences further investigation into who was trafficking the drugs and any deeper investigation into the supplier was not conducted. The investigation was complicated, not only by the fact that the offence occurred over two different geographical areas of police jurisdiction, but also because of the resources available, and other operational pressures.

**Available forensic data for investigation:**

* Bitcoin transactions
* Dark Web market URL-addresses
  + https://www.deepdotweb.com/marketplace-directory/categories/top-markets/
  + Addresses can be collected from text documents/Google maps (need to be international addresses of which one is linked to our municipality and another is linked to a municipality close by)
  + Bitcoin address can be collected from any online/offline wallet (Exodus, Jaxx, Ledger Nano ...) or online from Coinbase.com/LiteBit.eu/Bitcoin.org ...
  + Bitcoin transactions can be collected from https://www.blockchain.com/explorer (a bitcoin address is needed)
* Open source data

**Conventional practices followed by LEA:**

Primary investigation as per standard police work – information gathering and observation, if sufficient evidence gathered then this can lead to arrest, house search etc. In the primary phase formal investigation is limited to information held on police databases. Online research is at the discretion and aptitude of the investigator and there is no formal procedure associated with this.

* Trace the cryptocurrency transactions
* Establish or disprove links between persons identified in the case
* Establish the level of interaction between persons identified as being linked.
* Identify other persons organisations linked to the case

### Scenario 4: Counterfeit/falsified medicine, drugs and NPS (responsible partner: GDCOC)

**Narrative:**

In July 2016, Jane was browsing the internet and looking to buy drugs from a darknet marketplace. Access to such markets can be achieved in a number of ways. She was looking for a surface website that provided listings of ‘onion’ addresses for darknet markets, thus enabling ready access. She accessed one of the markets through an invitation (this is the only way possible to enter that market). While browsing for the drugs she was looking for she found that each of the websites had its own method of classifying what it had on sale, she was looking for opioids. Utilising anonymization services to browse the web without revealing her identity or location, Jane used ‘GRAMS’ as one of the search engines for Tor-based darknet markets. She entered the market and registered with an email [janedoe@jane.com](mailto:janedoe@jane.com) and the username, displayed on the market was DoeJane. She found the section in the marketplace marked as ‘opioids’ and established contact with the trader. The trader’s nickname was ‘Smithsopioids’.

They started to communicate using secure applications, which removed the risk of such communications being intercepted. The messages between the seller and the buyer were done using a crypthographic system, such as public key encryption. The payment was again through anonymous transactions. Jane sent the money through a crypto-wallet and the virtual currency was Monero. She received the drugs one week after she done the transaction, as agreed with the seller.

**Available forensic data for investigation:**

Suggested open source or constructed data, some suggestions can be found in section 7.1

**Conventional practices followed by LEA:**

Primary investigation as per standard police work – information gathering and observation, if sufficient evidence gathered then this can lead to arrest, house search etc. Looking for / at: Witness information, phone ownership, timeframes, names, roles on the Organised Crime Group. No specific online protocol.

* Trace the cryptocurrency transactions
* Establish or disprove links between persons identified in the case
* Establish the level of interaction between persons identified as being linked.
* Identify other persons organisations linked to the case

### Scenario 5: Counterfeit/falsified medicine, drugs and NPS (responsible partner: NPN)

**Narrative:**

None given

**Available forensic data for investigation:**

* Open source data

**Conventional practices followed by LEA:**

* Cases which have not met a legal minimum (complaints, anonymous information, witness information from other case) – all information will be gathered
  + i.e. GPS, nickname, username, organisation names, names and roles within the network, timeframes
* Checked in the available systems
* Determine if any related personal information is available
* Determine if the personal information can be used in internet searches
  + Risks of sharing search information in browser etc.
* If useful information found, investigator to determine
  + what can be used for further investigation
  + is information related to the investigation or a false positive
* Information downloaded and made available in a forensic way
* At agreed points Public Prosecutor contacted for legal approval

To be able to search for and link:

GPS, nickname, username, organisation names, names and roles within the network, timeframes

### Scenario 6: Counterfeit/falsified medicine, drugs and NPS (responsible partner: KWPG)

**Narrative:**

NPS and drugs trafficking is one of the fastest evolving branches of trafficking in Poland. One of the main reasons for this situation is the constantly developing market for such goods. There are some individuals who want to make profit this way, but one resourceful man saw an opportunity for himself and built an enterprise which was working on the edge of the law. He was selling an incredible amount of NPS substances, he had fully operational internet sites (and a back-up in case the authorities shut down his main site), accountants, co-workers and he was investing, earned money in to his semi legal business. The internet site on which he was making transactions was named www.odczynnikichemiczne.net.pl. He was importing the NPS from China, Middle East, Balkans, and other countries. He was selling in quantities that ranged from one gram to even a tonne to one buyer. In a year time his profits could be counted in millions of PLN. He accepted payments in cash, money transfers, money orders, crypto currency, and foreign currency. He had several methods of money laundering and several dozen co-workers each with a large number of customers. Some of those co-workers expanded the range of offered items adding prescription drugs mainly anabolic steroids and potency drugs. Those prescription drugs were produced or branded as products from Malesia, Turkey, Bulgaria, Russia, Ukraine. The anabolic steroids were in form of pills and injections.

**Available forensic data for investigation:**

* Open source data

**Conventional practices followed by LEA:**

Once the official investigation starts, operational officers may still be involved searching for further information but the prosecutor is now leading. In this particular investigation, the computers and all the data from the website are considered available (from the owner of the website that is considered arrested).

The main direction of the investigation is to identify and trace all the individuals who have purchased items form the website. It is also necessary to prove that the website owner is launching new products dangerous to life or health, and that he is aware this is the case.

Based on the data available it is not always possible to trace the individual purchasers, so investigators have to conduct manual internet searches for particular phrases found on the websites and any other small pieces of information available about each buyer. Part of this search for information included trying to find out where photos or movies were taken and also identify the people in them.

* Analysis of links
* Trace purchasers
* Analysis of currency payments
* Similarities of phrase
* Similarities of images
* People in images
* Places in images
* Unusual quantities

### Scenario 7: Counterfeit/falsified medicine, drugs and NPS (responsible partner: KWPG)

**Narrative:**

Since the start of Europe’s migrant crisis in 2015, the number of immigrants in Poland has also increased. It has been reported that some immigrants who either due to their criminal background or by the lack of other opportunities and difficult situation decide to get involved in crime.

This criminal activity sometimes takes the form of drug, NPS and pharmaceuticals trafficking. Organised criminal groups arrange flats, cars, satnav systems and phones, offering “easy money” to the immigrants described above, describing the merchandise as legal. The job involves answering the phone and delivering the goods to the buyer, collecting extra supplies as parcels from post lockers when stocks run low. The customers are targeted on internet sites, forums and social media. These advertisements are often sent from a variety of IP addresses but the layout and phrases are similar to each other. The prices are set and the immigrant keeps the delivery fee as earnings. If the immigrant is arrested they have not met and do not know the names or faces of those running the operation.

Often to ensure the loyalty of the immigrant the criminal gang running the operation will keep their passport or identity papers.

**Available forensic data for investigation:**

* Images have been provided that also appear on the internet; the task would be to find other instances of the image and or similar images.
* IP addressses
* Graphics
* Phrases
* Pictures
* Products
* Open source data

**Conventional practices followed by LEA:**

1) Judge the information –if it is reliable, if there are more than one similar advertisements with similar phone numbers posted in the internet.

2) Start operational work –gathering intelligence, other information, confirming the source, trying to establish all individuals who are involved, checking the phone billings of numbers given in ads, If necessary tapping or other ways of surveillance, such as observe the package, change the package, etc.

3) When all the information is enough to launch the case, all information is transferred to the prosecutor’s office

4) Official investigation starts, operational officers may still be involved and search for information, but from now on the prosecutor is involved

5) In this use case, if officers manage to establish the individuals and prosecutor agreed the perpetrators may be arrested and brought before the judge who decides if they should be arrested and detained for 1, 2 or 3 months

6) Police officially gather evidences and lead the case which may means hearing the witness, examining data sources like criminals’ phones, criminals’ computers, analyse all internet data and present conclusions to the prosecutor

7) Bring the accusation to prosecutor

* Analyse data from various sources looking for similarities
  + Graphics, phrases, pictures, products etc.
* Trace emails, websites etc.
* Link all the above information

Draw connections and hierarchy from the above information

## Use Case 2 – Weapons trafficking

Trafficking of weapons and firearms within Europe, especially in support of terrorist activities, has increased in the last few years to the point that it is one of the nine priority crime areas under the surveillance of Europol. Weapons trafficking, is often a part of wider criminal activity, exploiting existing criminal routes and networks, thus exposing this activity can often disrupt much larger criminal organisations. Unlike some more transient illegal commodities, weapons and firearms are durable which means once they have entered the criminal arena they can be circulated and used for many years, causing a long term problem. Firearms are frequently used by organised crime groups when trafficking human beings to intimidate and coerce their victims.

The regulation of firearms across Europe is extremely complex, and relatively simple modifications can render a particular weapon legal or illegal, which makes it more difficult to distinguish between legal and illegal firearms trafficking. As with all criminal activity, the anonymity offered by Deep Web and Dark Net market places has attracted those trafficking weapons. This also allows a wider variety of weapons to be offered than could be accommodated in a physical trade.



Figure 5: Weapons and Firearms

### Scenario 1: Cyberspace Weapons Trafficking (responsible partner: AoC)

**Narrative:**

Investigationson the dark Netreveal a number of advertisements for illegal firearms, though these can be hard to spot. Sometimes the advertisers are selling legal replica firearms or the website may use some kind of wildcard technique in order to only display real information to those who know how to access it. Mostly blockchain payments in cryptocurrencies including bitcoin and Ethereum

**Available forensic data for investigation:**

* +++

**Conventional practices followed by LEA:**

Specialist unit for terrorism and extremism would use investigators with specialist expertise in firearms and weapons. Also specialist cyber investigation teams as detailed in use case 1 scenario 1.

* Identifying people or objects from images
* Finding similar images
* Finding similar phrases
* Linking people with images, objects, websites etc.

### Scenario 2: Hidden parts (responsible partner: KWPG)

**Narrative:**

Following information from an informant, several adverts from internet auction sites and local announcement sites were investigated where the vendors was offering components of firearms – for example trigger mechanisms. These advertisements were posted on general interest sites rather than specialist forums where you would normally expect them. The vendor was contacted anonymously by an investigator posing as a buyer. The vendor asked several questions to try and find out if the buyer was genuinely trying to buy an illegal firearm. Once he was satisfied this was the case the transaction was initiated. The transaction was split into several stages with payments at each stage and more parts sent to assemble the illegal firearm. The parts were concealed in several parcels, sent from a variety of different addresses, mixed with other items and sent to a mailbox not a residence. Investigation of the adverts showed they were all similar in layout and style. The seller had attempted to conceal this connection by posting the adverts from different devices and from a variety of public hotspots.

**Available forensic data for investigation:**

* Emails
* IP addresses

**Conventional practices followed by LEA:**

Standard policing methods, no mention of specialist protocols or tools.

* Identifying people or objects from images
* Finding similar images
* Finding similar phrases
* Linking people with images, objects, websites etc.
* Link postings to times, locations etc.



Figure 6: Glock 9, disassembled

### Scenario 3: Mock ups (responsible partner: KWPG)

**Narrative:**

This scenario is an example of criminals often being involved in more than one type of crime. In this instance the individual is arrested for identity theft and fraud. During the search of his address several firearms, sniper rifles and black powder weapons are found, along with over a 1000 rounds of ammunition in various calibres. Searches of his computer and mobile phone revealed evidence that he was obtaining weapon parts, weapons and ammunition from a website. He was then contacting people with similar military interests and exchanging or trading weapons and ammunition with them. However, on the internet pages the weapons and ammunition were presented as mock-ups or components, e.g. bullets shown as empty casings. This case is further complicated by the fact that the black powder gun – as a collector’s item – is legal in Poland, though this loophole is starting to be abused by football fans who are obtaining and using them illegally.

**Available forensic data for investigation:**

* Emails
* IP addresses

**Conventional practices followed by LEA:**

Standard policing methods, no mention of specialist protocols or tools.

* Identifying people or objects from images
* Finding similar images
* Finding similar phrases
* Linking people with images, objects, websites etc.

### Scenario 4: Show off (responsible partner: KWPG)

**Narrative:**

During a routine trawl of social media, police officers found a man’s profile with photos of him posing with a weapon. Comments on the photos boasted of how powerful the weapon was. Also within the comments were phrases that constituted hate speech. Further investigation into the person revealed that he may be remodelling sport or air weapons to use combat ammunition. When the man’s apartment and workplace were searched multiple collecting and sporting weapons were recovered along with some noise makers. Ammunition, weapon parts and equipment used for adapting weapons and ammunition manufacturing devices were also secured. Documentation seized showed that weapons had been bought legally in the Czech Republic and Poland, remodelled and then sold. Experts confirmed that some of the weapons had been adapted for combat ammunition.



Figure 7: Illegal firearms

**Available forensic data for investigation:**

* Emails
* IP addresses

**Conventional practices followed by LEAs:**

Standard policing methods, no mention of specialist protocols or tools.

* Trawl for suspect images
* Trawl for hate speech
* Match similar images
* Match phrases
* Trace purchases / purchasers
* Link individuals / organisations and show level of interaction

### Scenario 5: Gun Education (responsible partner: KWPG)

**Narrative:**

This scenario is not really a crime more an example of possible missed opportunities to spot criminal or terrorist activity early and therefore prevent a major crime. An individual wishing to purchase a weapon searches the internet for technical information about the weapon and firearms: kind, specifications, construction details, etc. So far this is legal with no warning flags, however, as the questions get more specific – can the weapon be divided into sections, how these sections can be transported, what is the best weapon for long range shooting etc. now the questions are getting more concerning and could point to potential criminal or terrorist activity. The problem being most of the replies are via private message or through unrelated email addresses from domains outside the national jurisdiction such as gmail.com (USA) or mail.ru (Russia).

**Available forensic data for investigation:**

* Email addresses
* URLs
* Forum posts, etc.

**Conventional practices followed by LEAs:**

This is currently a largely untapped area for policing as the data volume is too large to realistically allow speculative investigations to provide a reasonable investment on time. Though some forums are monitored at least occasionally this is a manual task so extremely resource intensive.

* Ability to search for phrases
* Ability to search private messages
* Ability of searching mail servers outside of national jurisdiction

## Use Case 3 – Trafficking related to terrorism funding

Terrorist organisations require large amounts of funds in order to support themselves and their illegal activities. Unfortunately, the rise in online commerce and all that it facilitates in both domestic and international trade, not only makes it easy for legitimate businesses but also provides many opportunities for illegal funding. There are many ways in which terrorist organisations can obtain funding some of the most common are:

**Terrorist-related financial networks:** either through the same organisations or between terrorist organisations, exchanging goods cash or services. The money transactions are usually through blockchain or cryptocurrencies to preserve anonymity and hide from LEAs.

**Fake charities:** appearing to be about a major humanitarian issue such a medical assistance to war victims they lure in potential donors and can collect significant amounts of money. These are often well researched and credible professionally constructed sites, with activities, objectives and false information. Donations to these sites are normally on the surface web made through legal means such as bank transfer, PayPal and credit card. The money is then transferred via various means to the well-hidden terrorist organisation accounts.

**Fund-raising campaigns:** these are run over the public and Deep Web as well as the Dark Net. These campaigns are specifically targeted at other terrorists, extremists or sympathisers and often only last for a few days or even hours in order to escape notice by LEAs. The majority are established in the Deep Web or on Dark Nets again to escape notice from LEAs.

### Scenario 1: Terrorism Funding 1 (responsible partner: LPV)

**Narrative:**

An asylum seeker staying in an asylum shelter begins to show behaviour that indicates either they are a terrorist or that they are being radicalised. This person is making use of the facilities of the asylum shelter such as phone/smartphone and computer. Social services then notify the police of this disturbing change in behaviour and are asked to intervene.

**Available forensic data for investigation:**

* A great number of videos that are hosted on YouTube site
* A variety of keywords are provided and can be used in order to collect visual content such as videos and images from online sources (google images, dark web, YouTube)
* Links between people and levels of interaction
* Trace websites, email addresses etc.

**Conventional practices followed by LEA:**

We interview the representative of the social service and the asylum shelter, and other concerned persons if necessary to form an image of the situation and the target. It is the investigators choice to perform an internet investigation, and the investigator also determines the extent of this investigation.

We then interrogate the target, perform a house search and confiscate any material deemed useful for the investigation. Based on additional elements, the person will be asked to leave the country, will be incarcerated or let go.

### Scenario 2: Terrorism Funding 2 (responsible partner: LPV)

**Narrative:**

A young girl at school starts to display signs of radicalisation, such as starting to wear a burka. The family is known to be very isolated and does not wish to engage in conversation.

**Available forensic data for investigation:**

* A great number of videos that are hosted on YouTube site
* A variety of keywords are provided and can be used in order to collect visual content such as videos and images from online sources (google images, dark web, YouTube)

**Conventional practices followed by LEA:**

The school’s head master will be interviewed, the target’s head teacher and other concerned persons if necessary to form an image of the situation and the target. It is the investigators choice to perform an internet investigation, and the investigator also determines the extent of this investigation.

An effort will be made to talk to the family if possible and start some form of assistance for the family if necessary. If this procedure is not effective, the target will be interrogated, and possibly a house search will be performed for further evidence gathering and also confiscation of elements relevant to the investigation.

Based on what additional elements are found, this could lead to the person being incarcerated (or locked up in an institution for minors) or let go.

* Search for similar images
* Search on IP addresses
* Dark web search



Figure 9: Terrorist call for funding

### Scenario 3: Terrorism Funding 3 (responsible partner: LPV)

**Narrative:**

This case is similar but involves a returning TFT. This individual displays disturbing behaviour and signs of radicalisation. He lives a very isolated lifestyle and does not wish to engage in conversation.

**Available forensic data for investigation:**

* A great number of videos that are hosted on YouTube
* A variety of keywords are provided and can be used in order to collect visual content such as videos and images from online sources (google images, dark web, YouTube)

**Conventional practices followed by LEA:**

Initially we will interview neighbours and monitor where the suspect’s vehicle is driven. If this continues or heightens the level of concern we will try and speak to the suspect if possible. If the above does not lead to a satisfactory conclusion then the suspect will be interrogated and possibly a search of their premises will be conducted. Following these procedures and depending on what items are found, this could lead to the being prosecuted or released.

* GPS data
* Vehicle location
* Internet sites visited
* Internet searches conducted
* Links between people and hierarchy of interaction
* Association with keywords

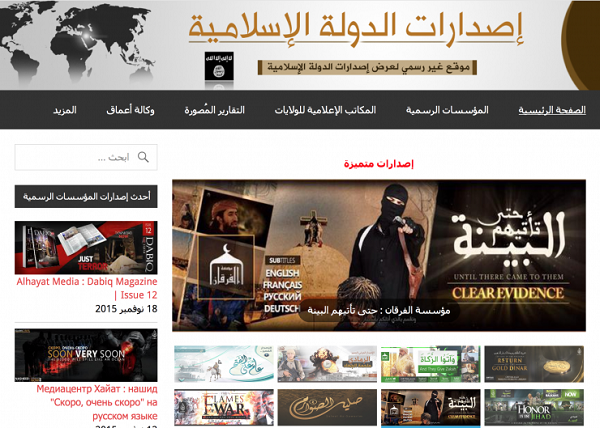


Figure 10: Fundamentalist website

### Scenario 4: Terrorism Funding (responsible partner: NPN)

**Narrative:**

Concerns an ongoing investigation so details are withheld.

**Available forensic data for investigation:**

None provided



Figure 11: Propaganda page

**Conventional practices followed by LEAs:**

Detection:

* Banks and other financial institutions have a legal obligation to identify and report certain possibly illegal (suspect) transactions. These are scanned, investigated and classified by the Financial Intelligence Unit. They determine whether the transaction is possibly terrorism related.
* Subjects might be known as terrorism related and facilitating. Based on that a financial investigation can be started
* Detection can be the result of actual investigations or accidental discoveries.
* There is currently no systematic, broad scanning of internet / social media.

Selection

* The signal is investigated (light means) by the police to enhance the case
* This means full inquiry upon the police information and inclusion of open source (internet / social media, etc.) to supplement the information. Related persons (when known) are included. No information from taxes and other governmental institutions may be included at this point.
* This case is brought to a committee (incl. DA) to determine the further action: official investigation by the police and/or combined action with other governmental units; no further action, etc.

Investigation police

* When the decision is made that a full police inquiry will start additional means will be used including telephone/wiretap, historical communication data/user information related to the phone number, beacon information, full financial information, etc. will be used. Telephones, computers etc. are possessed and investigated when possible. These investigations are largely done by hand.
* Audio information, but also written texts, documents, etc., in a foreign language is translated by a human interpreter (this is done to make sure we use the right context of the language and words used)
* Information is gathered from other governmental institutions (taxes, etc.), banks. Currently limited use is made of information from deep web, block chain, etc.
* Broad scanning of internet / social media
* Links with other watch lists
* Similarities between watch lists
* Comparison of audio data
* Physical location of phone
* Full digital footprint

### Scenario 5: Terrorism Funding (responsible partner: CAST/DSTL)

**Narrative:**

This case study relates to an investigation into payments sent from the UK to a Turkish based suspected terrorist financer. Initial enquiries identified that the facilitator had originally used WESTERN UNION to receive funds from multiple countries and senders. After WESTERN UNION closed his account, he was sent money via MONEYGRAM. The investigation into the WESTERN UNION payments resulted in the conviction of husband and wife, who had sent money to their nephew, who was associated with Da’esh.

After this, NTFIU commenced an investigation into the MONEYGRAM payments.

One payment to the facilitator of £1,500 was sent by a person who was arrested and interviewed, and his account indicated that he had been asked to send the money by person C, with whom he went to College and University.

Person C was arrested and multimedia devices belonging to her were seized at her home address in August 2016. She was interviewed and admitted that she had asked another person to send the money to her husband as it was not culturally appropriate for a female to send the money. She said that her husband was in TURKEY having started a new life, and added that she had no idea that he was or had been involved in terrorism.

She said she had never sent any other money to her husband, and claimed the money came from her bank account, and money she had laying around the house. NTFIU downloaded numerous multimedia devices belonging to person C. An examination of one of the phone handset’s identified a chat conversation, between her and her husband. The message and chat contained imagery of her husband (person D) holding an AK47 firearm and relaying Da’esh type rhetoric. Further messaging recovered showed him making specific requests to person C asking her to purchase items from AMAZON for him and other persons located with him.

An examination of person C’s bank and AMAZON accounts showed she had purchased the items that her husband had requested. Further enquiries with WESTERN UNION showed that a third party had also sent other money transfers (to three different named persons in TURKEY) for person D which totalled just over £10,000. Person C had also used a former female colleague, to send a money transfer of £2,000 to another suspected terrorist funding facilitator.

Person D’s wife was further interviewed, during this interview she still claimed not to know that her husband was or is involved in Terrorism or armed conflict, despite detailed pre-interview disclosure to her legal advisor. Person C changed her account part way through the interview and claimed that after she found out that her husband was involved in terrorism she didn’t send any further money or property to him.

Person C was charged with two counts of being concerned in arrangements to fund terrorism, one count to cover money, and another count to cover items purchased through AMAZON for her husband She appeared at Westminster Magistrates Court, and the Central Criminal Court. The case was sent to Woolwich Crown Court for trial.

Person C subsequently pleaded guilty to the offences and was sentenced to 5 years 3 months imprisonment.

**Available forensic data for investigation:**

+++

**Conventional practices followed by LEA:**

UK LEA has information sharing agreements with most financial providers and LEA’s across Europe and the rest of the world. These procedures enable the sharing of ‘suspicious’ transactions which are subsequently investigated. Intelligence of Terrorist organisations activities is collected across a whole host of LEA agencies. This information is assessed and graded to establish if offences are identified. In addition UK LEAs have web searching tools for identifying ‘key word’ searching to identify terrorist activity. This data is then reviewed by the investigation team to identify specific evidence.

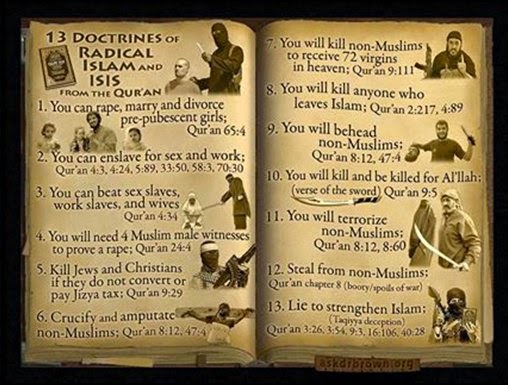


Figure 12: ISIS doctrine

* Keyword searches
* Links between people and hierarchy
* Image searches

### Scenario 6: Mikser (responsible partner: Aoc)

**Narrative**

This describes a common method of terrorism funding used repeatedly by various groups. The victim (Juliusz) is drawn in to a surface website having received several adverts for technical goods offered at extremely discounted prices – often up to 90% reduction. Juliusz received these adverts from both websites and emails. This was the first point at that Juliusz was suspicious, but the discount was too tempting. The website had an online payment system and Juliusz could pay by normal means such as his Mastercard, Maestro or similar. The next point that was suspicious was when the confirmation of the transaction sent to Juliusz via SMS and the price was shown in Chinese Yuan rather than the transaction he had completed in Euros. Juliusz waited two months for his goods during which time the website owner provided a very plausible and consistent order tracking system that seemed to show the status of the delivery – held by customs etc. After this time with still no delivery of the goods Juliusz threatens to contact the authorities. At this point the website owner takes down the website and disappears. This long after the original transaction was authorised neither bank could stop the transaction and the terrorist organisation got their money.

**Available data**

* False advertisements
* Images
* Phrases

**Current practice**

This would follow the lines of conventional policing:

* following the money trail as far as possible,
* looking for similarities in
  + images
  + phrases
  + layout
* tracking associates
* identifying other victims
* Follow the money trail
* Tracking associates
* Identifying other victims
* Looking for similarities in images, phrases, layout

# User Requirements

## User Functional Requirements

ANITA must be able to access information from the Public Web, Deep Web and Dark Net. All output must be forensically sound with robust audit trails and ability to import or accept links/references to other chain of custody records.

According to the user needs, as they have been illustrated through the previous section, ANITA is designed to satisfy the functional requirements reported in the following table. The requirements are categorically presented.

Table 1: ANITA functional requirements

|  |  |
| --- | --- |
| **User Requirements** | **Description** |
| **Data sources and stream analysis** | |
| Data source risk assessment in the Surface Web, Deep Web and Dark Nets | ANITA must provide methods and indicators for assessing risk associated with data sources. These risks are twofold, both the veracity of the data and the likelihood that the interrogation originates with an LEA. The methodology employed must consider the position of the source (Surface Web, Deep Web or Dark Net) and refine its policies accordingly. It must also consider the update frequency of the data and website appearance. |
| Black markets discovery and monitoring | A set of services for crawling relevant contents from Deep and Dark Nets, focussed on the discovery and monitoring of black markets. These will utilise data mining techniques within selected sources to identify specific products, volumes or behaviours. |
| Content acquisition from Surface Web and pre-processing | Based on open technologies like Apache Nutch and Apache Camel. This crawler will be able to acquire content both from open sources on Surface Web, like RSS, Social Networks (Twitter, Facebook), Web, Blogs and from local sources like files in the most common formats (i.e.: pdf, word, excel, power point, mail etc.) stored in folders or repositories. The crawler will be able to extract heterogeneous resources and specific taxonomies will be defined for contents pre-processing. |
| Blockchain analysis for illicit activity discovering | ANITA will contain methods for the detection, tracking and monitoring of monetary transactions, for the purpose of recognising trafficking activities. Whilst focussed on cryptocurrencies it will not be limited to them. This data can then be included with other information in order to identify users of black markets and reveal their identities. |
| Construction of source network and filtering | Using the output from the tools described above ANITA will be able to reconstruct links and relationships from the sources discovered in Surface Web, Deep Web and Dark Nets. Specific filtering mechanisms will be built in to constrain initial responses to a preliminary overview of relevant connections. These links can then be used to target and inform further investigation. |
| **Big Data analysis and analytics** | |
| Multilingual Text analysis | ANITA will incorporate an advanced Big Data oriented textual resource oriented framework incorporating services for categorization and semantic information extraction from huge amounts of textual resources. This will be enhanced with capabilities to extract advanced information such as temporal references and relationships between extracted entities. The implementation will be according to taxonomies defined around the online trafficking domain and will contain adaptions for short texts and misspellings to target the characteristic language found on the Deep Web and Dark Nets. |
| Image and Video analysis | Tools to extract relevant information form visual content – both images and videos – from online and offline sources. The algorithms will automatically detect potentially useful information such as objects, concepts and events. This will be achieved by a combination of: object recognition; concept detection and event detection. |
| Multilingual automated translation | ANITA will incorporate a translation function that will not only translate languages into English but also allow translation from English back into another language. This will then make the incorporation of new (to the system) languages easier and relatively straightforward if the new language is closely aligned with an existing language in the system. |
| Multilingual speech to text | A speech to text tool will be available to analyse audio data for speech and then transcribe that to text. The tool will accept many source languages. |
| Illegal trafficking trend analysis | ANITA will provide a set of tools to analyse the outputs of the multilingual tools outline above and extract trends from that data. The results will be presented with accompanying statistics in order to guide and support investigators. The tools will generate daily, weekly and monthly reports and will also be able to respond rapidly in critical situations. |
| Visual Indexing | Incorporated into the system will be an indexing framework for images and videos. This framework will allow the use of global as well as local descriptors within each visual frame. This will allow the incorporation of contextual knowledge and concept analysis. These tools will create effective descriptors while minimising their complexity. |
| **Knowledge generation and reasoning** | |
| Knowledge modelling for illegal trafficking | ANITA will define a set of ontologies to model all crime aspects including events, suspicious and illegal activities, threats, people, organisations, places, black-markets and illegal shops, products and their relationships. |
| Black markets and illegal shops and products tracking | ANITA will deliver an inference and reasoning tool that will correlate the data generated by the other services. This information will then be linked and reconstructed to provide a narrative storyline of the criminal activity. This will also be forward looking and able to apply the patterns and clusters discovered to other monitored markets and products. |
| Reasoning mechanisms for criminal network reconstruction | A tool to reconstruct criminal networks using advanced reasoning mechanisms and therefore extract hidden relationships from monitored sources. This information will then be used to create hierarchical (vertical) and matrix (horizontal) structures of criminal organisations. This data will then be used to analyse new sources and facilitate the creation of new targeted crawling tasks to ‘fill in the gaps’ of sparse data sets. |
| Knowledge based browsing, search and retrieval | ANITA will incorporate an advanced semantics based service for search and retrieval that will be available at all stages of the retrieval process. This service will incorporate queries expressed in natural language as well as keyword based search techniques. |
| Extraction of evolving knowledge from deep neural network representations | The system will automatically assess and rectify the knowledge it produces, and validate it against existing knowledge bases. The system will have the capacity to learn novel hypotheses from the observed data and to infer new beliefs based on these hypotheses. |
| **Integration of human factor in the analysis loop** | |
| Implicit and explicit user feedback capturing | ANITA can incorporate various methods to collect interaction data with the user. These would include, eye tracking, head movements, facial expressions and mouse movements. These can then be used to interpret the demeanour of the user and use this information to iteratively refine the system’s Big Data analytics services. |
| Adaptive user modelling for cognitive states estimation | ANITA will develop an adaptive model of the human user able to infer the performance of the user whilst interacting with the system. |
| Incorporation of conscious and subconscious user feedback in deep learning representations | A system will be incorporated whereby the voluntary and involuntary actions and reactions of the user are used to reinforce and influence the results and inferences generated by the system |
| Knowledge transfer to new officers | ANITA will incorporate a recommender system that will learn from expert users and transfer those techniques to novice users. This will be achieved through a database of expert users that that relates different actions and content with positive outcomes. There will be the facility of creating different user profiles and the amount of assistance will be targeted to the specific needs of each user. |
| **Novel applications for LEAs and System Integration** | |
| Big Data infrastructure | Compliant with Big Data principles, ANITA will be scalable and flexible; it will be based on a Service Oriented Architecture, including interfaces to facilitate the integration of services outlined above. A Privacy-by-Design approach will be adopted in order to ensure compliance with legal and ethical constraints and regulations at both National and European level. |
| Novel applications for illegal trafficking | ANITA will incorporate applications equipped with user-friendly and interactive widgets able to fulfil LEAs’ requirements and boosting investigative operations allowing officers to verify and validate Big Data analysis results, report events, browse and analyse trafficking activities and potential links among them. These applications will make available useful suggestions to users coming from reasoning algorithms to provide insights for the reconstruction of temporal and causal sequence of events and activities allowing investigators to understand the specific criminal dynamics. Collaborative services to enable and improve active collaboration within and among LEAs will be integrated into the applications. |
| Advanced Big Data Visual Analytics | ANITA will incorporate advanced visual analytics services to represent patterns and trends in online trafficking crimes. Specific visualisation metaphors will be implemented to improve LEAs capabilities in understanding the dynamics of online criminal networks. Interactive widgets for big data visualization will support investigation activities and will help LEAs during the decision-making processes. Furthermore such widgets will provide users with analytics on-demand services. |
| Chain of custody and evidence export | The ANITA system will automatically document the chain of custody and export of evidence. All phases of the information workflow will be analysed in order to define the chain of custody architecture, which will ensure the integrity of information stored into the system, according to the guidelines and best practices provided by LEAs. Mechanisms to ensure data protection, according to the access control policies and taking into account legal, ethical and privacy regulations and constraints will be defined and implemented. Maintenance of integrity will be ensured following specific documented policies. If data integrity is correctly respected, users will be able to export stored information as evidence with all its chain of custody history, to use in court, or to be shared with other LEAs. |

## Functional Requirements in Use Cases

Having specified in detail all the use cases and required functionalities of the ANITA project, the mapping of the required functions in each scenario is presented in Table 2. In the previous sub-section, an initial description of the required ANITA services was presented, regarding the envisaged research methods that are to be used. After thorough discussions with the end-user and technical partners, Table 2 has been filled in such a way that it depicts the real needs of end-users.

Table 2: Functional requirements per Use Case and Scenario

|  | UC-1 (Pharmaceuticals) | | | | | UC-2 (Firearms) | | | | | UC-3 (Funding) | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Service | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| Data sources and stream analysis | | | | | | | | | | | | | | | |
| Data source risk assessment in the Surface Web, Deep Web and Dark Nets | **✓** | **✓** | **✓** | **✓** | **✓** | **✓** | **✓** | N | N | **✓** | **✓** | **✓** | **✓** | **✓** | **✓** |
| Black markets discovery and monitoring | **✓** | **✓** | **✓** | **✓** | **✓** | **✓** | **✓** | N | N | **✓** | **✓** | **✓** | **✓** | **✓** | N |
| Content acquisition from Surface Web and pre-processing | **✓** | **✓** | **✓** | **✓** | **✓** | **✓** | N | N | N | **✓** | **✓** | **✓** | **✓** | **✓** | **✓** |
| Blockchain analysis for illicit activity discovering | **✓** | **✓** | **✓** |  | **✓** | **✓** | N | N | N |  | **✓** | **✓** | **✓** |  | **✓** |
| Construction of source network and filtering | **✓** | **✓** | **✓** | **✓** | **✓** | **✓** | N | N | N | **✓** | **✓** | **✓** | **✓** | **✓** | **✓** |
| Big Data analysis and analytics | | | | | | | | | | | | | | | |
| Multilingual Text analysis | **✓** | N | N | N | N | **✓** | N | N | N | N | **✓** | **✓** | **✓** | N | N |
| Image and Video analysis | **✓** | **✓** | **✓** | N | N | **✓** | **✓** | **✓** | **✓** | N | **✓** | **✓** | **✓** | N | N |
| Multilingual automated translation | **✓** | N | N | N | **✓** | **✓** | N | N | N | N | **✓** | **✓** | **✓** | N | **✓** |
| Multilingual speech to text |  | N | N | N | N |  | N | N | N | N | **✓** | **✓** | **✓** | N | **✓** |
| Illegal trafficking trend analysis | **✓** | N | N | N | N | **✓** | N | **✓** | N | N | N | N | N | N | N |
| Visual Indexing | **✓** | ? | ? | N | N | **✓** | **✓** | **✓** | **✓** | N | ? | ? | ? | N | N |
| Knowledge generation and reasoning | | | | | | | | | | | | | | | |
| Knowledge modelling for illegal trafficking | **✓** | **✓** | **✓** | N | N | **✓** | N | N | N | N | **✓** | **✓** | **✓** | N | N |
| Black markets and illegal shops and products tracking | **✓** | **✓** | **✓** | N | N | **✓** | N | N | N | N | N | N | N | N | N |
| Reasoning mechanisms for criminal network reconstruction | **✓** | **✓** | **✓** | N | N | **✓** | N | N | N | N | **✓** | **✓** | **✓** | N | N |
| Knowledge based browsing, search and retrieval | **✓** | **✓** | **✓** | N | N | **✓** | **✓** | **✓** | **✓** | N | **✓** | **✓** | **✓** | N | N |
| Extraction of evolving knowledge from deep neural network representations | **✓** | ? | ? | N | N | **✓** | N | N | N | N | ? | ? | ? | N | N |
| Integration of human factor in the analysis loop | | | | | | | | | | | | | | | |
| Implicit and explicit user feedback capturing |  | N | N | N | N | **✓** | **✓** | **✓** | **✓** | N | N | N | N | N | N |
| Adaptive user modelling for cognitive states estimation | **✓** | N | N | N | N | **✓** | N | N | N | N | N | N | N | N | N |
| Incorporation of conscious and subconscious user feedback in deep learning representations |  | N | N | N | N | **✓** | N | N | N | N | N | N | N | N | N |
| Knowledge transfer to new officers | **✓** | **✓** | **✓** | N | N | **✓** | **✓** | **✓** | **✓** | N | **✓** | **✓** | **✓** | N | N |
| Novel applications for LEAs and System Integration | | | | | | | | | | | | | | | |
| Big Data infrastructure | **✓** | N | N | **✓** | **✓** | **✓** | N | N | N | **✓** | N | N | N | **✓** | **✓** |
| Novel applications for illegal trafficking | **✓** | N | N | N | N | **✓** | N | N | N | N | N | N | N | N | N |
| Advanced Big Data Visual Analytics | **✓** | N | N | **✓** |  | **✓** | N | N | N | **✓** | N | N | N | **✓** | N |
| Chain of custody and evidence export | **✓** | N | N | **✓** | **✓** | **✓** | N | N | N | **✓** | N | N | N | **✓** | **✓** |

Discussions with the LEAs showed that each of the use cases would benefit from the tools to a greater or lesser extent. The amount of benefit derived from each tool would be entirely dependent on the exact circumstance of the scenario in which it was used and robustness of the tool itself. The key to the table is as follows:

* **✓** – denotes a positive response from the LEA
* N – denotes a negative response from the LEA
* ? – denotes an ambiguous response from the LEA

The following comments were recorded along with the responses.

|  |  |
| --- | --- |
| Service | Comment |
| Data sources and stream analysis | |
| Data source risk assessment in the Surface Web, Deep Web and Dark Nets | Legality (at this point in the investigation) should preferably be a criterion in the selection (Privacy by Design). |
| Big Data analysis and analytics | |
| Image and Video analysis | Extract text from images e.g. we found several images taken with a smartphone of cryptocurrency wallet addresses. It would be useful if ANITA could detect and identify these types of images. |
| Knowledge generation and reasoning | |
| Extraction of evolving knowledge from deep neural network representations | Knowledge offered for addition to the knowledge base should be supported by adequate chain of evidence information. |
| Integration of human factor in the analysis loop | |
| Incorporation of conscious and subconscious user feedback in deep learning representations | Possibly ambiguous reactions of users’ needs to be correctly interpreted. Some visible signals might mean completely different things in different situations or for different persons. This depends of course very much on the quality of the models/algorithms used, the quality and richness of the observations, etc.  Provisions have to be made to avoid training the system with (unconscious) judgements of persons involved based on misconceptions, prejudices (including discrimination, etc.) and false assumptions. Perhaps this could be tackled by interpreting this data, transforming it to explicit knowledge (supported by chain of evidence), which can, when validated, be used to improve the learning process. |
| Novel applications for LEAs and System Integration | |
| Chain of custody and evidence export | Automation, completeness, adequate handling of probability, dynamic level of detail (zoom in out) |

There is broad agreement between the data attributes mentioned across all use cases. Table 3 below shows the amount specific attributes were mentioned across the submitted scenarios.

|  |  |
| --- | --- |
| Attribute | No. of Scenarios |
| Name | 83.33% |
| Nickname | 41.67% |
| Username | 33.33% |
| Phone | 66.67% |
| Email | 58.33% |
| Text | 66.67% |
| Images | 58.33% |
| Video | 41.67% |
| Audio | 16.67% |
| IP Addresses | 58.33% |
| Social media | 25.00% |
| Deep web access | 58.33% |
| Dark Net access | 75.00% |
| Cryptocurrency transactions | 66.67% |
| Similarities | 41.67% |
| GPS | 16.67% |
| Time | 33.33% |
| Organisations | 50.00% |
| Role | 41.67% |
| Associates | 25.00% |
| Affiliates | 16.67% |
| Bank accounts | 25.00% |
| Financial institution data | 8.33% |
| Size of transaction | 16.67% |
| Vehicle data | 16.67% |
| Physical address | 25.00% |

Table 3: Attribute frequency

To assist the formulation of test data sets that do not contain real people and are therefore less sensitive in respect of GDPR, the attributes given above are presented by use case. Table 4, shows the attributes specifically mentioned in use cases 1, 2 and 3.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | | Attribute | Use Case 1 | Use Case 2 | Use Case 3 | | Name | **✓** | **✓** | **✓** | | Nickname | **✓** | **✓** | **✓** | | Username | **✓** | **✓** | **✓** | | Phone | **✓** | **✓** | **✓** | | email | **✓** | **✓** | **✓** | | Text | **✓** | **✓** | **✓** | | Images | **✓** | **✓** | **✓** | | Video |  | **✓** | **✓** | | Audio |  |  | **✓** | | IP Addresses | **✓** | **✓** | **✓** | | Social media | **✓** |  | **✓** | | Deep Web access | **✓** | **✓** | **✓** | | Dark Net access | **✓** | **✓** | **✓** | | Cryptocurrency transactions | **✓** | **✓** | **✓** | | Similarities |  | **✓** | **✓** | | GPS | **✓** |  |  | | Time | **✓** | **✓** | **✓** | | Organisations | **✓** | **✓** | **✓** | | Role | **✓** | **✓** | **✓** | | Associates | **✓** | **✓** |  | | Affiliates | **✓** |  |  | | Bank accounts | **✓** | **✓** | **✓** | | Financial institution data |  |  | **✓** | | Size of transaction |  |  | **✓** | | Vehicle data | **✓** | **✓** |  | | Physical address | **✓** | **✓** |  | |

Table 4: Attributes by use case

# Datasets to be used

## Introduction

Paramount to developing and testing the tools and services encompassed within ANITA is having real or at least realistic data to work with. Significant effort has been allocated to the project to investigate methods and possible issues around data gathering. The ANITA datasets will enable the testing of the systems and will be the basis for validating the overall project, thus assessing the system performance and its accuracy by the completion of the project.

By creating a dataset shared by all partners there will be a consistent platform to exchange data with common sources of information for all involved. Thus tools and systems developed independently will be evaluated against a common dataset ensuring consistency when evaluating the application. All partners and end users will contribute towards the datasets that will be used to develop and test the system.

After discussion with the end users and analysis of the scenarios submitted in response to the use cases it is clear that there is significant overlap between the data sets among the different scenarios under the same use case. Therefore datasets will be created that will be possible to be used within all scenarios under the same use-case for development and testing of the overall ANITA project.

The process of collecting the required data will encompass the following five steps which are also illustrated in +++

* Specify the data requirements by analysis of use cases and scenarios
* Discuss with the project end-users and clarify the procedural aspects of data gathering
* Analysis and discussions about available data that can be gathered from end-users repositories or open sources and data selection
* Conformance with ethical rules and legislation that stem from national laws and directives across different countries in Europe
* Setting and finalizing ANITA datasets

## Use case 1 data collection

Prior to the finalisation of suitable data sets for ANITA, the steps outlined above were followed: data requirements; inquiry and discussions with the project end-users for clarifying all the procedural aspects of data gathering; discussions about available data from end-users repositories or open sources; conformance with ethical rules that are held in different countries and finalizing the project datasets.



Figure 13: Considerations for ANITA datasets

### Data requirements in use case 1

For describing the data requirements for use case 1 a momentous step is to provide a clear definition, whereby all the potential aspects are thoroughly discussed. Therefore, datasets are useful only when there exists a concise definition of the use case that describes all the functionalities required to address the use case. In more detail, the addressed functionalities and services definition that are required for all the use cases are fully described in Section 3. Once properly defined, data need to be collected to enable the development and testing of the system functionality with respect to the specifications of each use case. Initially, there are a number of aspects to be considered on the data requirements and before setting up the datasets:

* What is the most effective way to collect data for UC1
* What questions have to be discussed with the end-users
* What are the specific data required for UC1
* What are the possible legal and ethical ramifications that may occur from gathering the data for UC1
* What is the source of data for UC1, and in which way is it collected
* Time plan for the availability of data and when it is needed
* What elements need to be considered in the process of data collection for UC1

The questions posed above cover a broad scope, so the pertinent topics are expanded below.

### Procedural aspects of data gathering

Thorough discussions were held between the partners and end users in order to form the best possible basis for a robust data set useable not only for this use case but potentially also transferrable across other use cases. The responsible partners providing source information for the UC1 datasets are AoC, LPV, GDCOC and KWPG. The forensic dataset comprises:

* Examples of phishing emails downloaded from the deep web.
* Two Facebook accounts
* Transaction data
* Two phone numbers
* Cryptocurrency wallets
* Visual content associated with social media accounts

### Procedures followed for collecting data

Describe in detail how the various datasets have been formulated in UC1 (list of keywords, sites, sources, etc.).

### Legal and ethical aspects

As noted in section 4.2.2, due to the personal and or sensitive nature of the data involved, and the fact that much of it is still part of active LEA investigations any real data that is made available must be handled with the utmost care. Therefore, not only must all ethical and privacy stipulations be adhered to, but also extreme care must be taken to avoid potential ramifications from its use.

ANITA will ensure that any data made available, will be receipted under all the legal procedures required by the partners and LEAs. ANITA will also ensure all the required approvals will be obtained during the course of the project, especially when concerning data capturing sessions, trials and demonstrations. To this end, ANITA will submit all the necessary documentation, including informed consent and actor role forms.

The ethical rules and legislation that stem from national laws and directives will be fully respected. ANITA will abide by all the different regulations required across the involved countries.

Table 5: Datasets - Use Case 1

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Dataset Name ID or similar | Short Dataset Description | Dataset Owner (Country) | Data formats | Are Personal data (\*) included? | Are sensitive data (\*\*) included? | Are Police & Judicial Data (\*\*\*) included? | Have data subjects given consent and for what use of the data? | Possibility to anonymize dataset |
|  | Phishing emails from deep web | Serbia |  |  |  |  |  |  |
|  | Various social media and cryptocurrency data | Belgium | Various | Yes | Yes |  |  |  |

*\* 'personal data' shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity;*

*\*\* personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, or concerning health or sex life.*

*\*\*\* data related to a) suspected criminal offences, b) criminal sentences, c) criminal and administrative sanctions, penalties, fines or d) any other police measures in place against an individual. If so, for which of these categories?*

## Use case 2 data collection

The procedure that was followed for developing the datasets for UC2 is similar to that already described for UC1. To finalize the setting of UC2 the following steps were involved: data requirements; inquiries and discussions with project end-users for clarifying all the procedural aspects of data gathering; analysis of available data in user’s repositories as well as legal and ethical aspects.

### Data requirements in Use Case 2

As with Use Case 1, a thorough discussion of the requirements was conducted with the end users and the definitions given in Table 1 were verified. The following aspects were also taken into account:

* What are the specific data required to address UC2 functionalities
* What is the most effective way to collect data for UC2
* Who is needed to be involved in the process of data collection in UC2
* What are the legal and ethical implications from gathering the data of UC2
* What questions needs to be discussed with the end-users
* When the data will be available to set ANITA datasets regarding UC2

### Procedural aspects of data gathering

Thorough discussions were held between the partners and end users in order to form the best possible basis for a robust data set useable not only for this use case but potentially also transferrable across other use cases. The responsible partners providing source information for the UC1 datasets are AoC and WPG. The forensic dataset comprises:

* Examples of phishing emails downloaded from the deep web.
* Two Facebook accounts
* Transaction data
* Two phone numbers
* Cryptocurrency wallets
* Visual content associated with social media accounts

### Procedures followed for collecting data

Describe in detail how the various datasets have been formulated in UC1 (list of keywords, sites, sources, etc.).

### Legal and ethical aspects

All the foreseeable implications in law and legislation about data exchange for setting ANITA datasets have been extensively discussed with partners. Therefore, ANITA will respect all the regulations that accompany the data from each country.

Table 6: Datasets - Use Case 2

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Dataset Name ID or similar | Short Dataset Description | Dataset Owner (Country) | Data formats | Are Personal data (\*) included? | Are sensitive data (\*\*) included? | Are Police & Judicial Data (\*\*\*) included? | Have data subjects given consent and for what use of the data? | Possibility to anonymize dataset |
|  |  | Serbia |  |  |  |  |  |  |

*\* 'personal data' shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity;*

*\*\* personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, or concerning health or sex life.*

*\*\*\* data related to a) suspected criminal offences, b) criminal sentences, c) criminal and administrative sanctions, penalties, fines or d) any other police measures in place against an individual. If so, for which of these categories?*

## Use case 3 data collection

The process of creating the datasets for UC3 requires steps that are similar with those involved in UC1 and UC2. Therefore, following a similar procedure, the datasets for UC3 was established so as to enable the evaluation and test of the functionalities that are described in Section 3.1.

### Data requirements in Use Case 3

During the project meetings and through the discussions that took place, the end-users and technical partners clarified all the requirements for the UC3 datasets. The use case and scenarios served to inform the content of the datasets in such a way that they will be useful for technical partners to develop their tools. Among others, the following questions were answered:

* What queries need to be discussed and further clarified with end-users
* What are the specific data that are required for UC3
* How are the data for UC3 collected
* What are the legal and ethical implications from gathering data for UC3
* What partners need to be involved in the collection process

Each question was separately discussed among the partners in the consortium to realize the actual requirements for UC3 datasets before proceeding to its creation.

### Procedural aspects of data gathering

Thorough discussions were held between the partners and end users in order to form the best possible basis for a robust data set useable not only for this use case but potentially also transferrable across other use cases. The responsible partners providing source information for the UC1 datasets are LPV, NPN, DSTL and AoC. The forensic dataset comprises:

* Examples of phishing emails downloaded from the deep web.
* Two Facebook accounts
* Transaction data
* Two phone numbers
* Cryptocurrency wallets
* Visual content associated with social media accounts

### Procedures followed for collecting data

Describe in detail how the various datasets have been formulated in UC1 (list of keywords, sites, sources, etc.).

### Legal and ethical aspects

All the foreseeable implications in law and legislation about data exchange for setting ANITA datasets have been extensively discussed with partners. Therefore, ANITA will respect all the regulations that accompany the data from each country.

Table 7: Datasets - Use Case 3

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Dataset Name ID or similar | Short Dataset Description | Dataset Owner (Country) | Data formats | Are Personal data (\*) included? | Are sensitive data (\*\*) included? | Are Police & Judicial Data (\*\*\*) included? | Have data subjects given consent and for what use of the data? | Possibility to anonymize dataset |
|  | Facebook accounts from closed case | Bulgaria |  | Yes |  |  |  |  |

*\* 'personal data' shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity;*

*\*\* personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, or concerning health or sex life.*

*\*\*\* data related to a) suspected criminal offences, b) criminal sentences, c) criminal and administrative sanctions, penalties, fines or d) any other police measures in place against an individual. If so, for which of these categories?*

## Publicly available data

ANITA will require data from sources other than those held by the end-users, because the project will require an abundance of data for development and evaluation of the system. For this reason, some public repositories have been identified as useful to the technical partners in ANITA. Most of these datasets are publicly available to possible users with only minor restrictions about their usage, if any.

The publicly available datasets that have been identified as having multimedia content that can be exploited for ANITA system development are provided in Table 8.

Table 8: List of publicly available datasets

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Dataset** | **Data typology** | **Dataset size** | **Privacy or license of the data** | **Input formats** |
| **Video re-identification (iLIDS-VID)** | Multivariable, categorical | .300 identities from two (2) disjoint camera views  .Data size: 648M | Open | Mat and .txt files |
| **FALL** |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Joachim has listed various EU Law enforcement datasets – not sure on access restrictions.

# Conclusion

The ANITA project aims at developing automated tools for extracting and analysing heterogeneous forensic data from multiple sources. Throughout the development process a user-centric approach will be used, where end-users will participate in every step, collaborating with technical partners towards the initial definition and further refinement of the tools’ specifications. At the end of this intensive process, investigators are expected to be able to use the system for retrieving a wealth of information related to the trafficking of arms and pharmaceuticals, as well as other sources of terrorism funding. Extracted information will be then subject to a deep analysis using the rest of the ANITA tools through a user-friendly interface.

The user-centric approach that will be followed ensures that the functionalities of the system suit the needs of the end users in their day to day work. The user-centric approach relies on well-defined use cases and this document has extensively emphasized their clear definition so as to demonstrate the usability and added value of the overall system. A high-level description of the use cases is presented through short narrative stories, bridging real world needs with the envisioned system functionalities, as agreed between end-users and technical partners of the consortium. This document reports on work performed in the context of WP4 to reach consensus between developers and users on the practices and representative use cases that will be followed.

The three use cases discussed delineate -through a detailed description- the overall context in which the ANITA system can be used in real life scenarios. Besides use cases definition, a preliminary analysis of the system functionality is also presented along with the services that are required for realizing that functionality. Moreover, the datasets that will be utilized throughout the project for developing and testing purposes are briefly mentioned in Section 4. Data available from the end-users, data from online sources as well as public open datasets and constructed data are used in setting up the overall ANITA database. Discussions with end-users regarding the availability of specific data were performed alongside the use case definition and thus the presented use cases have a direct connection with data available in the ANITA database.

# References

# Annex

This annex contains the collected keywords and data sources for each of the use cases.

These lists are provided in such a way that they should be easy to translate into csv files.

## Use case 1 Pharmaceuticals

### Keywords

4-Methylthioamphetamine, Acid Pig, AK-47, Alkyl nitrates, A-PVP, ayahuasca, ayawaska, blowing weed, Buprenorphine, BZP, cannabis, cannabis in squeeze pouch, cannabis on a table, Cathinones, Cocaine, Codeine, Crazy Dick, Czeszący Grzebień, DMT, Drugs, Ecstasy, efedron, Ethereum, Ethereum mobile wallet jaxx transactions, Ethereum transactions, Gamma-butyrolakton (GBL), giebel, giebla, gieblowanie, Good Shit, Grabarz, gruba lola, Gumi Jagoda, Hardcore, Hebt gij ganja, Heroin, Hex-en, Ibogaine, Ja Rush, Jebaczka, JP2GMD, Ketamine, kicia, klexedron, koniczynka, kot, kulomiot, Kunt gij iet fixe, Kunt gij iet fixen, Kutango Sex 1, Lady Salvia, LSD, magiczna roślinka, M-Cat, Mcpp, MDPK, MDPV, mef, mefacja, mefcia, mefedron, mefka, Mescaline, Methadone, Methamphetamine, Methylphenidate, metkat, Metylokatynon, miau, Mocarz, muchomorki, Mushrooms, Nasa, Nootropics, Opiate, Opioid, Oxycodone, pasibrzuch, pFPP, piperazynka, pipka, Polish PORNSTAR, Psychedelics, puder, pussy stench, puszek, salvia divinorum, salwinoryna A, Sence, Spice, Sztywny Misza, tajfuny, TFMPP, wyrwidąb, yage, yajé, zerwikaptur, zolty ogien.

### data to generate connected to a persona

Bitcoin Transactions, Conversations including keywords or phrases, Dark Web URLs, Ethereum transactions, Facebook profile, Google profile, Images of transactions, Images of wallets, IMEI numbers, Instagram profile, Medium profile, Tumblr profile, Twitter profile, Vkontakte profile, Youtube profile

### Further sources

[www.deepdotweb.com/marketplace-directory/categories/top-markets/](http://www.deepdotweb.com/marketplace-directory/categories/top-markets/), Bitcoin.org, Coinbase.com, <http://6khhxwj7viwe5xjm.onion>, <http://tochka3evlj3sxdv.onion>, <http://valhallaxmn3fydu.onion>, <http://wallstyizjhkrvmj.onion>, <http://www.legalchem.pl/pl/>, <https://forum.dopalamy.com/topic/198-naj%C5%9Bmieszniejsza-nazwa-dopalacza-wg-was/>, <https://hyperreal.info/talk/mdpv-t12355-980.html>, <https://kolekcjoner.nl/sklep>, <https://synthetics.pl/>, <https://theexploaders.com/index.html>, <https://www.gov.uk/government/publications/controlled-drugs-list--2/list-of-most-commonly-encountered-drugs-currently-controlled-under-the-misuse-of-drugs-legislation>, <https://www.liveleak.com/view?i=1eb_1515399044>, <https://www.talktofrank.com/drugs-a-z>, <https://www.youtube.com/watch?v=crlRFpXhlpo>, <https://www.youtube.com/watch?v=GbHuaZn073w>, <https://www.youtube.com/watch?v=Hd5HJLx-BpI>, <https://www.youtube.com/watch?v=nJhAtKtkkpc>, <https://www.youtube.com/watch?v=TTavb16W7Ds>, <https://www.youtube.com/watch?v=UIFAUgYq7Qk>, LiteBit.eu, www.blockchain.com/explorer

## Use case 2 Weapons Trafficking

### Keywords

AK47, Ammunition, Arms Trade, Arms trafficking, bron, colt, dziurkacz, Firearms, flinta, fuzja, gana, gnat, gun, kalasznikow, kaliber, karabin, klamka, kolt, kopyto, maszynka, parabelka, parabellum, pestki, pistolet, pompka, rura, , Smuggling, sztucer, Undermine democracy, Weapons

### data to generate connected to a persona

Gun serial numbers, Gun model, Gun origin, Gun characteristics, New or used guns

### Further sources

<http://tuu66yxvrnn3of7l.onion>, <http://2kka4f23pcxgqkpv.onion>, <http://armsmhmd4c3hb5xu.onion>, <http://blackpowder.pl/>, <https://www.youtube.com/watch?v=ziK-Xxf3UL0>, <https://www.youtube.com/watch?v=qV8ZJPvU-_k>, <http://www.armybazar.eu/pl/>, https://zafanatrzeciastrona.pl/post.darknet-kanal-sprzedazy/

## Use case 3 Terrorism Funding

### Keywords

Account, Al Qaeda. Al Shabaab, ammonium nitrate, AQAP, AQMI, Attack, blades, Bomb, Bombs, chemical, collection, Daesh, Daesh symbols, dollars, eco terrorism, fund, Home grown, how to make a bomb, IED, IED, Isis, ISIS, Isis flag, Isis symbols, JIHAD, knife, knives, martyrs, money, mosque, payment, plot, Radical, Sharia4Belgium, suicide, Syria, Taliban, target, terror, terrorist, weapons, weapons, Wedding

### data to generate connected to a persona

Bank accounts, Blackberry Messenger, chaton, credit cards, Discord, Facebook Messenger, Facebook profile, FACETIME, Google Allo, hike, Instagram, Kakao, Kik, Line, Signal, Silence, SilentPhone, Skype, Slack, Snapchat, Telegram, Textra, Threema, TWITTER, Viber, Voxer, WeChat, WhatsApp, Wickr Me, Wire

### Further sources

<http://terroristdomains.com/>, Facebook, Google, <http://terroristdomains.com/>, Instagram, TRIP ADVISOR, Twitter