Deliverable D3.1
Date of document – December/2017 (M15)

D3: Report on OC and terrorism in Cyberspace

WP 3, T 3.1, T 3.2, T 3.3, T 3.4, T 3.5, T 3.6

Authors: Stefan Rilling (Fraunhofer), Jingquan Xie (Fraunhofer); Michael Osborne (IBM), Michael Baentsch (IBM), Axel Tanner (IBM); Francesco Calderoni (Transcrime), Martina Elena Marchesi (Transcrime), Valentina Calcagno (Transcrime); Michael Wolfowicz (HUJI), Simon Perry (HUJI)

Modelling the Processes leading to Organised crime and TerrOrist Networks
FCT-16-2015
### Technical References

<table>
<thead>
<tr>
<th>Project Acronym</th>
<th>PROTON</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Title</td>
<td>Modelling the PROCesses leading to Organised crime and TERRORist Networks</td>
</tr>
</tbody>
</table>
| Project Coordinator | Ernesto Savona  
Università Cattolica del Sacro Cuore  
ernesto.savona@unicatt.it |

| Project Duration | October 2016 – September 2019 (36 months) |

| Deliverable No. | D 3.1 |
| Dissemination level | PU |
| Work Package       | WP 3 - Organized Crime and Terrorism in Cyberspace |
| Task               | T 3.1, T 3.2, T 3.3, T 3.4, T 3.5, T 3.6 |
| Lead beneficiary   | 3 (Fraunhofer) |
| Contributing beneficiary(ies) | 3 (Fraunhofer), 4 (IBM) |
| Due date of deliverable | 31. December 2017 |
| Actual submission date | 22. December 2017 |

1. PU = Public  
   PP = Restricted to other programme participants (including the Commission Services)  
   RE = Restricted to a group specified by the consortium (including the Commission Services)  
   CO = Confidential, only for members of the consortium (including the Commission Services)

### Document history

<table>
<thead>
<tr>
<th>V</th>
<th>Date</th>
<th>Beneficiary</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30.10.2017</td>
<td>Fraunhofer</td>
<td>Stefan Rilling, Jingquan Xie</td>
</tr>
<tr>
<td>2</td>
<td>18.11.2017</td>
<td>Fraunhofer</td>
<td>Stefan Rilling, Jingquan Xie</td>
</tr>
<tr>
<td>3</td>
<td>22.11.2017</td>
<td>Fraunhofer, IBM, Transcrime</td>
<td>Stefan Rilling, Jingquan Xie, Michael Osborne, Michael Baentsch, Axel Tanner, Valentina Calcagno</td>
</tr>
<tr>
<td>4</td>
<td>05.12.2017</td>
<td>Fraunhofer, IBM, Transcrime</td>
<td>Stefan Rilling, Jingquan Xie, Michael Osborne, Michael Baentsch, Axel Tanner,</td>
</tr>
</tbody>
</table>
Summary

This deliverable presents the findings and reports for all tasks within WP3. Each of the six tasks conducted within this work package is presented within a separate section. The work done in T3.1, the systematic review of cyber-related OC and terrorist activities, and technological means for their study forms the theoretical foundation of WP3’s overall goals. The qualitative-quantitative study conducted within T3.4 gives an insight in the social media activities of lone-wolf terrorists and shows possible indicators of the move from radicalization of belief to radicalization of action. The technical work conducted in WP3 is represented by the tasks T3.2, T3.3 and T3.5. A data crawling and analysis platform was developed and used within the tasks. The major outcome of this system is the generation of personality profiles from textual data collected from various online resources.
# Table of content

0 **SUMMARY** 3

1 **INTRODUCTION** 12

2 **TASK 3.1: SYSTEMATIC REVIEW OF CYBER-RELATED OC AND TERRORIST ACTIVITIES, AND TECHNOLOGICAL MEANS FOR THEIR STUDY** 13

2.1 **BACKGROUND** 13
2.2 **ONLINE ACTIVITIES** 14
2.3 **DEFINING TERRORISM AND ORGANISED CRIME** 18
2.3.1 **TERRORISM AND RADICALISATION** 18
2.3.2 **ORGANISED CRIME** 20
2.4 **RATIONALE OF THE SYSTEMATIC REVIEW** 24
2.4.1 **OBJECTIVES AND RESEARCH QUESTIONS** 25
2.5 **METHODOLOGY** 25
2.6 **FINDINGS OF THE SYSTEMATIC REVIEW** 29
2.6.1 **ONLINE ACTIVITIES OF TERRORIST NETWORKS** 29
2.6.2 **ONLINE ACTIVITIES OF ORGANISED CRIME** 44
2.6.3 **METHODOLOGIES AND SOFTWARE** 59
2.7 **DISCUSSION AND CONCLUSIONS** 60
2.7.1 **ONLINE ACTIVITIES OF TERRORIST AND ORGANISED CRIME NETWORKS** 60
2.7.2 **POPULAR METHODOLOGIES IN RESEARCHING ONLINE ACTIVITIES** 62
2.7.3 **SIMILARITIES AND DIFFERENCES BETWEEN TERRORIST AND CRIMINAL NETWORKS** 63
2.7.4 **POLICY IMPLICATIONS** 64

3 **TASK 3.2: INNOVATIVE STUDY - FINDING THE DARK WEB SIGNPOSTS** 66

3.1 **THE EXPERIMENTAL PLATFORM** 66
3.1.1 **SYSTEM ARCHITECTURE** 66
3.1.2 **BACK-END DESIGN** 67
3.1.3 **DATA COLLECTORS** 71
3.1.4 **IN-DATABASE ANALYSIS** 76
3.1.5 **DEPLOYMENT** 77
3.1.6 **THIRD-PARTY LIBRARIES** 77
3.2 **TERMS AND DEFINITIONS** 78
3.3 **THE DARK WEB** 79
3.3.1 **SURFACE WEB** 79
3.3.2 **DEEP WEB** 80
3.3.3 **DARK WEB** 80
3.3.4 **RELEVANT TECHNOLOGIES** 80
3.4 **DARK WEB SIGNPOST IDENTIFICATION** 82
3.5 **PROJECT RESULTS** 83
3.5.1 **TWITTER** 84
3.5.2 **REDDIT** 84
3.6 **PERSONALITY AS A SIGNPOST** 85
3.7 **DATA INPUT TO PERSONALITY INSIGHTS ANALYTICS** 87

3.8 **BIG FIVE** 88
3.8.1 **AGREEABLENESS** 89
PEOPLE WHO SCORE HIGH...

ALTRUISM / ALTRUISTIC

FIND THAT HELPING OTHERS IS GENUINELY REWARDING, THAT DOING THINGS FOR OTHERS IS A FORM OF SELF-FULFILLMENT RATHER THAN SELF-SACRIFICE.

COOPERATION / ACCOMMODATING / COMPLIANCE

DISLIKE CONFRONTATION. THEY ARE PERFECTLY WILLING TO COMPROMISE OR TO DENY THEIR OWN NEEDS TO GET ALONG WITH OTHERS.

MODESTY / MODEST

ARE UNASSUMING, RATHER SELF-EFFACING, AND HUMBLE. HOWEVER, THEY DO NOT NECESSARILY LACK SELF-CONFIDENCE OR SELF-ESTEEM.

MORALITY / UNCOMPROMISING / Sincerity

SEE NO NEED FOR PRETENSE OR MANIPULATION WHEN DEALING WITH OTHERS AND ARE THEREFORE CANDID, FRANK, AND GENUINE.

SYMPATHY / EMPATHETIC

ARE TENDER-HEARTED AND COMPASSIONATE.

TRUST / TRUSTING OF OTHERS

ASSUME THAT MOST PEOPLE ARE FUNDAMENTALLY FAIR, HONEST, AND HAVE GOOD INTENTIONS. THEY TAKE PEOPLE AT FACE VALUE AND ARE WILLING TO FORGIVE AND FORGET.
4.1 Measuring the social media presence of gangs in Europe 111
4.2 Analysis of gang-related online contents¶ Fehler! Textmarke nicht definiert.
   4.2.1 Approach

5 Task 3.4: Innovative study - Radicalisation in cyberspace and radical social media networks 119
5.1 Introduction 119
5.2 Theoretical framework 120
5.3 Data and methodology 123
5.4 Results 124
5.4.1 Differential associations 124
5.4.2 Frequency 125
5.4.3 Radicalness 126
5.4.4 Post type 127
5.4.5 Differential reinforcement 129
5.4.6 Patterns of activity 130
5.5 Discussion 134
5.6 Summary 135

6 Task 3.5: Innovative study - Terrorist-related contents in cyberspace 137
6.1 Identification and collection of terrorist-related online contents 138
   6.1.1 Radicalization signals in social media 139
   6.1.2 Analysis of Twitter data 139
   6.1.3 Analysis of YouTube data 142
   6.1.4 Analysis of texts from the Rumiyah magazine 145
6.2 Assessing mechanisms of online propaganda 146
   6.2.1 Personality traits clustering for social media data 147
   6.2.2 Personality profiles from propaganda text 153
6.3 Summary 154

7 Task 3.6: Policy makers’ contribution 155
7.1 Europol 155
7.2 Unodc 156
7.3 Eucpn 157
7.4 Wodc 157
7.5 Actions resulting from the policy maker’s contributions 158

8 Agents and interactions 159
8.1 Agent persona library 159
   8.1.1 Agent personas and data privacy 160
   8.1.2 Agent persona template 161
   8.1.3 Collecting data for personas 162
8.2 Benefits to this approach 162
8.3 Adapting the TARA methodology 163
8.4 TARA model 165
   8.4.1 TARA profiles 165
   8.4.2 TARA attributes 166
8.5 Adapting the TARA methodology 167
8.6 Agent persona attributes 169
   8.6.1 TARA related 169
   8.6.2 Linguistic analysis 170
ANNEX D – STUDIES ON METHODOLOGIES AND SOFTWARE

217
List of Tables

Table 1 – Modes for exploitation of the Internet .......................................................... 15
Table 2 – Types of criminal networks covered by this systematic review based on McGuire’s framework (2012) .............................................................................. 23
Table 3 – List of databases and search strategies ....................................................... 27
Table 4 – Aims of propaganda messages .................................................................. 36
Table 1 Summary of the PROCeeD RESTful web services, which are categorised by the associated HTTP operations like GET, PUT, etc. ................................................................. 85
Table 2 Summary of the operations specified in the PROCeeD system to ensure correct interplay with collectors ................................................................. 85
Table 1 Distribution of unique/total .onion links in various subreddits (reddit communities) ................................................................. 85
Table 2 Distribution of unique/total .i2p links in various subreddits (reddit communities) ................................................................. 85
Table 3 Personality Trait - Agreeableness – Facets ................................................. 89
Table 4 Personality Trait - Agreeableness - Range of characteristics .................... 90
Table 5 Personality Trait - Agreeableness - Primary and secondary dimensions – part 1 .............................................................................................................. 90
Table 6 Personality Trait - Agreeableness - Primary and secondary dimensions – part 2 .............................................................................................................. 90
Table 7 Personality Trait - Conscientiousness – Facets .......................................... 91
Table 8 Personality Trait - Conscientiousness - Range of characteristics ............ 92
Table 9 Personality Trait - Conscientiousness - Primary and secondary dimensions – part 1 .............................................................................................................. 92
Table 10 Personality Trait - Conscientiousness - Range of characteristics .......... 92
Table 11 Personality Trait - Conscientiousness - Primary and secondary dimensions – part 2 .............................................................................................................. 92
Table 12 Personality Trait - Extraversion – Facets .................................................. 93
Table 13 Personality Trait - Extraversion - Range of characteristics ..................... 94
Table 14 Personality Trait - Extraversion - Primary and secondary dimensions – part 1 .............................................................................................................. 94
Table 15 Personality Trait - Extraversion - Primary and secondary dimensions – part 2 .............................................................................................................. 94
Table 16 Personality Trait - Emotional range – Facets .......................................... 95
Table 17 Personality Trait - Emotional range - Range of characteristics ............... 96
Table 18 Personality Trait - Emotional range - Primary and secondary dimensions – part 1 .............................................................................................................. 96
Table 19 Personality Trait - Emotional range - Primary and secondary dimensions – part 2 .............................................................................................................. 96
Table 20 Personality Trait - Openness – Facets ...................................................... 97
Table 21 Personality Trait - Openness - Range of characteristics ......................... 98
Table 22 Personality Trait - Openness - Primary and secondary dimensions – part 1 98
Table 23 Openness - Primary and secondary dimensions – part 2 ........................ 98
Table 24 Personality Trait - Needs ............................................................................ 99
Table 25 Personality Trait - Values .......................................................................... 99
Table 5: List of keywords used to identify radical content ...................................... 139
Table 26 TARA Attributes ....................................................................................... 167
Table 27 Persona Attributes – TARA related ......................................................... 169
Table 30 Persona Attributes – BIG 5 Personality Traits ....................................... 170
Table 31 Persona Attributes: Personal Needs Model ............................................. 170
Table 32 Persona Attributes: Personal Values Model ............................................ 171
Table 5 – List of databases and search strategies .................................................. 192
Table 6 – Search categories and related search terms ............................................. 193
Table 7 – Databases and related queries ................................................................. 195
Table 8 – Contributing experts .............................................................................. 198
Table 9 – Studies included TN & OC .................................................................... 202
Table 10 – List of studies included on terrorist networks ....................................... 211
Table 11 – List of studies included on criminal networks ....................................... 215

List of Figures

Figure 1: The Structure of WP3. The tasks T3.2, T3.3 and T3.5 comprise the technical work conducted with the implemented data analysis system. ................................. 12
Figure 2 – Query structure .................................................................................... 26
Figure 3 – Flow chart of the selection of relevant literature on terrorism ............... 28
Figure 4 – Flow chart of the selection of relevant literature on organised crime .... 29
Figure 5 – Types of terrorist networks .................................................................. 30
Figure 6 – Activities of terrorist networks per methodology .................................. 30
Figure 7 – Types of criminal networks .................................................................. 44
Figure 8 – Activities of criminal networks in cyberspace ...................................... 45
Figure 9: The architecture of the PROCeeD system. It is a web application with two main blocks: the service server with the database and the collectors that crawl various kinds of information on the Web. .................................................................. 67
Figure 10: The database schema of the PROCeeD core table structure.................. 69
Figure 11: The database schema of the table structure for the management of the crawled YouTube data. ...................................................................................... 71
Figure 12: The UML class diagram of the inheritance relationships of PROCeeD collectors ............................................................ 74
Figure 13: Illustration of the database schema for the Twitter collector. .............. 75
Figure 14: Schematic overview of the data collection process for YouTube video descriptions and comments. The list of keywords acts as the input of the automated extraction system. .................................................................................... 75
Figure 15: A part of the defined database views to analyse the YouTube video and the comments crawled by the YouTube collector. .................................................... 77
Figure 16: Agent Persona Creation ....................................................................... 102
Figure 17: Personality Insight from single person.................................................. 107
Figure 18: Aggregate Personality Insight form multiple persons ............................ 108
Figure 19: Sunburst visualization of personality insights ....................................... 109
Figure 20: The web presence of the Hells Angels MC showing the various chapters over Europe (top), and within the particular countries (bottom) .................... 112
Figure 20 Timeline for specific accounts mentioned in the daesh stream - for details see text ........................................................................................................... 141
Figure 21 Enlarged view of the previous figure. ...................................................... 142
Figure 22: Example of a YouTube search. The search term is entered into the text field on the top of the web page; a list of videos with according links is presented. .... 143
Figure 23: Proportion of video descriptions with matching radicalization indicating keywords (left), and video comments (right) ...................................................... 144
Figure 24: Distribution of word counts in video descriptions (top left) and video comments (top right). Median and distribution of the word count in video descriptions (bottom left) and comments (bottom right). The scale on the Y-axis is logarithmic. 145
Figure 25: Screenshot of the Rumiyah magazine (English version), issue 10, p. 38.

Figure 26: Assessment of the main personality characteristics of the actors involved in social media presences. The personality characteristics are derived from textual data extracted from social media and web pages. The personality traits associated with a specific text item are analyzed for clusters to get an impression of the main personality structures of the content originators.

Figure 27: Personality cluster sizes for PCA transformed data (left) and the raw data (right) for video comments and descriptions.

Figure 28: Silhouette coefficient plots for the personality clusters of the comment data.

Figure 29: Silhouette coefficient plots for the personality clusters of the video description data.

Figure 30: Personality clusters derived from the WPI analysis of the video description texts. Cluster 0 is shown on the left, cluster 1 is shown on the right.

Figure 31: Personality clusters derived from the WPI analysis of the video comments texts. From top left to bottom right: Cluster 0 – Cluster 7.

Figure 32: Sunburst visualizations for the results of the WPI analysis from the different texts from the Rumiyah magazine. Top left: Interview from issue 5. Top right: Interview from issue 10. Bottom left: Interview from issue 12. Bottom right: Narrative “A Mujahid Memories”.

Figure 8: Agent Persona building blogs.

Figure 26: TARA Scheme overview.

Figure 27: TARA Scheme overview.
1 Introduction

The main objective of the work package “Organized Crime & Terrorism in Cyberspace” is an analysis of how OCTN influence and exploit the internet to reach their goals. In the process, the usage of social networks is of particular interest. One particular outcome of WP3 will be the input for an agent based model based on the personality structure of the agents involved (the so-called agent personas), which describes the links between cyberspace and the factors leading to OCTNs. Besides this, a software, which allows the collection and analysis of various social media and internet data sources, will be implemented and used within WP3.

Work package 3 is structured in 6 different tasks, among these are three tasks involving mainly technical work (T3.2, T3.3, T3.5). The tasks T3.1 and T3.4 comprise theoretical work and data analysis. Task T3.6 will connect the policy makers to the outcome of the other tasks.

Figure 1: The Structure of WP3. The tasks T3.2, T3.3 and T3.5 comprise the technical work conducted with the implemented data analysis system.
2 Task 3.1: Systematic review of cyber-related OC and terrorist activities, and technological means for their study

This systematic review aims at systematically reviewing the existing knowledge on the exploitation of cyberspace and Information and Communication Technologies (ICTs) by organised crime and terrorist networks (OCTNs).

2.1 Background

The exploitation of cyberspace by OCTNs for communication and illicit activities has gained growing attention in the last years. This is embedded in a wider process of globalisation that has been facilitated by advancements in ICTs, especially since the 1980s (Berend, 2016; Steger, 2014; Sassen, 2007). Globalisation had an impact on legal economies, and illegal markets benefited from the same facilitators (Friedrichs & Rothe, 2014). ICTs introduced some new criminal opportunities and new typologies of activities requiring effective counter strategies (Pyrooz, Decker, & Moule, 2015; Lavorgna, 2015a). The exploitation of cyberspace by organised crime groups (OCGs) and terrorist networks (TNs) takes different forms. The use of the Deep Web to engage in illegal activities represents the most visible example (Everett, 2015; Holt, Smirnova, Chua, & Copes, 2015; UNODC, 2014; Europol, 2014b). However, other types of online activities have been documented. This includes the facilitation of internal and intergroup communication, and the use of the Internet for self-portrayal and image cultivation. Street gangs, for example, use social networks to fulfil their “symbolic needs” (Pyrooz, Decker, & Richard K. Jr, 2015). The dissemination of terrorist-related information is known for its role of “force multiplier” and “radicalisation accelerant” in extremists’ radicalisation (Neumann & Stevens, 2009; Royal Canadian Mounted Police, 2009; Sageman, 2004). Occasionally, non-systematic online recruitment by OCTNs have been reported both in the case of OCGs (Lavorgna, 2015a; Décary-Hétu & Morselli, 2011; Kleemans & de Poot, 2008; King, Walpole, & Lamon, 2007) and TNs (Basra, Neumann, & Brunner, 2016; Vidino & Hughes, 2015). Both OCGs and TNs employ various anti-forensics techniques such as...
steganography, encryption and proxies, to counter law enforcement actions (Dela, 2016; Veerasamy & Grobler, 2011; Labi, 2006; Lau, 2003). With regard to TNs, some authors suggest that the unlimited access to information on the web has facilitated distant training (e.g. building explosives) and allowed terrorist to obtain information needed for planning attacks (Weimann, 2004). The use of social networks by TNs gained particular attention, with some recent studies analysing their propaganda strategies (e.g. ISIL use of Twitter: Berger & Morgna, 2015; Magdy, Darwish, & Weber, 2015).

To clarify the rationale of the current systematic review and define its scope, the current chapter elaborates on the definitions of online activities, terrorism and OC. This preliminary literature review framed the objectives of the review and facilitated the definition of keywords and queries for the systematic search of literature.

## 2.2 Online activities

The ICTs have several characteristics that create an opportunity structure for exploitation of cyberspace by OCTNs. Nowadays access to Internet is easy, affordable and providing global reach, enabling OCTNs to reach large audiences and the most vulnerable targets (Holt, Bossler, & Seigfried-Spellar, 2015; Koops, 2011; Weimann, 2004). It allows anonymity and distant interaction with affiliates and victims, removing potential physical and social barriers (Sandywell, 2010). Furthermore, ICTs facilitate the manipulability of data and software with minimal costs, the use of different type of media (e.g. text, video and audio), and the shaping of coverage in traditional mass media (Koops, 2011; Weimann, 2004). The advent of ICTs turned into a valuable asset both data itself and the ability to analyse it, and allows rapid innovation cycles of criminal strategies and tools (Koops, 2011). In addition, authors agree that the Internet poses structural constrains to effective guardianship, and lacks any form of government control (Koops, 2011; Sandywell, 2010; Weimann, 2004).

OCTNs exploit to a varying degree these characteristics of the Internet, depending on the goals they pursue. Therefore, the various uses of the Internet are listed separately for OCGs and TNs, but similarities are pointed out. With regard to the terminology, the systematic review uses “online” or “cyberspace activities” to denote the various possible uses of ICTs. These terms are broader than concepts like “cybercrime”, “computer crime”, “technologically-enabled crime” or “high-tech crime”. While some authors have adopted a wide understanding of the term “cybercrime” as existence of illegal behaviour somehow related to a computer (Yar, 2005), online activities also encompass behaviour, which is not illegal per se, but contributes to the OCTNs’ goals.

Criminals and terrorists could exploit ICTs in three main ways (Holt et al., 2015):

1) as communication tool and means to form online subcultures;

2) as a mechanism to commit a crime;

3) as incidental tool facilitating the offence.
Some authors consider that the activities carried out by OCTNs in cyberspace are not inherently different from those they carry out in settings unrelated to cyberspace (McCuster, 2006; Grabosky, 2001). The innovation rather concerns the technological means that facilitate the perpetration of crimes. Much like technology in general, these means are constantly changing, innovating and adapting (Grabosky, 2001).

Within the three modes of exploitation of ICTs, the literature suggests that there are several types of activities of OCGs and TNs (see Table 1) (Europol, 2017; Nagy & Mezei, 2016; Veerasamy & Grobler, 2011; Goodman, Kirk, & Kirk, 2007; Weimann, 2004).

Table 1 – Modes for exploitation of the Internet

<table>
<thead>
<tr>
<th>TNs Type of activity</th>
<th>OC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ICT as a communication tool and means to form online subcultures</strong></td>
<td></td>
</tr>
<tr>
<td>Internal communication within group members</td>
<td>Internal communication within group members</td>
</tr>
<tr>
<td>Recruitment and mobilization</td>
<td>Recruitment and mobilization (gangs and DTOs)</td>
</tr>
<tr>
<td>Networking</td>
<td>Networking</td>
</tr>
<tr>
<td>Propaganda and psychological warfare</td>
<td>Dissemination of criminal culture (gangs and DTOs)</td>
</tr>
<tr>
<td>Training</td>
<td></td>
</tr>
<tr>
<td><strong>ICT as a mechanism to commit a crime</strong></td>
<td></td>
</tr>
<tr>
<td>Perpetration of computer-focused crimes (e.g., use of malware to commit identity fraud; cyberattacks against critical infrastructures)</td>
<td>Perpetration of computer-focused crimes (e.g., use of malware to commit identity/payment fraud; cyberattacks against financial institutions/critical infrastructures)</td>
</tr>
<tr>
<td><strong>ICT as incidental facilitator of criminal offences</strong></td>
<td></td>
</tr>
<tr>
<td>Money laundering and financing</td>
<td>Money laundering and financing</td>
</tr>
<tr>
<td>Information gathering and planning of physical attacks</td>
<td></td>
</tr>
</tbody>
</table>

Internal communication and networking

With regard to TNs, ICTs could not only connect members of the same terrorist cell, but also facilitate intergroup communication and organisation (Weimann, 2004). This could happen through targeted websites, forums and social networking sites (Veerasamy & Grobler, 2011). According to Weimann (2004) affiliations of partially independent cells form terrorist networks. These cells would be increasingly likely to be organising through ICTs-enabled communication in a horizontal rather than vertical matter. The reduced transmission time, low costs and encryption possibilities could all work to the benefit of the internal communication and coordination of organised crime and terrorist networks (Nagy & Mezei, 2016; Weimann, 2004). Veerasamy and Grobler (2011) and Halopeau (2014) highlight the role of anti-forensic tools and VPN services in particular in the communication among terrorist members.
**Propaganda and psychological warfare**
As terrorism is also a form of psychological warfare, terror organisations have tried to use the Internet to this purpose (Keene, 2011; Woods, 2007; Weimann, 2004). They spread disinformation, threats and horrifying images of their actions, and to that end they produce various multimedia materials and use different media channels (Halopeau, 2014; Weimann, 2004). Woods (2007) considers TNs to show themselves more capable than they are in reality. Social media allowed terrorists to secure their publicity by avoiding the thresholds of traditional media outlets. The control over their own propaganda content allows terrorist groups to manipulate their image and that of the enemy, and shape how they are perceived by audiences (Keene, 2011).  

**Recruitment and mobilisation**
The terrorists’ propaganda serves also as an initial step for the recruitment of supporters to play an active role in terrorist activities (Woods, 2007). Weimann (2004) suggests that terrorist groups track information about the visitors of their websites and contact the ones that seem most interested in their cause. Internet interactive nature would allow recruiters to actively reach supporters in chatrooms and cybercafes, and use algorithms for targeted propaganda to different audiences (Keene, 2011; Woods, 2007). Apart from promoting community environment, terrorist groups would offer monetary incentives for low income supporters to join their cause (Woods, 2007). According to Weimann (2004) there are also cases where supporters advertise themselves and contact the terrorist organisation themselves through the Internet.  
With regard to OC, social networking sites allow gang members to interact with a wider circle of people –with variable levels of awareness and intent– who would otherwise not be exposed to their lifestyle (Décary-Hétu & Morselli, 2011). There is an increasing trend among gang organisations to post videos of the advantages of being a gang member (Décary-Hétu & Morselli, 2011).  

**Training**
In addition to the propaganda and recruitment narratives, terrorists’ online platforms provide instruction manuals and videos on technical and tactical topics, such as explosive making, small arms handling, hostage taking, operational security (Stenersen, 2008; Weimann, 2006). The Internet’s interactive features provide opportunity for following online courses and communicating with trainers (Weimann, 2006).  

**Perpetration of computer-focused crime**
The term “cyberterrorism” was originally coined by Collin (1997) as convergence of “cybernetics” and “terrorism”. Scholars suggest that cyberterrorism could inflict actual damage on society’s critical infrastructure, such as telecommunications, electricity networks, water supply and financial institutions (Denning, 2000; Collin, 1997). What differentiates cyberterrorism from hackers’ cyberattacks is the political motivation justifying the need to inflict human and economic losses (Denning, 2000). Denning (2000) argues that two factors could increase the probability of cyberterrorism attacks: the vulnerability of targets and the capability of terrorists to acquire technological skills. Cyberterrorist activities could also be part of the strategy of physical
terrorist attacks, as presumably happened during the London and Madrid bombings (Woods, 2007). Regarding OC, literature has focused on the question to what extent hackers have created more complex organisational networks and whether offline OCGs have started committing cybercrimes (Leukfeldt, Lavorgna, & Kleemans, 2016; McGuire, 2012; Broadhurst, Grabosky, Alazab, & Chon, 2014).

**Perpetration of traditional organised crime activities**

The opportunities provided by the anonymity of the ToR network and the development of cryptocurrencies led to the development of dark net markets, where drugs, counterfeit pharmaceutical products and arms are traded (Europol, 2017). As in the physical world these activities are mainly within the realm of OCGs, it could be assumed that OC has moved the trade to the cyberspace (Lavorgna, 2015a; K. Choo & Smith, 2008).

**Money laundering and financing**

Terrorist networks use the Internet to raise financial resources (Veerasamy & Grobler, 2011; Woods, 2007; Weimann, 2004). The funding strategies include asking supporters for donations, mimicking humanitarian charities, online auctioneering and criminal activities such as credit card fraud and drugs trafficking (Veerasamy & Grobler, 2011; Woods, 2007; Weimann, 2004).

Williams (2001) recognised the increased likelihood of the use of Internet for money laundering of OC proceeds. According to Lilley (2002) the most common scheme for criminals to launder proceeds is through the numerous available virtual casinos or through claiming that they operate a gambling website and thus establishing a banking relationship. An online auction scheme could involve either the selling of goods originating from crime commission or no movement of the goods after transferring the money (Nagy & Mezei, 2016). The advent of virtual currencies provided further opportunities for money laundering due to their the greater level of anonymity, their global reach and the unclear responsibility for AML compliance and supervision (FATF, 2012).

**Information gathering and planning of attacks**

Apart from training information, the Internet could also offer information needed for the planning of physical attacks – infrastructure details, maps, satellite photos, email distribution lists (Veerasamy & Grobler, 2011; Woods, 2007; Weimann, 2004). Investigative material on terrorist groups and their training manuals confirm their reliance on legally and cost-free publicly available information in planning attacks (Woods, 2007). In addition, there are cases of illegally obtaining confidential information through hacking intrusions in computer systems (Woods, 2007). During the execution phase of the attack, ICTs serve as an operational command and control tool (Veerasamy & Grobler, 2011).

The abovementioned online activities are closely interrelated, often happen simultaneously and could be difficult to differentiate. However, the description and categorisation of the various cyber activities of OCTNs is not the aim of the current systematic review.

The organised character of execution of these activities will remain central to the discourse of the systematic review. Specific cybercrimes perpetrated through the use of crimeware will be considered as long as they are carried out
by OCTNs. The same applies for traditional offline OC activities, which could be facilitated through the use of Internet, as drugs and human trafficking. Only studies testing the organised nature of online perpetration of these crimes will be included in the final selection. The systematic review does not cover any activities perpetrated by actors not falling within the concepts of OCGs or TNs, as described in the following section.

2.3 Defining terrorism and organised crime

The definition of OCTNs will assist the scope of the study and the development of the exhaustive list of keywords for the systematic search queries. Regarding terrorism, this systematic review relies on widely accepted definitions of terrorism and radicalisation. With regard to OC, the systematic review opts for the definitions of traditional OCGs, but also takes into account recent theoretical understandings of the emergence of cybercriminal networks. This approach guarantees a thorough review providing a wide picture of OCTNs exploiting ICTs.

2.3.1 TERRORISM AND RADICALISATION

There is no universally agreed definition of terrorism. The most widely shared idea is that “[t]errorism is, first and foremost, a method” (Matusitz, 2013, p. 4). The concept of terrorism is socially constructed (Barnett & Reynolds, 2009) and highly variable depending on the political and historical contexts, and the dominant ideologies (Crenshaw, 2010, p. 598). The amount of studies on terrorism significantly increased since the 1980s, leading to a fragmentation of technical and operative definitions (Jackson, 2016; Matusitz, 2013; Crenshaw, 2010; Jongman & Schmid, 2005; Hoffman, 1998; Laqueur, 1977). This variety of definitions has been recognised as one of the key obstacles to the advancement of this field of study (Crenshaw, 2010).

Schmid and Jongman (2005) collected and systematically analysed more than one hundred definitions of terrorism, but were not able to synthesise a satisfying definition. However, they reported the concepts most commonly associated to terrorism: violence (83.5%); political goals (65%); causing fear and terror (51%); indiscriminate targeting (21%); victimisation of civilians, non-combatants, neutrals and outsiders (17.5%).

Hoffman (1998) argued in favour of a functional change of perspective, from seeking an all-inclusive ‘positive’ definition of terrorism to establishing ‘negative’ criteria that distinguish terrorism from other forms of violence. Following this approach, Ganor (2010) developed three pillars of characteristics with corresponding exclusion criteria: (1) the use of violence, excluding non-violent protests (e.g. strikes, peaceful demonstrations, tax revolts); (2) the political nature of the goal, allowing to mark a distinction with general delinquency; (3) the civilian targets, excluding other forms of political violence (e.g. guerrilla warfare, civil insurrection, etc.).

Faced with the problem of defining terrorism, Project PROTON developed an operational definition of terrorism. This definition is (1) general and inclusive,
encompassing all acts of terrorism, across time and space, and (2) specific in identifying and circumscribing the object of the study.

_Terrorism is the unlawful use of violence or threat of violence against persons, as well as serious damage or threats to property, critical infrastructure or systems, carried out by non-state actor organizations, members or supporters of such organizations, small groups or individuals who are motivated by religious, political, or other ideological beliefs, and aim to install fear in and coerce governments or societies in pursuit of the furtherance, advancement or promotion of goals that are usually political, social, religious or ideological._ (Wolfowicz, Hasisi, & Ovadia, 2016)

The concepts of “radicalisation” and “violent extremism” are also clarified to provide an exhaustive portrait of the overall terrorism process. Radicalisation is “the process whereby individuals come to hold radical views in relation to the status quo” (Barlett, Birdwell, & King, 2010, p. 1). An important aspect is that process does not always lead to violence. Whereas cognitive (non-violent) radicalisation is the adoption of the radical ideas per se, violent radicalisation occurs when an individual takes the additional step of employing violence to further their extreme views (Vidino & Brandon, 2012). Therefore, to provide more clarity, some authors choose the term “radicalisation into violent extremism or terrorism” (Clutterbuck, 2015). The Council of Europe (2015) considers violent extremism as the promotion, support and commitment of acts that defend a particular ideology that opposes core democratic principles, and that may lead to terrorist attacks.

The understanding of the term Islamic terrorism is linked to the concepts of Islamism and jihadism. The Islamism ideology considers that Islam is not just a religion, but a political and legal code of conduct (Mozaffari, 2007). Islamists do not accept the separation between religion and state, and would like to reorder the government and society in accordance with Islamic law (Mozaffari, 2007). Islamism inspires various extremist Muslim organisations. In the 1990s these organisations focused on mainly national causes. However, since 2001 their actions have become transnational to evolve into the “global jihad” ideology (Wiktorowicz, 2005). Islamists believe that they need to wage “jihad”, “defined as armed struggle against the enemies of Islam, including non-Muslim nations, and the current rulers of Muslim states who have supplanted God’s authority with their own.” (Rabasa, Pettyjohn, Chez, & Boucek, 2010, p. 2)
Organisational forms of current Islamic networks include decentralised and highly mobile structures (Rabasa et al., 2010). Current actions include plotting of attacks in Europe by returning jihadists, who are trained and indoctrinated. Smaller-scale attacks perpetrated by relatively autonomous local cells are also common (Rabasa et al., 2010).

### 2.3.1.2 Far-Right Extremism and Terrorism

Far-right extremism and terrorism encompass a diverse range of groups with different beliefs and goals (Caiani & Parenti, 2013). Common ideological features of contemporary far-right groups include racism, ultra-nationalism, xenophobia and opposition to liberal democracy (Ramalingam, 2014). Far-right groups include neo-Nazi terrorist cells, anti-Islam activists, registered political parties, youth street gangs and informal groups, gathering around sports events (Caiani & Parenti, 2013). Not all types of groups defend their ideological belief through terrorist attacks. Some groups have a record of low-level acts of violence and hate crime, while others commit deadly attacks, such as arson or violent assaults. Apart from that, far-right extremists engage in a range of activities, such as vandalism, racist, abusive slogans and graffiti (Ramalingam, 2014).

### 2.3.2 Organised Crime

Organised Crime (OC) is also a contested concept. Academics, public institutions and international organisations have long been debating on the definition of OC (von Lampe, 2015; Calderoni, 2008; von Lampe, 2008; Symeonidou-Kastanidou, 2007; Hagan, 2006; Finckenauer, 2005; Hagan, 1983). The label “organized crime” was first adopted in the early Twentieth Century in the U.S. referring to a “criminal class” defined as capable of develop and maintain “crime as a business” due to the inefficient local judicial enforcement and a de facto immunity (von Lampe, 2001, p. 104). Since then, the expression OC has gained more prominence. According to Paoli (2014b, p. 3376) two perspectives succeeded over time: “(1) a set of stable organizations illegal per se or whose members systematically engage in crime and (2) a set of serious criminal activities, and particularly the provision of illegal goods and services, mostly carried out for monetary gain”. The shift of focus from groups to activities would have followed the evolution of OC. In addition to the mafia-type organisations, actors of the transnational arena could be included, such as those groups and networks illegally operating in the markets of prohibited psychoactive drugs, human beings and international fraud, and to some extent organised cybercrime (Kleemans, 2014).

This work refers to the definition of “organized criminal groups” adopted by the international community within the Palermo Convention:

‘Organized criminal group’ shall mean a structured group of three or more persons, existing for a period of time and acting in concert with the aim of committing one or more serious crimes or offences established in accordance with this
For the purpose of this work, two different forms of OC will be considered: (a) traditional OC and (b) the new forms of cybercriminal organisations.

### 2.3.2.1 Traditional Organised Crime

Traditional OC refers to various forms of criminal groups, which could be divided in four broad categories: Mafia-type organisations, drug trafficking organisations (DTOs), street gangs, and other OCGs.

**Mafia-type criminal groups** share four main characteristics: (1) longevity, (2) organisational and cultural complexity, (3) the claim to exercise a political dominion over its areas of settlement, and (4) their ability to control legitimate markets (Paoli, 2014a). The term “Mafia” originally referred to the Sicilian organised group, and then metonymy extended its use. It is now applied to other large-scale and long-lasting criminal organisations worldwide (Varese, 2006, 2011). Some examples are the Sicilian Cosa Nostra, the Calabrian ‘Ndrangheta, the Italian-American La Cosa Nostra, the Chinese Triads, the Japanese Yakuza, and the Thieves-in-law in the ex-Soviet Union (USSR) (Paoli, 2014a; Varese, 2006).

**Drug trafficking organisations** (DTOs) are “complex organizations with highly defined command-and-control structures that produce, transport, and/or distribute large quantities of one or more illicit drugs” (U.S. Department of Justice, 2010, p. 10). Drug trafficking markets are the most profitable, demand-driven illicit markets worldwide (Savona & Riccardi, 2015; UNODC, 2010). Due to this high profitable potential, a variety of actors have alternated in supplying illegal drugs, including OCGs (Natarajan, Zanella, & Yu, 2015; Benson & Decker, 2010; Desroches, 2007; Dorn, Levi, & King, 2005; Pearson & Hobbs, 2001; Eck & Gersh, 2000). According to Dorn, Levi and King (2005), it is possible to make a distinction between: Business groups, which are mainly profit-driven; Political-driven groups, financing their political activities through direct involvement in trafficking activities or indirect “taxes” imposed to traffickers/producers.

A **street gang** is “any durable, street-oriented youth group whose involvement in illegal activity is part of its group identity” (Klein & Maxson, 2006, p. 4). According to the Eurogang research network (Weerman, Maxson, Esbensen, Medina, & van Gemert, 2009), this definition highlights the four main characteristics of street gangs. Street gangs would be (1) “durable groups”, meaning there should be continuity in group activities regardless members’ turnover. They are (2) “street-oriented” groups since their members spend a lot of time outside traditional public spaces such as home, school and work. Most of members are (3) “youth”, meaning teens or early twenties, and they (4) “behave illegally as part of the group identity”, participating to proper illegal activities and crimes and losing personal identity and identifying themselves with their group’s social identity.

**Other OCGs** fall within the definition of OC, including groups involved in trafficking in human beings, illicit cigarettes trade, firearms trafficking and
trafficking of environmental resources. Maritime piracy has also been considered as OC.

2.3.2.2 Cybercriminal Networks

The development of ICTs over the last decade has influenced both old-school hackers and the criminal syndicates of the physical world. Some authors have started to explore the organisation of cybercriminals and the involvement of traditional organised crime in cybercrimes (Leukfeldt, Lavorgna, et al., 2016; Lavorgna, 2015a; Broadhurst et al., 2014; McGuire, 2012). The networks operating online might not properly fit the definition adopted in the Palermo Convention. However, some of them would have achieved some level of organisational complexity, and would operate as loose networks based on temporary ties between members (Leukfeldt, 2014).

Relying on a sample of cybercrime groups, McGuire (2012) proposed a typology framework based on the extent of cyber activity and the structural sophistication. The first type within McGuire’s categorization (2012) operate exclusively online and is further differentiated into swarms and hubs. The swarms are disorganised organisations, which operate without leadership, but share a common purpose (e.g. the Anonymous) (Broadhurst et al., 2014). Unlike swarms, hubs are more organised and are often controlled by a hub of core criminals and a periphery of accomplices. According to Broadhurst et al. (2014) the organisations behind cryptomarkets such as the Silk Road fit this definition. The second type within McGuire’s typology (2012) are the hybrids, which carry out criminal activities both online and offline. In case of clustered hybrids, the criminals are well-organised in committing specific online crimes, (e.g. credit card skimming groups) (Soudijn & Zegers, 2012). Similarly to swarms, extended hybrids are more diffuse, although they have some level coordination that allows them to operate successfully for online and offline (e.g. child pornography groups) (McGuire, 2012).

The main activities of type III of McGuire’s framework (2012) are in the physical world, but they may extent some of them to the online environment. Hierarchies represents the traditional OC moving into the online world (McGuire, 2012). The crimes they are engaged in online are usually linked to the expertise of the group in the physical world (Hill & Marion, 2016). Aggregate groups are hardly organised and are usually formed for limited periods for generic purposes (McGuire 2012). They use of technology on an ad hoc basis, mostly to get affiliates to participate. Broadhurst et al. (2014) illustrate this type with the use of mobile phones to coordinate public disorder. By differentiating the criminal networks active online on the basis of their level of organisation, McGuire’s framework (2012) supported the authors in determining the scope of the systematic review (see Table 2). As the focus of the review is on the organised perpetration of online activities, it will not cover swarms, extended hybrids and aggregate groups.
Table 2 – Types of criminal networks covered by this systematic review based on McGuire’s framework (2012)

<table>
<thead>
<tr>
<th>Formal organisation</th>
<th>No formal organisation</th>
<th>Covered by this systematic review</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type I</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operating exclusively online</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hubs (e.g. Silk Road)</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Swarms (e.g. the Anonymous)</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td><strong>Type II</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operating both online and offline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clustered hybrids (e.g. credit card skimming groups)</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Extended hybrids (e.g. child pornography groups)</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td><strong>Type III</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operating offline, occasionally online</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hierarchies (e.g. traditional OCGs)</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Aggregate groups (e.g. public disorder gatherings)</td>
<td></td>
<td>No</td>
</tr>
</tbody>
</table>
2.4 Rationale of the systematic review

Among the emerging threats to the international community, terrorism and OC stand out due to the potential large scale and extreme harmfulness of their activities (Loeber & Farrington, 2014; Edwards & Gill, 2002). Nevertheless, the literature often argues that the extent of political, public and media attention is at odds with the lack of empirical research on these topics (Kleemans, 2008). Cyberspace is hard to police due to a lack of regulation, traceability, and the possibility for anonymity (UNODC, 2016; Everett, 2015; Europol, 2014a). Moreover, the dynamic and resilient nature of cyberspace leads to a difficulty in the application of traditional counter-strategies. New websites, platforms and virtual identities quickly replace those disrupted by the intervention of law enforcement agencies (Bingham & Buxton, 2015; Kelion, 2014; Maras, 2014; Europol, 2014b). New technological solutions for evading law enforcement constantly replace previous ones (Balduzzi, Cincaglini, Goncharov, & McArdle, 2013).

At the same time, the visibility of online activities –at least of some of them– and communication provide a unique opportunity for researchers and law enforcement agencies to monitor and collect information about the evolution of OCTNs-related cyber-activities (Décary-Hétu, 2016; Mandala, 2016; Wall & Williams, 2013). Cyberspace is attractive to researchers as “there remain aspects of organizational life in cyberspace that resemble the terrestrial world” (K.-K. R. Choo & Grabosky, 2014, p. 483). Tracing and monitoring online activities can help understanding OCTNs-related dynamics. The monitoring of signposts and entry points to the Deep Web, for example, has raised law enforcement interest (Spitters, Verbruggen, & Staalduinen, 2014; National Research Council, 2005). Other aspects include research on drug-related online activities (Bingham & Buxton, 2015; Winstock, 2015; Choi, Earl, Park, & Della Giustina, 2014; Maras, 2014; Martin, 2014).

A systematic review will provide results useful at two different levels: (a) provide evidence of online visibility of OCTNs, thus helping the understanding of the related dynamics; (b) collect supporting information about OCTNs online activities which can help the development of effective counter-strategies for law enforcement actors.

For reaching these goals, this systematic review goes back to the basis of academic knowledge, providing a comprehensive portrait of the existing evidence-based knowledge on the activities that mark the exploitation of cyberspace by OCTNs. This will lead to two different contributions:

- Systematically collecting evidence-based knowledge, thus providing academic basis for WP3 related activities.
- Analysing the collected results to identify the types of online activities carried out by OCTNs in cyberspace, providing support to the construction of strategies for online crime prevention, crime monitoring and crime policing.
2.4.1 **OBJECTIVES AND RESEARCH QUESTIONS**
The objective of this systematic review is to provide information that could help identifying OCTNs related online activities. This systematic review aims at exploring and investigating the following research questions:
1) Which are the most commonly reported online activities of OCTNs?
2) What methods are typically employed to study the online activities of OCTNs?
3) What are the similarities and differences between online activities conducted by OCGs and TNs?

What are the implications to advance prevention policies can be derived from the literature on OCTNs online activities?

2.5 **Methodology**
This review applied a systematic approach in identifying and selecting relevant literature.\(^1\) It relied on three sources of information gathering: (1) the electronic database search, (2) the contribution of experts\(^2\), and (3) the lists of references of the selected studies. Eleven social sciences databases and one general computing database ensured the multidisciplinary results of the searches. *Fehler! Verweisquelle konnte nicht gefunden werden.*

---
\(^1\) Methodological details: *Fehler! Verweisquelle konnte nicht gefunden werden.*
\(^2\) The authors of the systematic review are grateful to all the experts providing their contribution:
- Manuela Caiani, Associate Professor at the Institute of Scienze Umane e Sociali at Scuola Normale Superiore (SNS).
- Eric Rutger Leukfeldt, Senior Researcher Cybercrime at NSCR & Director Cybersecurity & SMEs at The Hague University of Applied Sciences.
- David Décary-Hétu, Assistant Professor at University of Montreal.
- Jytte Klausen, Fellow at The Wilson Center in Washington D.C.
- Rolf van Wegberg, PhD researcher at the Faculty of Technology, Policy and Management of Delft University of Technology and Research Scientist at TNO.
Table 3 reports the list of databases and sub-databases, indicating the technique applied for the search. A two-fold query structure ensured systematic and thorough results (Figure 2). The first part of the query represents the common element of both searches, including keywords related to the concept of online activities. The Boolean Operator “OR” separates the keywords within this category. The second part of the query differentiates between organised crime and terrorism. It includes key words relevant to the definitions of organised criminal groups and terrorist networks adopted by this systematic review (Paragraph 2.3). The Boolean Operator “AND” connects the two parts of the query.

Figure 2 – Query structure
Table 3 – List of databases and search strategies

<table>
<thead>
<tr>
<th>Database</th>
<th>Sub-database</th>
<th>Search Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBSCO</td>
<td>Criminal Justice Abstracts Full text</td>
<td>Abstract or Author-Supplied Abstract</td>
</tr>
<tr>
<td>Open Grey</td>
<td></td>
<td>Full Text</td>
</tr>
<tr>
<td>ProQuest</td>
<td>Social Sciences Premium</td>
<td>Abstract</td>
</tr>
<tr>
<td></td>
<td>NJCRS</td>
<td>Abstract</td>
</tr>
<tr>
<td></td>
<td>PsycInfo</td>
<td>Abstract</td>
</tr>
<tr>
<td></td>
<td>Abi/Inform</td>
<td>Abstract</td>
</tr>
<tr>
<td></td>
<td>International Bibliography of the</td>
<td>Abstract</td>
</tr>
<tr>
<td></td>
<td>Social Sciences</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Public Health Database</td>
<td>Abstract</td>
</tr>
<tr>
<td></td>
<td>Military Database</td>
<td>Abstract</td>
</tr>
<tr>
<td></td>
<td>EconLit</td>
<td>Abstract</td>
</tr>
<tr>
<td></td>
<td>PsycArticles</td>
<td>Abstract</td>
</tr>
<tr>
<td></td>
<td>Computing Database</td>
<td>Abstract</td>
</tr>
<tr>
<td>PubMed</td>
<td></td>
<td>Title and Abstract</td>
</tr>
<tr>
<td>Scopus</td>
<td></td>
<td>Title, Abstract and Keywords</td>
</tr>
<tr>
<td>Web of Science</td>
<td></td>
<td>Title, Abstract and Keywords</td>
</tr>
</tbody>
</table>

The results of the databases search (excluding newspaper articles) were:
- OC: n=10,588 records
- TR: n=23,463 records

Covidence software automatically removed the duplicates in the two databases of records. A team of trained researchers applied shared criteria for excluding irrelevant literature by screening titles and abstracts. At a first step, researchers excluded literature falling without the scope of this systematic review, i.e. not pertaining to terrorism, organised crime, Internet crimes and crime facilitated by the Internet. At a later step, researchers excluded non-empirical literature (theoretical), and literature from non-scientific sources (magazines). At least two researchers evaluated each study. A senior researcher resolved the conflicts.

The experts’ contribution and the review of bibliographies assessed the consistency of the electronic databases and validated the final list of studies selected. The number of documents meeting the criteria for the full-text screening were:
- OC: n=59
- TR: n=263

The full text screening applied inclusion/exclusion criteria based on methodological and publication characteristics (Figure 3 and Figure 4). The final evaluation of the methodologies of the relevant studies followed the guidelines from the Critical Appraisal Skills Programme (CASP).³ CASP inclusion criteria included (a) the complete elaboration of the study design, (b) a systematic data collection method and a rigorous data analysis, and (c) a clear statement of findings.

Out of the initial records, N=25 met all criteria for inclusion in the systematic review for organised crime, and N=41 for terrorist networks.

**Figure 3 – Flow chart of the selection of relevant literature on terrorism**
2.6 Findings of the systematic review

The findings are presented separately for terrorism and organised crime.

2.6.1 Online activities of terrorist networks

Forty-one empirical studies discussing the online activities of terrorist networks met the final inclusion criteria of the systematic review. Approximately half of the studies (n=20) rely on a mixed (quantitative and qualitative) method approach, and the remaining are qualitative (n=10) or quantitative (n=11). The majority of papers (n=27) are published in peer-reviewed journals and 5 are conference papers. There are per 3 entries for Ph.D. dissertations, think-tank reports and book chapters.

Islamic radicalisation and Islamic terrorism is the most discussed topic, reflecting the overall dominance of the Islamic terrorism discourse in the initial query results. Far right radicalisation is the second most relevant topic. All the remaining studies make no distinctions on ideological grounds (Figure 5).
The majority of the studies selected discuss several activities (Figure 6). Social networking, propaganda and recruitment are closely related, and some authors examine them together. The internal communication within a terrorist group and the networking between terrorist networks are the most discussed issues within the selected studies. The topics of propaganda, recruitment and mobilisation are also well represented. Six empirical studies provide some insights into training and the planning of attacks. There is little empirical analysis of the terrorist involvement in computer crimes, in contrast to the high number of anecdotal literature trying to conceptualise “cyber terrorism”. Only three empirical studies focused on computer crimes and cyber terrorism. The systematic review did not identify any empirical study on online financing of terrorist activities and money laundering.

2.6.1.1 INTERNAL COMMUNICATION AND SOCIAL NETWORKING

The use of ICTs for facilitation of internal communication and intergroup networking is one of the main topics of the selected studies. The facilitated communication and networking between terrorist organisations could serve as a force-multiplier in the advancement of their common goals and affect their structure (Veer asamy & Grobler, 2011; Weimann, 2004). 22 studies empirically test some of these hypotheses.
2.6.1.1 Intergroup networking

Six of the selected studies apply social network analysis to the intergroup organisation and communication of far-right organisations in Italy, Germany, Spain, U.K., U.S., and France. The research process included the identification of the most relevant far-right groups in each country, followed by the application of snowball techniques to the friendly links on their websites (Caiani & Parenti, 2013; Burrus, Smith, & Strahm, 2012; Tateo, 2005). Internet links might not be an ideal indicator of the true online relationships between organisations – they tend to be denser and less selective than other forms of communication (Burrus et al., 2012). Nevertheless, the authors of the six studies agree that the links between websites provide the groups with an effective mode to coordinate their political efforts (Caiani & Parenti, 2013).

In almost all countries there was a growing use of the Internet by far-right organisations to form online networks between themselves (Caiani & Parenti, 2013; Caiani & Wagemann, 2009). At the same time, there were significant variations among countries in the density and centralisation of the networks (Caiani & Parenti, 2013, 2011; Caiani & Wagemann, 2009). The German network was the densest and most centralised, with the leading role played by political parties (Caiani & Parenti, 2013; Caiani & Wagemann, 2009). The Italian and British networks seemed to be also centralised, but they were more segmented (Caiani & Parenti, 2013; Caiani & Wagemann, 2009; Tateo, 2005). According to Caiani & Parenti (2013, 2011) these differences reflected offline national characteristics, such as the openness of the cultural and political environment to far-right discourse. Content analysis also showed that far-right extremist groups tend to use the Internet to build collective integrity (Gerstenfeld, Grant, & Chiang, 2003). Around 80% of the websites studied by Gerstenfeld et al. (2003) contained external links and 49.7% linked to a website of a group falling within another category. Gernstenfeld et al. (2003) argues that this publicly visible communication made them appear more influential and organised, but also less extreme.

With regard to Islamic terrorism, significant differences on the network structures emerged depending on the geopolitical context analysed. Three papers applied social network analysis to three different communication channels of jihadist groups – Twitter, YouTube, and organisations’ websites (Al-Saggaf, 2016; Klausen, Barbieri, Melnick, & Zelin, 2012; Xu, Chen, Zhou, & Qin, 2006). Xu et al. (2006) use the hyperlinks between terrorist organisations’ websites to compare the networks in Latin America, the U.S. and the Middle East. The comparison concluded that the Middle-Eastern jihadist network was the largest one, with the highest average degree, indicative of a strong intention of the groups to cooperate with each other (Xu et al., 2006). The authors recognise that the results were linked to the diverse ideologies of the Latin American and US-based terrorist groups, making them less likely to cooperate. The high online interconnectedness of the Middle East Islamic groups was also confirmed by Janbek’s (2009) quantitative content analysis of 30 websites. Al-Saggaf’s study (2016) focuses only on the twitter activity of the six suspected radicals without looking at the online behaviour of their followers. Despite the limited sample, Al-Saggaf (2016) concludes that there is a significant, but incoherent interaction within the networks of the six users.
The study of Klausen et al. (2012) on the British al-Muhajiroun’s YouTube network examines whether the YouTube accounts linked to the jihadist group constituted a centrally directed network.\textsuperscript{4} To test this hypothesis, the study compared the activity of the jihadists to the one of another political movement, serving as a case-control (Klausen et al., 2012).\textsuperscript{5} The jihadist network had more outward-directed relations between each other than the other political movement. No controlling actor emerged in the network, but nodes with medium to high number of connections (Klausen et al., 2012). The results on betweenness centrality suggested that the network was built to resist law enforcement disruption and had a high degree of coordination. This is consistent with the assumption that al-Mihajiroun played an organisational role in the management of the examined YouTube propaganda accounts (Klausen et al., 2012). In a research on twitter networks of Western foreign fighters Klausen (2015) observes the same trend of highly controlled content and tight integration of individual and official group accounts.

\textbf{2.6.1.1.2 Extent of the use of ICTs for communication}  
While it is reasonable to believe that terrorist groups use ICTs to communicate internally, only one study quantifies the exact degree of this use. Gill et al. (2017) developed a database of 223 UK and US terrorist offenders centred on their online behaviours. To that end, they relied on both official and open sources, and limited their search to the period between 1990 and 2014. The analysis of the database showed that approximately 29\% of terrorist offenders communicated online with other extremists and the majority opted for the use of e-mails (15\%). The other two most preferable communication channels were chat rooms (9\%) and discussion forums (8\%). Far-right extremists were around 2.5 times more likely to use the Internet for communication with affiliates than jihadists (Gill et al., 2017). The use of online forums is the main topic of five of the selected studies. Torres-Soriano (2013) provide quantitative and qualitative data on the 7 most influential jihadi internet forums in the period January 2011 to June 2012. The success of the forums was dependent on their ability to sustain cooperative relations with terrorist organisations, which also determined their hierarchical structure (Torres-Soriano, 2013). This conclusion is confirmed by Torres-Soriano (2016b), in which the author analysed the private communications on the Ansar al-Mujahideen forum.\textsuperscript{6} Three tiers appeared in the hierarchical forum structure – administrators, moderators and users (Torres-Soriano, 2013). While the administrators sustained the links with the patron terrorist organisation, moderators would ensure the ideological coherence of the forum communication (Torres-Soriano, 2013). Although the system encouraged

\begin{footnotesize}
\begin{enumerate}
\item Al-Muhajiroun was a terrorist Salafi Jihadist organisation, based in the United Kingdom and linked to international terrorist groups.
\item The political movement used as a case-control is the Texas Tea Party, which is an American conservative movement supporting the reduction of the US national debt.
\item Ansar al-Mujahideen is a militant group operating in Pakistan that carries out attacks against Pakistan government forces.
\end{enumerate}
\end{footnotesize}
participation through gamification techniques, the active users were only 11% of all forum users (Torres-Soriano, 2013). Similarly, a policy report showed that terrorist networks’ forum discussions were active only for a short period of time and comments represented only a minor portion of forum users (Erez, Weimann, & Weisburd, 2011). In addition, terrorists’ forums experienced an average of 3.1 outages per 18 months, proving the high vulnerability of these platforms to disruptions (Torres-Soriano, 2013). The private communication between forum administrators showed that some of the main causes of instability of the forums were the shortage of human and material resources to sustain their presence (Torres-Soriano, 2016b). Administrators and moderators had poor internet connection and lack of sufficient resources to cover basic operational costs. In addition, the more popular the website was, the higher the risk of infiltration and cyberattacks were (Torres-Soriano, 2016b, 2013). Apart from discouraging the approval of new users, this threat also resulted in the banning of current active users and removal of posts (Torres-Soriano, 2013). Torres-Soriano (2016b, 2013) concluded that the lack of direct personal contact and the persistent threat of enemy infiltration have compromised the main benefits of forum communication platforms. On the contrary, Zelin and Fellow (2013) consider that the overall forum communication is durable because takedowns of major forums are mitigated through increased activity on smaller platforms. Some authors have also considered that jihadi groups have started to recognise the need to reach Westernised and younger audiences through multiple languages communication in forums (Torres-Soriano, 2013; Zelin & Fellow, 2013; Ducol, 2012). Ducol (2012) examines the French-speaking jihadisphere, and Zelin and Fellow (2013) compares the English-speaking forums to the Arabic ones. Both studies concluded that the non-Arabic forums had fewer registered users, postings and unique visitors than the Arabic ones, which suggested that the jihadi online penetration in the West was not as widespread as popular opinions imply (Zelin & Fellow, 2013; Ducol, 2012). At the same time, examining the thread-to-post ratio Zelin and Fellow (2013) noted that administrators were prioritising English-language communication at a higher rate than there was audience for it. Ducol (2012) recognised that although the French-speaking forum network was small, there was a circle of forum users that were particularly active.

2.6.1.1.3 Online vs. offline communication

Ducol (2012) raised the question of the connections between the online and the offline networks. Ansar al-Haq forum users learnt of the existence of the platform from online sources (51%) and real life social ties (41%) (Ducol, 2012). Despite the limitations of his case study approach, Ducol (2012) observed a dynamic interconnection between online jihadist communities and pre-existing “physical world” Islamism networks. Similarly, the 15 radicalisation case studies of Von Behr et al. (2013) suggested that online communication did not replace the need for in-person meetings, but only complemented it. Gill et al.’s (2017) observed that terrorists communicating

---

7 Ansar al-Haq is a prominent jihadist forum in French language.
online with co-ideologues were 3.89 times more likely to have had also non-virtual meetings. Gill and Corner’s (2015) study of lone actors confirmed that each type of interaction reinforced the likelihood of the other. Drozdova and Samoilov’s (2010) analysed the correlation between Al-Qaeda communication patterns and terrorist attacks to identify the patterns that signal forthcoming attacks. A database of 496 communication instances showed that the volume of interaction fluctuated over time and there was no consistent correlation between the peaks in communication and the attacks. However, there were differences between “low-tech” and “high-tech” communication patterns prior to attacks.\footnote{Low-tech communication is simple and relies on people and physical objects for limited transactions. Examples include courier and face-to-face communications, handwritten notes, regular mail. High-tech communication comprises advanced and typically complex modern IT enabled means, such as Internet, mobile and satellite, phones, global positioning, etc.} Increases of the volume of low-tech communication occurred way earlier than the attacks, which suggested that low-tech communication may be an indicator of an upcoming terrorist attack. In contrast, high-tech peaks usually occurred during the time of the attacks (Drozdova & Samoilov, 2010).

2.6.1.2 PROPAGANDA AND PSYCHOLOGICAL WARFARE

Propaganda – i.e. the spreading of misleading information, serving to promote a political cause – is one of the main outputs of terrorist networks’ online presence (Chatfield, Reddick, & Brajawidagda, 2015). Nineteen studies discuss various aspects of the online propaganda of terrorists.

2.6.1.2.1 The terrorists behind online propaganda

Two studies focused on the extent to which offenders convicted for terrorism spread online propaganda. Gill et al.’s (2017) quantitative study found out that 15% of the terrorists within the sample disseminated information online. The majority of these terrorists only circulated the content produced by others and did not generate the materials themselves. Some lone actors were not just consumers of online extremist content, but were actively spreading propaganda to convince others of their beliefs or to indicate their commitment to the cause (Gill & Corner, 2015). Some of them produced their own propaganda material and used their attack as a marketing tool for its dissemination.

Three studies provided some insights into the extent to which terrorist networks disseminate propaganda. Tsfari and Weimann’s (2002) analysis of 29 extremists showed that by 2002 most of the organisations with a website were based in Third World countries and had national, revolutionary or religious character. Server and domain data showed that the websites of all anti-government organisations were operated outside the states against which they were working (Tsfati & Weimann, 2002). This was not the case with far-right organisations, whose websites were managed within the home countries (Caiani & Parenti, 2013, 2011). The content analysis of their websites showed...
that more than 60% of them published ideological articles and papers, and 50% of them had a news section, displaying media coverage of their activities (Caiani & Parenti, 2013). The frequency of the dissemination of far-right propaganda, however, varied among countries and among type of extreme-right groups (Caiani & Parenti, 2011, 2009).

With regard to use of Twitter for propaganda, Klausen’s (2015) study concluded that 70% of Western foreign fighters retweeted content from other users instead of generating content themselves. Klausen (2015) considered that Western foreign fighters as “producers rather than the consumers of impact” by looking at the ratios between followers and following, and comparing it to the general Twitter users. While the majority of Twitter accounts belonged to individuals, some of them were managed on behalf of organisations or were created in the name of deceased (Klausen, 2015). The possibility that a number of persons stood behind a single Twitter account was suggested also by Chatfield et al. (2015). Twitter also brought the greater involvement of female supporters in propaganda activities (Klausen, 2015).

2.6.1.2.2 Target audience

Two authors made a distinction between propaganda towards insiders and outsiders (Chatfield et al., 2015; Caiani & Parenti, 2013). Janbek (2009) used the language of the website as an indicator for the intended audience. He argued that the organisations were targeting audiences beyond the Arab world due to the prevalence of contents in Western languages. The presence of press releases on 23 sites was indicative of their efforts to reach traditional media outlets. Janbek (2009) concluded that almost all websites used rhetoric targeting men; less than half of them contained some content aimed at women. They were trying to reach younger audiences through multimedia content.

Mair (2017) used the language and internet connectivity as indicators of the intended audience of the twitter activity of the terrorist group Al-Shabaab. The author concluded that the main audience groups were the general Kenyan population, the terrorist sympathizers and the Kenyan government. During their terrorist attack the West and the media were not prioritized as targets of terrorists’ propaganda. Offline channels for propaganda were still preferred in areas with low internet connectivity (Mair, 2017).

A review of 29 terrorists’ websites in 2002 suggested that most terrorist networks attempted to appeal to (Tsfati & Weimann, 2002):

- **Local supporters** – evidenced by the use of local languages, the presence of sale and the detailed description of internal politics;

- **The international public** – indicators of this were the use of a second language and the presence of basic facts about the organisation;

- **Traditional media outlets** – indicated by the presence of press releases and contact details of press offices;

- **Enemy audiences** – there were efforts to demoralize the enemy and create feelings of guilt on the websites.
2.6.1.2.3 Types and purpose of the propaganda messages

The systematic review adopts the list developed by Chatfield et al. (2015) to clarify the various types and purposes of propaganda messages. Table 4 shows propaganda goals with evidence from the selected studies. The religious narrative constitutes a great proportion of the online propaganda of Islamic Terrorist networks. Erez, Weimann and Weisburd’s (2011) analysis of 2112 forum discussions confirmed that the majority of them provided information on Islamic teachings through the sharing of downloadable materials (43%) or citing religious quotes (51%). These discussions varied from threads on general lifestyle habits to threads justifying the organisational activities. The content analysis of the forums suggested that the information spread also news or official press releases on recent developments and opinions on various topics. Janbek’s (2009) analysis of 30 Islamic Terrorist networks’ websites concluded that almost all of them included a mission statement, used a heavily religious rhetoric and took responsibility for previous attacks on the enemy. Half of the sites stressed the importance of active resistance and individual sacrifice through listing the names of martyrs and detailing their lethal operations. The organisations actively monitored their representation in mainstream media (Janbek, 2009). Similarly to forums, twitter messages also referred predominantly to jihadist dogma, mainly religious instructions (39%), reporting from the battle (40%) (Klausen, 2015).

Table 4 – Aims of propaganda messages

<table>
<thead>
<tr>
<th>Study</th>
<th>Target population</th>
<th>Propaganda platform</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spreading fear and intimidating Western public opinion</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awan (2017)</td>
<td>Society at the global level (outsiders)</td>
<td>Social media (Facebook and Twitter)</td>
<td>1264 messages of ISIL containing physical threats</td>
</tr>
<tr>
<td>Salem et al. (2008)</td>
<td></td>
<td>Videos on websites</td>
<td>Videos documenting and claiming responsibility for attacks</td>
</tr>
<tr>
<td><strong>Supporting the moral legitimacy of terrorism violence</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tsfati &amp; Weimann (2002)</td>
<td>Society at the global level (outsiders); Potential supporters (insiders)</td>
<td>Websites</td>
<td>Terrorists legitimising narratives: &quot;no other choice&quot; motive; demonization of the enemy; terror as &quot;weapon of the weak&quot;</td>
</tr>
<tr>
<td>Salem et al. (2008)</td>
<td></td>
<td>Videos on websites</td>
<td>Videos on the suffering and injustice caused by the West</td>
</tr>
<tr>
<td><strong>Engaging wider audience with terrorism ideologies and actions to shift individuals' perspectives</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tsfati &amp; Weimann (2002)</td>
<td>Society at the global level (outsiders)</td>
<td>Websites</td>
<td>Presentation of the main goals; Refraining from detailing violent activities; Nonviolence rhetoric and claim of peaceful solutions</td>
</tr>
<tr>
<td>Ismail (2009); Janbek (2009)</td>
<td></td>
<td>Websites</td>
<td>Construction of an image different from the one projected by traditional media</td>
</tr>
<tr>
<td><strong>Generating support from local communities or target population</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Janbek (2009)</td>
<td>Potential supporters (insiders)</td>
<td>Websites</td>
<td>Terrorist networks stressed the importance of active resistance and individual sacrifice through the presentation of martyrs and their activities</td>
</tr>
</tbody>
</table>
Ismail (2008) examined the websites of Hezbollah and the Palestinian Islamic Jihad to see how they used them to construct ideologies and counteract the U.S. mainstream ideologies. The study argued that the organisations’ websites aimed at legitimising their activities as emanations of patriotic and religious duties and construct them as mainstream and widely accepted components of the Islamic world. The two terrorist networks used the virtual space to construct an image that differed significantly from the one projected by traditional media, and thus counteract the Western narrative (Ismail, 2008).

A substantial proportion of the jihadist twitter activity aimed at exaggerating the power of the jihadist movement, while mocking the enemy forces and blaming them for civilian causalities and destruction of infrastructure (Chatfield et al., 2015). A similar rhetoric of focusing on the negative influence of the “Western occupation” was identified by Drissel (2015), who studied the twitter and website activity of the Taliban insurgency in the first half of 2012. The culture-based “clash of civilisations” was framed within the category of “global justice”, which referred to various issues such as foreign intervention, colonialism, universal human rights and economic inequality (Drissel, 2015). The suffering and injustice caused by the Western forces were also a main topic of the video propaganda of jihadists (Salem, Reid, & Chen, 2008).

An analysis of 100 ideological articles suggested that ISIL’s strategy is to balance between the violence topic and themes of ideology and pragmatism (Derrick, Sporer, Church, & Scott Ligon, 2016). This reflected the strategy of ISIL to acquire and hold territory through a mixture of violence-seeking and pragmatic leadership (Derrick et al., 2016).

With regard to the objective of twitter postings during a terrorist attack, Mair (2017) quantified that almost half (49%) of the tweets were aimed at furthering the ideology of the Al-Shabaab network. The other most common objectives were to justify the attack (21%), to generate news content (20.7%) and to threat with future attacks (19%) (Mair, 2017). The focus of terrorists during an attack in a third-world country was not on anti-Western rhetoric and
threats against the West, in contrast to the general patterns of online activities of other terrorist networks.

2.6.1.2.4 Approaches to the spreading of online propaganda

The change of modes for online propaganda of terrorist networks followed the evolvement from Web 1.0 to Web 2.0. This reflected the way scholars approached the topic of online activities of terrorist networks and their choice of virtual platforms to be studied.

The majority of Islamic Terrorist networks’ resort to extensive use of multimedia materials on their websites. Janbek’s (2009) website analysis suggested that the majority of terrorist groups used audio (86%) and video (76%) files. The multimedia content aimed at educating the users and motivating them to support the movement (Janbek, 2009). Approximately 60% of the websites also featured pictures depicting organisation’s military power, martyrs and the people of the country of operation (Janbek, 2009). 63% of the videos (n=60) examined by Salem (2008) were documentaries, claiming responsibility for attacks. Their production features and cultural expressions (e.g. religious quotes, exchange of kisses) were framed in a way that helped targeted audiences identify with the Islamic movement (Salem et al., 2008).

Klausen (2015) argues that social media has freed Islamic terrorist networks from the dependency on traditional media. However, most authors agree that Twitter will not replace the forums as the main propaganda distribution channels among jihadists, as the content is more easily controlled on this type of platforms (Klausen, 2015; Zelin & Fellow, 2013). In this regard Janbek (2009) argues that social media instruments do not automatically benefit terrorists but require human resources that are capable of using their potential. His examination of Al-Qaeda communication in the period 1998-2015 suggested that the terrorist group used twitter posts solely for publishing links for downloading materials and did to manage to interact with its followers or turn content viral (Janbek, 2009).

The use of pictures on Twitter is widespread and they range from “selfies of the foreign fighters” to images of dead bodies (Klausen, 2015). The coding of the images showed that the pictures correspond to the messaging of the texts and tried to provide evidence of the success of the jihadists on the battlefield (Klausen, 2015). On the contrary, Mair (2017) noted the absence of images in the tweets of Al-Shabaab during one of its terrorist attacks. Mair (2017) argued that Al-Shabaab wanted to keep control of the narrative and appeal to a wider audience through refraining from showing atrocities. Awan (2017) also studied the propaganda tools of ISIS on social media by examining 100 Facebook pages and 50 different Twitter users. The most common social media tools were web links, retweets and shares, likes and hashtags, followed by the use of images and videos (Awan, 2017).

Two of the selected papers examined the quality and credibility of terrorist networks’ website propaganda. Qin et al. (2007) focus on the advancement of web usage by assessing the technical sophistication, content richness and web interactivity of terrorists’ websites. The authors compared the terrorist websites with that of US federal government ones to evaluate the online capabilities (Qin et al., 2007). With regard to technical sophistication, the
government websites were significantly more advanced than the terrorist ones in terms of using dynamic programming languages, but were equally good in applying advanced HTML techniques and embedded multimedia techniques. In terms of content richness the government websites provided a larger volume of multimedia content, which Qin et al. (2007) attributed to the low capacity and instability of terrorists’ servers. While US government institutions scored higher in one-to-one-level web interactivity, terrorist websites were better in community-level interactivity. Qin et al. (2007) concluded that terrorist networks had invested a substantial amount of resources in building technical expertise to create their web infrastructure and use it for propaganda. Similarly, the four terrorist organisations studied by Spinks (2010) were advanced in the design and use of websites for propaganda dissemination. Spinks (2010) used the overall design, functionalities and information reliability as indicators of the credibility of websites. The findings suggested that the websites of terrorist groups had the potential to be effective in reaching their goals. The differences in the resources of particular groups and the level of their appeal were reflected in the credibility of their online presence (Spinks, 2010).

2.6.1.3 Recruitment, Mobilisation and Self-Radicalisation

One of the goals of propaganda is to mobilise supporters and recruit new ones. While it is difficult to determine what part of terrorist networks’ online messages focus exclusively on these goals, seventeen studies discuss Internet recruitment strategies. Recruitment does not limit itself to involvement in violent activities, but also covers engagement activities such as promotion of group events and calls for donations (Davies, Bouchard, Wu, Joffres, & Richard, 2015; Caiani & Parenti, 2013).

While radicalisation and recruitment are two distinct processes, they are closely related (Davies et al., 2015). Radicalisation is linked to the receptiveness of people to recruitment which, by contrast, involves also practical steps to get involved in a terrorist network. According to Davies et al. (2015), the Internet has blurred the distinction between radicalisation and recruitment, and terrorists’ strategy adapted to encourage self-recruitment through radicalisation.

2.6.1.3.1 Recruiters

Two studies investigated the extent to which convicted terrorists tried to recruit supporters online. Only 9% of the sample of terrorism offenders studied by Gill et al. (2017) sought to recruit other people online. Gill and Corner (2015) noted that some lone actors originally tried to recruit affiliates, but were unsuccessful in their efforts. The study provided examples of lone actors that created websites and actively reached for supporters (Gill & Corner, 2015).

Davies et al. (2015) developed a recruitment and extremist scale to analyse the extent to which extremist organisations used recruitment narratives on their official websites. The review of the sites of 8 groups showed that groups engaged into local conflicts were more passive in their online recruitment efforts than groups with global causes (Davies et al., 2015). The websites
active in recruitment avoided discussing violence, while the ones with moderate to low recruitment efforts were more likely to debate it openly, predominantly in a defensive context. The study also suggested three levels of recruitment messages (Davies et al., 2015):

1) Development of a narrative around the topic of oppression and injustice;

2) Justification of violence through provision of evidence of violence against them and explaining that they should act to protect themselves;

3) Offering of opportunities for people to support and join the cause.

In his evaluation of terrorists’ Twitter activity during an attack, Mair (2017) suggested that the online recruitment activity of the Al-Shabaab terrorist group corresponded to their overall recruitment strategy, i.e. to refrain from recruiting foreign fighters. Accordingly, they were not using the success of the attack to publish mobilising messages.

2.6.1.3.2 Potential recruits

Awan’s (2017) study of social media presence of ISIL concluded that there were seven types of offenders’ behaviour on these platforms: cyber mobs, loners, fantasists, thrill seekers, moral crusaders, narcissists and identity seekers. Two of the most commonly identified online behaviours – “thrill seekers” and “moral crusaders” – were most prone to recruitment, as they expressed their desire to join and fight with ISIL (Awan, 2017). The group of identity seekers were the ones to which ISIL recruitment efforts appealed the most.

According to Conway and McInerney (2008) the mean age of the jihadist supporters on Youtube was 26.7 and 86% of them were within the 18-34 years range. However, some of the supporters were lying about their ages to avoid restrictions on “inappropriate content” and were most probably under 18 years old. Jihadists active on Youtube were mainly located in Western countries – US (42%), UK (15%), Canada (8%), Germany (7%). This could be attributed to the diaspora communities in these countries (Conway & McInerney, 2008).

Marcellino et al. (2017) tried to measure the appeal towards Daesh in Egypt through a combination of linguistics and statistical methods. The findings suggested that only a small percentage (1 to 2%) of Twitter users in the region echoed Daesh’s discourse. However, when there was support, it was intense (Marcellino et al., 2017).

2.6.1.3.3 Far-right mobilisation

Four studies on far-right examined the use of the websites for mobilisation. Caiani and Parenti (2013, 2011, 2009) concluded that approximately one-fifth of the extreme right organisations in Europe and the US were trying to mobilise their supporters by publishing event calendar and information on meetings, concerts and demonstrations. Half of these groups shared events organised by other extreme-right groups (Caiani & Parenti, 2013, 2011, 2009). The overall use of such mobilisation techniques varied between countries. The Spanish groups, for example, were less active in their mobilisation efforts than the Italian ones (Caiani & Parenti, 2011, 2009). According to Gerstenfeld’s (2003) review of far-right groups’ websites, 54.8% sold merchandise items
and 44.1% included membership forms, which suggested that they heavily relied on the Internet as a recruitment tool.

The study of Ray (2000) on far-right recruitment efforts towards children refuted empirically the proposition of literature that far-right groups are proactive and aggressive in targeting children. Ray (2000) concluded that the extreme right groups were opportunistic in their recruitment of children.

2.6.1.3.4 Self-radicalisation
The internet connectivity facilitates the formation of terrorist cells and inspires self-radicalised individuals lacking prior history of extremist activity (Johnson et al., 2016). The authors try to provide quantitative predictions about the evolution of online support by relying on longitudinal records of cyber support activity for ISIS and comparing it to that of civil protestors. Although pro-ISIS aggregates were ad-hoc group of followers that had never met, they had the potential to adapt in order to extend their lifetime and increase their size (Johnson et al., 2016). These adaptations were a response to the risks of government monitoring and shutdowns and included name changes, flipping between privacy settings and reappearances of profiles (Johnson et al., 2016).

The study of Scrivens (2017) identifies the most radical users of web-forums through semi-automated tools of POS tagging and sentiment analysis. While detecting the most extreme users does not mean that they are the most dangerous, it might prevent their “echo chamber” effect on easily influenced readers (Scrivens, 2017, p. 3). The findings did not highlight a specific “most radical users” typology. However, the radical score developed by Scrivens (2017) proved flexible enough to assess the online sentiment activity.

Two studies examine the role of Internet as a facilitator of radicalisation. Conway and McInerney (2008) examine Youtube videos to determine whether they facilitate the radicalisation with or without prior predisposition for jihadist activity. The analysis showed that Youtube might provide a platform for bottom-up radicalisation, which covered all individual consumption of online jihadist materials via random or targeted browsing (Conway & McInerney, 2008). Von Behr’s (2013) examination of 15 cases of terrorism shows that the internet enables but does not accelerate the process of radicalisation. Several of the reviewed terrorist cases had periods of inactivity on the Web, which suggested that there were external factors accelerating the radicalisation. In all cases the radicalisation process included virtual and physical communication with like-minded individuals, which suggested that online radicalisation was not a self- but a social process (von Behr et al., 2013).

2.6.1.3.5 Transition from online to offline supporter
Participation in the forums might provide sufficient credentials to become part of armed forces of terrorist groups (Torres-Soriano, 2016b). The contacts created on the forums enabled online supporters to establish physical contacts with members of terrorist networks, plan travels and ultimately join them on the ground. Based on the internal communication of Ansar Al Mujahideen’s forum, Torres-Soriano (2016b) argued that the cyber-jihadists often transited to the world of the terrorist armed forces and played the role of representatives of the forum on the ground. This transition was also a result of dissatisfaction with the everyday repetitive tasks of forum administration,
which did not correspond to the jihadist ideal projected on the forums (Torres-Soriano, 2016b). Stenersen (2008) noted that requests to join jihad were extremely rare in open forums. The few cases either received a general uninformative answer or some open source information on the countries where the terrorist networks were active. Since 2006 administrators have started to close such forum threads and to advise against posting “how to join” questions (Stenersen, 2008).

2.6.1.4 INFORMATION GATHERING, TRAINING AND PLANNING OF ATTACKS

Six of the selected studies discuss the use of Internet for information gathering, training and planning of attacks. The Internet could be used for training in two main ways: as a library of training manuals, and as an interactive platform for discussion of training-related issues and communication with online trainers (Stenersen, 2008). The section also discusses how online training is linked to the preparation of terrorists’ attacks. Two of the studies used open-source data to compile dataset of convicted terrorists and/or lone actors, and assessed the extent to which they used the Internet for training (Gill et al., 2017; Gill & Corner, 2015). Around 46% of lone actors learned through online source information about a target, how to prepare the attack, or overcome technical problems (Gill & Corner, 2015). A smaller percentage (32%) of convicted terrorists used online resources such as bomb-making instructions, Inspire magazine, and surveillance manuals to prepare for their attacks (Gill et al., 2017). The police also confiscated copies of videos of previous terrorist attacks, body armour and chemicals purchased online (Gill & Corner, 2015). Differences among the ideological groups emerged: far-right terrorists were approximately 4 times more likely to prepare for an attack with online sources and 3.4 times more likely to learn online than jihadists. Online learning was also positively correlated with the probability of non-virtual interactions with like-minded individuals (Gill et al., 2017).

Three studies discuss the types and contents of jihadi training materials. Training materials were usually published in the form of encyclopaedias, focusing either on a particular topic (e.g. weapons, explosives) or a combination of topics (e.g. The Encyclopaedia of Preparation) (Stenersen, 2008). Serials and periodicals issued on a regular basis were also common. Stenersen (2008) noted that these materials were often combining ideological with military preparation and were progressing from basic information to advanced subjects. They were also interactive, encouraging contact with the serial’s issuer and the submission of questions and homework (Stenersen, 2008). Forums offered discussions on technical skills, referencing to manuals (Erez et al., 2011). While the format and design of training materials were constantly improving, the same could not be said about the content (Stenersen, 2008). The threat of law enforcement monitoring explains the reluctance of jihadists to share new ideas and technologies online.
Some authors agree that videos are widely used as training materials (Salem et al., 2008; Stenersen, 2008; Salem, Edna, & Chen, 2006). An exploratory study of 20 jihadi extremist groups’ videos showed that 90% were related to violent attacks and included planning sessions with maps and logistic preparation (Salem et al., 2006). A subsequent content analysis of jihadi videos noted that suicide attack and beheadings videos illustrated the overall process of execution – planning, preparation and execution (Salem et al., 2008). Stenersen (2008) concluded that not all terrorist networks had the capacity to produce high-quality videos. The best quality videos were products of Hizballah, which were then modified and made available on jihadists’ websites to serve Al-Qaeda goals (Stenersen, 2008).

Most authors agreed that online training covered only basic skills and knowledge and was a step towards real-life training (Stenersen, 2008; Salem et al., 2006). Terrorist networks such as Al-Qaeda did not always make and organised effort to train their supporters (Stenersen, 2008). The majority of their online training materials were products of sympathisers.

Only 5% of convicted terrorists signalled online their intentions to commit a violent act prior to the attack itself (Gill et al., 2017). Lone actors were more likely to signal an attack online, or to leave an online note to friends (Gill & Corner, 2015) The terrorists convicted for attacks against government targets used the Internet for preparation more than the ones targeting civilian targets (Gill et al., 2017). Lone actors also rely on the Internet for avoiding technical problems, such as failures to build bombs (Gill & Corner, 2015).

### 2.6.1.5 Perpetration of Computer-Focused Crimes

The advancement of ICTs has led to an increasing threat of terrorists’ attacks perpetrated through cyberspace and to the introduction of the term “cyber terrorism”. Some authors argue that jihadist supporters perpetrate various types of cybercrimes (e.g. credit card fraud) to finance terrorist activities (Weisburd, 2008). The systematic review suggests that empirical research on the topic is scarce, which could be related to: (a) the difficulties in attributing cybercrimes and cyberattacks to terrorists, and (b) the difficulty of acquiring data.

Yunos and Ahmad (2014) provided a guideline for classifying a cybercrime as an act of cyber terrorism. They conceptualise the term “cyberterrorism” on the basis of 22 face-to-face interviews with experts. They used a three-step coding analysis to define six different perspectives for conceptualising cyber terrorism: motivation, target, method of attack, domain, impact, and tool of attack. Cyberspace is the domain in which the terrorist attack is perpetrated, while the method of attack refers to how the goals are achieved (Yunos & Ahmad, 2014). The majority of the interviewed experts validated the proposed framework. Some made recommendations, such as the combining of “method and domain”.

Erez et al. (2011) analysis of 2000 forum threads suggested that the largest portion of discussions on illegal activities involved computer crimes (3% of the total forum discussions). These computer crimes involved illegal software downloads (2.2%), breaking into sites and/or servers (0.4%), defacement of websites (0.3%), DDoS attacks and theft of credentials (0.1%) (Erez et al.,
The research concluded that discussions on computer-focused crimes were uncommon and there was scarce evidence of their use for financing of terrorist activities (Erez et al., 2011). Ray (2000) reaches similar conclusions with regard to forums of far-right groups. He suggests that there is no real awareness of the hacking culture (Ray, 2000). Even the websites that contained some hacking information would not provide links to “crackers” platforms.

2.6.2 ONLINE ACTIVITIES OF ORGANISED CRIME

N=25 empirical studies discussing met the final inclusion criteria. The majority of studies (n=16) rely on a qualitative approach; the remaining on mixed qualitative and quantitative methods (n=7 each), and quantitative (n=2). The majority of the studies (n=24) are peer-reviewed papers. More than a half of the studies (n=16) focus on traditional organised crime networks, or on illegal activities traditionally involving sophisticated criminal organisations. The remaining studies focus on cybercriminal networks (n=9) (Figure 7).

Recruitment and networking are the main activities analysed in the included studies (n=14) (Figure 8); perpetration of computer-focused crimes (n=9); personal use of social networking sites by organised crime affiliates, dissemination of criminal culture and “indirect propaganda” (n=8); traditional offline activities facilitated by online means, mainly transit crimes (e.g., drug trafficking) (n=8).³

Figures 7 – Types of criminal networks

³ Some publications discuss more than one online activity.
2.6.2.1 PERSONAL USE OF SOCIAL NETWORKING SITES BY ORGANISED CRIME AFFILIATES, DISSEMINATION OF CRIMINAL CULTURE AND “INDIRECT PROPAGANDA”

The included literature has two main objectives:

1) Analysing the personal social media profiles of self-declared OC affiliates and supporters to verify their traceability and the types of content they share. The included studies focus on gangs (Pyrooz, Decker, & Moule, 2015; Hellemont, 2012) and Mexican DTOs (Dávila, 2016; Nix, Smith, Manjarrez, Petrocelli, & Rojek, 2016; Womer & Bunker, 2010).

2) Analysing the official and/or fan pages of OCGs on social media and other websites to provide insights into the online “official self-portrait” of the criminal organisations. The included studies focus on gangs (Moule, Pyrooz, & Decker, 2014; Morselli & Décary-Hétu, 2013; Décary-Hétu & Morselli, 2011), and Mexican DTOs (Dávila, 2016).

2.6.2.1.1 Personal social media profiles of self-declared organised crime affiliates and supporters

Drug Trafficking Organisations

All research on Mexican DTOs analyse the members’ public profiles on open sources, mainly social networking sites. The three studies identified comparable characteristics (Dávila, 2016; Nix et al., 2016; Womer & Bunker, 2010).

The users had no privacy issues in using open and public personal profiles. Most of the identified members self-admitted their affiliation to a drug trafficking organisation. There was sometimes an explicit reference to hierarchical positioning. According to Nix et al. (2016), more than the 60% of their sample (75 Facebook profiles) include their affiliated criminal organisation as their rank or occupation. Similarly, Womer and Bunker (2010) highlight that
the “narco-occupation” is specific for Mexican cartels; such self-identification is uncommon in other organisations. The Mexicans “advertise their affiliation as a profession with an income, in addition to a group membership.” (Womer & Bunker, 2010, p. 86)

Dávila (2016) frames this reporting propensity of narco-profession within a discourse on social acceptance. Poverty in Mexico is widespread, and “DTO-members are admired as anti-heroes” personifying successful business models (Dávila, 2016, p. 70). The drug trade is re-built upon an image of a normal labour market and a legitimate formal business (Dávila, 2016). The Sinaloa Cartel, for example, was particularly focused on the “rebrand” of the drug trafficking business (Dávila, 2016). The authors underline how the affiliates to drug trafficking organisations are not only tolerated but also supported and admired by their community (Dávila, 2016, p. 70). Narcos become aspirational role models; these online role-models emerge at the background of the offline fragmentation of DTOs, characterised by the emergence of small groups composed of discontent youths (Dávila, 2016).

Three main categories of web content shared by DTO affiliates emerged from the literature: (1) drug trafficking-related activities (2) self-glorifying portraits of criminal lifestyle and culture; (3) symbols showing affiliation and pride for the organisation.

The web content on drug trafficking-related activities includes multimedia and text messages, showing members’ everyday work as traffickers or their abuse of illegal substances (Dávila, 2016; Nix et al., 2016; Womer & Bunker, 2010). Nix et al. (2016) report cases of original photos of dead bodies and tortured persons. The females show their affiliation as “hawks” (be the lookout), sharing pictures with radios and cell phones; they also have (or pretend to have) the role of “panteras” (panthers), seducing rivals for negotiation or murder purposes (Dávila, 2016, p. 62). According to Womer and Bunker (2010), some characters wear masks, while others do not cover their identity. In some cases, even sensitive strategic and operational information might be shared, including locations and travel plans (Dávila, 2016; Nix et al., 2016). Some posts of illicit activities are accompanied by the tag “on the field” (Dávila, 2016, p. 66). The transnational dimension of the drug trade emerges from online evidence of travels for criminal purposes (Nix et al., 2016, p. 402).

The self-glorifying portraits of criminal lifestyle and culture include gangster style pictures, with guns as symbols of power, wealth, success, and status (Dávila, 2016; Nix et al., 2016; Womer & Bunker, 2010). Cartel members are surrounded by sexualised females and luxurious environment. Dávila (2016) describes the female self-portraying. Young and low-level female members post provocative selfies with guns “portraying themselves as seductive killers” (Dávila, 2016, p. 62). High-ranked elite members present themselves as “femmes fatales”, who are attractive, provocative and powerful (Dávila, 2016, p. 62).

DTO symbols showing affiliation and pride for the organisation include DTO names, logos, insignias, graffiti, narco-corridos and emblematic images (Dávila, 2016, p. 63; Womer & Bunker, 2010, p. 88), but also popular photos of known cartel members (Womer & Bunker, 2010, p. 88), posters, music and even poetry (Womer & Bunker, 2010, p. 90). Recognisable tattoos are present
with low visibility rates (Nix et al., 2016, p. 403). Authors agree that the members of DTOs use social networking sites “to build an image of power for themselves and the organization” (Dávila, 2016, p. 67).

**Street gangs**

Hellemont (2012)’s study on blogs of Bruxelles black African gangs and Womer and Bunker (2010)’s study on social media of Sureños gangs highlighted that illegal activities were less central to the discourse respect to DTOs. Similarly to DTOs, central types of web content were: (1) self-glorifying and portraits of criminal style and culture; (2) symbols showing affiliation and pride for the organisation.

Self-glorifying and portraits of criminal style and culture play with public’s imagination to build up fictional and stereotyped characters (Hellemont, 2012). Individuals make references to fictional criminals as Scarface, the Godfather, or members of famous gangs (e.g., Bloods, Grips) to show gangster appearance (Hellemont, 2012). Similarly to DTOs, guns and weapons played a key role in spreading portraits of power and status (Hellemont, 2012; Womer & Bunker, 2010). Wealth is also displayed through expensive branded goods and piles of money (Hellemont, 2012). Transgression is spread through references to alcohol and marijuana (Womer & Bunker, 2010).

Symbols showing affiliation and pride for the organisation include hand signs, blue bandanas, graffiti, tattoos, gang initiations, etc. (Womer & Bunker, 2010). Pyrooz et al. (2015) approaches the same topic from a different perspective, relying on interviews with gang members. Respondents self-declared to be online and to have technological skills, non-significantly different rates than non-gang youths. This is in line with contemporary social tendencies of young adults incorporating technologies in everyday life at a higher rate (Pyrooz et al., 2015). “Gang membership has no inhibitory influence on stifling these processes.” (Pyrooz et al., 2015, p. 48) The only slight difference from non-gang respondents was their longer internet presence at weekly rates, which would suggest the Internet satisfies the symbolic needs of gangs (Pyrooz et al., 2015). This validates the analysis of open sources, which likely captures at least a percentage of actual gang members.

In conclusion, “Online activities of gangs reflect symbolic rather than instrumental objectives” (Pyrooz et al., 2015, p. 49). However, Hellemont (2012) reminds that the online activities of gangs are not totally disconnected from the offline world. In his research, one-third of the authors of blogs appeared in public prosecutors’ databases for gang-related crimes in the offline world (Hellemont, 2012). The study of Pyrooz et al. (2015) confirms this view, suggesting that everything that happens online goes on Facebook or YouTube.

**Street gangs vs. DTOs**

Studies on personal profiles of (potential) members of Mexican DTOs and street gangs show similar web content and narratives. Facebook is the preferred social media website for both groups, maybe due to its own nature built upon personals’ contents, instead of social contacts (as for Twitter or MySpace).
A notable difference is that the social acceptance gained by narco-lifestyles in Mexico, ensuring a high propensity of DTO members to show up their criminal activities, their identity and cartel affiliation. This could be explained by their seek for social acceptance, and the normalisation of the narcoculture in the Mexican society (Dávila, 2016). On the contrary, street gangs favour the share of symbolic images and narratives. Due to this, the researchers had to be more cautious in assuming the affiliations due to the higher “rumour” of supporters and self-declared members. In this regard, the systematic review mitigates the risk of catching sampling distortions by including studies based on sources others than social networking sites.

2.6.2.1.2 Official and/or fan pages of organised crime groups on social media and other websites

**Street gangs**

The analysis of street gangs’ official and/or fan pages on social networking sites and other websites address their “web footprint”. The most famous groups across the world attract the highest number of members/followers; smaller gangs have smaller footprints (Moule et al., 2014; Morselli & Décary-Hétu, 2013; Décary-Hétu & Morselli, 2011). However, self-declared affiliates report that the online presence is heterogeneously distributed across the different gangs: not all highly organised gangs declared to be online. (Moule et al., 2014)

The web footprint of a gang also vary depending on the social network site analysed. In some cases the footprint improved or decreased over time (Morselli & Décary-Hétu, 2013). MySpace is characterised by higher content openly associated with criminal behaviours, and to more explicit and active conversation between users (Morselli & Décary-Hétu, 2013; Décary-Hétu & Morselli, 2011). The web footprint of a gang also vary depending on the social network site analysed. In some cases the footprint improved or decreased over time (Morselli & Décary-Hétu, 2013). MySpace is characterised by higher content openly associated with criminal behaviours, and to more explicit and active conversation between users (Morselli & Décary-Hétu, 2013; Décary-Hétu & Morselli, 2011). The web footprint of a gang also vary depending on the social network site analysed. In some cases the footprint improved or decreased over time (Morselli & Décary-Hétu, 2013). MySpace is characterised by higher content openly associated with criminal behaviours, and to more explicit and active conversation between users (Morselli & Décary-Hétu, 2013; Décary-Hétu & Morselli, 2011).

Showing gang colours and the symbols of group membership is also a frequent content. The pages share portraits of members with gang colours and symbols, participating into violent scenes, often carrying guns. They also report the “history of the gang, the achievements of its most prominent members, and an estimation of its size and power” (Morselli & Décary-Hétu, 2013, p. 163). This aims at building a powerful portrait of the gang. Gangs’ reputation is a central concept, and social networking sites fit the need of enhancing reputation among a wide public, better than traditional media channels. Pyrooz, Decker, & Moule (2015, p. 49) underline how gangs “use the internet much like an electronic graffiti wall”.

The online content also aims at legitimising the gang. For example, regarding the Hells Angels (Morselli & Décary-Hétu, 2013) the big population of supporters emphasised the legitimacy of their organisation. They also launched solidarity campaigns for their members, whose arrest was shown as an injustice.

Another reported activity is the emerging tendency to announce meetings and parties to the public, which are followed by posts, pictures and video documenting the event. This has occasionally been documented also in the case of drug trafficking organisations (Womer & Bunker, 2010).
Both studies remark the promotional intent of text and multimedia sharing. The videos “praise the ‘gangster way of life’” (Décary-Hétu & Morselli, 2011, p. 884). Guns, wealth, drugs, women, alcohol, and parties are the main recurrent themes. Content is highly sexualised, it shows violence and wealth, and it suggests criminal activities (Morselli & Décary-Hétu, 2013; Décary-Hétu & Morselli, 2011).

Nevertheless, the authors do not label this narrative building and criminal culture sharing as propaganda. They underline how they “tend to promote a certain way of life rather than promoting the criminal organization themselves.” (Décary-Hétu & Morselli, 2011, p. 886) Nevertheless, this type of content “suggest what awaits individuals should they decide to join.” (Décary-Hétu & Morselli, 2011, p. 884) The general tendency towards a growing space given to promotional content document “the influence and popularity of gangs among the general population and for a more limited subset of youth worldwide.” (Décary-Hétu & Morselli, 2011, p. 885).

The online presence fulfils “instrumental and symbolic needs of gangs”, but could also be “a luxury rather than a need” due to the threat from law enforcement actions associated with online visibility (Moule et al., 2014, pp. 1198–9).

### 2.6.2.1.3 Limitations

The literature presented has some limitations. The studies based on the individuals’ online profiles in many cases fail to clearly proof the effective affiliation of users to a criminal organisation, especially for individuals sharing generic contents. In some cases this might depend on the members’ awareness of the law enforcement threats associated with sharing identifiable contents (Décary-Hétu & Morselli, 2011). The studies analysing the official and/or fan pages face the same limitations. As many organisations benefit from a very large number of followers or friends, only a small percentage are actual members.

However, carrying out observational research online also presents advantages:

> “First of all it is easy to access and observe cyber-hotspots in the surface web. Second, it is not necessary to be online at the time when posting and discussion takes place. Third, useful research can be carried out without the need to interact with the group, therefore running the risk of influencing their behavior. Finally, it involves fewer ethical issues, given that participants in online activities are aware that their postings and conversations are being watched by people who do not reveal themselves, potentially by law enforcement officers.” (Lavorgna, 2015b, p. 360)

Concerning the results of the systematic review, the combination of literature based on multiple research approaches (secondary analysis of web contents, non-participant observation in cyberspaces, and semi-structured interviews) mitigated these limitations.
2.6.2.2 **RECRUITMENT AND NETWORKING**

Online recruitment is a controversial topic, characterised by a variable spectrum of conclusions depending on the definition of recruitment adopted. While literature on propaganda has already been presented, this section addresses the direct affiliation of new members by other members of the criminal network.

Two distinct patterns emerge from literature depending on the type of criminal network analysed: while direct online recruitment would play a marginal role among traditional OCGs, some instances of direct online recruitment emerge in the case of cybercriminal networks.

2.6.2.2.1 Online recruitment into traditional organised crime groups

All the authors agree that the dissemination of a criminal culture and “indirect propaganda” only play facilitating role in the recruitment of new members and affiliates among OCGs such as street gangs and DTOs. As previously highlighted “having access to such a wide pool of curious onlookers does facilitate any recruitment process that may follow beyond the sphere of the Internet” (Morselli & Décary-Hétu, 2013, p. 166) Law enforcement agencies could have played a role in the overestimation of online recruitment by interpreting the self-glorification of gangs and the spread of gangster culture as proper recruitment (Womer & Bunker, 2010). Nevertheless, offline social relations remain a central pillar of the recruitment process.

Regarding DTOs, Dávila (2016) notes that they are not actively using social media networks for recruitment. Rather the recruitment takes place through personal acquaintances due to the involvement of entire families and groups of friends in these organisations.

Regarding street gangs, poor evidence of direct online recruitment emerges. Décay-Hétu and Morselli (2011) report the presence of these groups online beyond the anecdotal evidence from newspapers, but no proactive recruitment strategies appear. The authors state that social networking sites are providing a new venue for people who share the same street gang lifestyle to come together. Only on the MySpace platform there was a limited set of content explicitly addressing the advantages of becoming a gang member. In their subsequent research, Morselli and Décay-Hétu (2013, p. 152) confirm these findings, and add that “[i]n regard to the visitors to such sites, there is no evidence that they are being tricked or manipulated in any way”, and that the promotion of a general gang street culture remains the main interpretation key of the shared content.

Partially supporting evidence for recruitment in cyberspace comes from the interviews conducted by Pyrooz et al. (2015), in which a low percentage (about the 8%) of the respondents reported that their gang recruited new members online “with current gang members more likely to answer yes, and older respondents and those further removed from the gang more likely to answer no.” (p. 490) This could at least suggest that recruitment happens through secret channels, which are not detected by research from open sources (Pyrooz et al., 2015). However, no specific definition of recruitment is provided by the authors, leading to uncertainty on the meaning attributed to the concept.
In conclusion, there is poor evidence of online strategies for targeted recruitment into traditional organised crime groups, at least in the visible cyberspaces. A more flexible interpretation of the concept of recruitment could better fit the complex relational opportunities enabled by the online cyberspace. According to Kleemans and Van de Bunt (1999), individuals become involved in criminal networks thought a “snowball effect” enabled by their social networks of contacts and relational opportunities. According to this theoretical framework, social networking platforms would enlarge the relational opportunities of individuals involved and fascinated by criminal subcultures.

2.6.2.2.2 Online recruitment into cybercriminal networks

Online cyberspace plays a more direct role in the case of recruitment into cybercriminal networks. As in the case of traditional organised crime groups, the concept of recruitment should be intended as a set of relational opportunities leading different individuals to converge in criminal groups. However, in the context of cybercriminal networks there are also cases of explicit online recruitment of affiliates and service providers for specific purposes.

European cybercriminal networks

A set of recent studies explore the Europe-based cybercriminal networks, all of them applying mixed methodologies relying on a variety of data sources: (a) online open sources, mainly digital forums; and (b) other traditional sources, as judicial cases, police files on criminal investigations, and interviews with key experts and law enforcement agents. These studies sometimes rely on overlapping data, and apply similar analytical methods and theoretical frameworks for testing different research questions. They focus on Dutch cases (Bijlenga & Kleemans, 2017; Leukfeldt, Kleemans, & Stol, 2017a, 2016; Leukfeldt, 2014), or complement them with insights form Germany, the U.K., and the U.S. (Leukfeldt, Kleemans, & Stol, 2017b, 2017c; Leukfeldt, Lavorgna, & Kleemans, 2016). The criminal networks analysed are involved in phishing and malware attacks.

The most common channels of recruitment into cybercriminal networks are social ties and online forums, with differences depending on the structure of the networks analysed and the role of their members. Cybercriminal networks have variable numbers of affiliates, from couple of units to more than 20 individuals. The ICTs-related skills vary depending on the single case, the structure of the network, and the position role. Despite the variations in their social organisation, some common structural characteristics emerge, which also influence the exploited channels of recruitment.

A framework of role positions consists of four typologies (Leukfeldt et al., 2017a, 2017b, 2017c; Leukfeldt, Lavorgna, et al., 2016; Leukfeldt, Kleemans, et al., 2016):

1) **Core members**, representing the core of the networks structure, without whom the crimes analysed would not take place. They represent the “social capital” of the networks with organisational and coordination functions, counting on their skills, their (variable) ICTs knowledge, and
their networks of contacts. The number of core members can vary from very low to more enlarge groups. In some cases there is a clear work division, while in other cases they share the organisational functions.

2) **Professional enablers**, composed of qualified providers of services, as the falsification of documents and the selling of malwares. They can be involved in the social network structure, or they could have a passive function of service providers.

3) **Recruited enablers**, who in many cases have strategic job positions and personal skills, as cashers, employees in banks, call centres, or postal offices. In some of the cases they declared to have been encouraged or forced by the core members to provide operational services.

4) **Money mules**, “passive” actors allowing the core members to use their bank accounts as transit money deposits in exchange of small amounts of money. The core members use the money mules to avoid being directly connected to the fraudulent transactions.

Social ties and the snowball effect play a central role in the recruitment of affiliates and enablers. As a general pattern, the “recruitment”, or better said, the co-offending of core members in a majority of the cases remained rooted in their offline social networks, synthesised in the formula “friends-of-friends”. Core members knew each other from the same neighbourhood, school, local or ethnic community, they could be long lasting friends or even lovers (Leukfeldt et al., 2017a, 2017c; Leukfeldt, Kleemans, et al., 2016; Leukfeldt, Lavorgna, et al., 2016). In other cases, they had overlapping criminal careers, being embedded in the same offline criminal circuits, former co-offenders in other illegal operations, or serving a sentence in the same prison (Leukfeldt et al., 2017a, 2017b, 2017c; Leukfeldt, Lavorgna, et al., 2016; Leukfeldt, Kleemans, et al., 2016; Leukfeldt, 2014). Social ties also play an important role in the recruitment of enablers, both professional and recruited. There are often references to the snowball effect “a friend knew a friend” who could provide malwares or fake documents (Leukfeldt et al., 2017a; Leukfeldt, Lavorgna, et al., 2016). Similarly, the bank employees or the call centre employees experts in the telephone calls, in charge of obtaining sensitive consumers’ information, could come to the same neighbourhood, or could be reached through common social ties (Leukfeldt et al., 2017a; Leukfeldt, Kleemans, et al., 2016). Bank employees were “approached by friends or acquaintances on the street” (Leukfeldt, 2014, p. 238). Work relationships were also particularly stressed as a key issue, especially in the case of the recruitment of ICTs experts and strategic-positioned employees (Bijlenga & Kleemans, 2017). Specialised individuals are needed in the case of cybercrime attacks, which require both ICT skills and social skills, as in the case of phone callers. Finally, the recruitment of money mules relies on the snowball effect. Core members or recruiters reached individuals from the same relational communities - vague acquaintances, friends-of-friends, etc. (Leukfeldt, 2014; Leukfeldt et al., 2017a, 2017b). In some cases, the money mules themselves offered their services (Leukfeldt, Kleemans, et al., 2016). They also prompted the
emergence of the figure of the recruiter as a hub of established networks of known money mules, in which new members were introduced through a snowball effect.

Social ties are the primarily channels of recruitment and co-offending. This is in line with the theoretical framework developed by Kleemans and Van de Bunt (1999). However, the online channel represented by digital forums “play an important role as digital offender convergence settings.” (Leukfeldt, Kleemans, et al., 2016) These forums “serve as a criminal marketplace where services are offered, thus functioning as ‘offender convergence settings’ where cybercriminals meet each other online.” (Leukfeldt, 2014, p. 232) A high number of partners with ICT skills could be potentially reached online, representative of a population of auto-selected individuals with ICTs skills. In this way “[f]orums enable a small group of core members to have a high impact.” (Leukfeldt, Kleemans, et al., 2016)

Bijlenga and Kleemans (2017) highlight how online meetings can also overcome some limitations of real world co-offending, as the need of gaining trust. Online criminal collaborations are not based on trust, which is substituted by an “online reputation” of anonymous individuals, enabling the collaboration among individuals on the basis of previous good experiences. In some cases, the collaboration can be limited to the exchange of information, or to the sale of malwares and stolen data. A second point is that social relationships are “highly clustered and may be restricted to people in the same region.” This could be a limitation in the case of cybercrime, which requires certain level of skills by offenders. Personal social relations could not be enough to fulfil the need of the social capital required by relatively complex criminal operations. Digital forums are again a good means to overcome these limitations, concentrating communities regardless physical and geographical borders. It is likely that the higher the level of technology involved is, the more contribution could come from digital ties. This element emerged from Leukfeldt, Kleemans and Stol (2016)’s analysis of Dutch criminal investigations.

In the literature examined, the online process of recruitment was correlated to the types of origin and growth of the networks analysed, and four types emerged:

1) Origin and growth completely on social contacts.

2) The core of the network is born and grows on social contacts, and forums contribute to the recruitment of specialised service providers.

3) Forums as the core base, and social contacts are used to recruit local criminals for operational purposes.

4) Origin and growth completely on forums.

In the case of (2) and (4), core members meet up on digital forums. They know each other from chat rooms, and they exchange technical information on cyber-related technologies. In other cases, the core of the network is composed of individuals who are “well known” by the cyber community for their skills, and only need very restricted external contribution from enablers and service providers. Case (2) is characterised by a core of the network based
on social ties, whose core members could have limited or no ICTs skills. They use forums for buying specific services, or for recruiting specialised enablers. In conclusion, social ties remain a central channel of recruitment and co-offending among European cybercriminal networks. However, online recruitment gains importance with respect to traditional organised crime groups, with a central role played by digital forums. Social ties still play a central role in online recruitment and co-offending. When individuals need a co-offender, an enabler, a money mule, the first channel explored is their own network. In this regard, digital forums represent a channel for expanding individuals’ social networks within ICTs specialists or cyber-punks, interested or likely to be involved in cybercrime. They have a “social function” of facilitating contacts and alliances (Leukfeldt et al., 2017c) in line with the theoretical framework of Kleemans and Van de Bunt (1999). Cybercriminal networks (Leukfeldt, 2014) differ from traditional organised crime members, who are seeking visibility on blogs and social networking sites. The use of cyberspace by cybercriminal networks has different drivers, and seeks different goals. Visibility is not needed, as operational and demand-supply drivers prevail as drivers for recruitment.

**Nigerian Yahooboyism**

Two studies included by the systematic review presented the case of Nigerian Yahooboyism cybercriminal networks. Adeniran (2008) carried out survey research on 400 youths (age 15-35) in the city of Lagos, and online participant observation on open sources. Tade and Aliyu (2011) carried out their research thought semi-structured interviews and focus groups on yahoo-boys in the University of Ibadan. “Yahooboyism” is an urban sub-culture of Nigerian youths. It was born within a context of poverty and unemployment, characterised by a “culture of fraud and corruption” (Adeniran, 2008, p. 369). Within this context young people embraced the ICTs inventions and spent a large part of their days online (Adeniran, 2008, p. 368). The weak legal system and the large diffusion of ICTs technologies led to the emergence of cybercriminal networks involved in online frauds. Yahooboyism represents a particular case highly connected to its originating geopolitical context. Unemployment is a key driver of the involvement in cybercrimes (Adeniran, 2008). These youths show elements of proud and justify their activities with the unjust social system (Adeniran, 2008) Cybercrime activities are seen as a “creative innovation linked to survival to cope with economic insolvency.” (Tade & Aliyu, 2011, p. 869) This parallels with the analysed Mexican drug cartels: the lack of socioeconomic development leads the population to find illegal channels of self-sustenance, and Yahooboy are socially recognised by their friends and age peers. According to Tade and Aliyu (2011) despite the digital nature of the crimes, real-world social ties represent the primary channel leading to co-offending and “recruitment”. Co-yahoo comrades, but also family ties, friendship, and love relationships are the core social bonds enabling people to become involved in cybercriminal networks. Real-world relationships remain central to the network structure, and co-offenders meet up in physical world to self-organise and carry out cyber-attacks.
Similarly to the European cybercriminal network cases, bank employees would facilitate transactions, and obtain sensitive customers’ information. It is likely that the employees themselves are central peer partners of the operations. In addition, insiders from the police can have a facilitating role, ensuring the release of convicted Yahooboys.

2.6.2.3 Perpetration of computer crimes

Computer crimes could refer to several types of crimes enabled or facilitated by ICTs. Some of them require higher human involvement, and ICTs only act as a facilitator. This is the case of online frauds where a person has to persuade bank customers’ to provide their sensitive data. In other cases, individuals play a more marginal role, as in the case of malware attacks to bank websites.

2.6.2.3.1 Traditional organised crime groups’ involvement in computer crimes

Despite the growing concern on the potential involvement of OCGs in cybercrime, the systematic review failed to identify any study specifically addressing this process. The only reference to OCGs and families comes from Leukfeldt, Lavorgna and Kleemans (2016)’s research on cybercriminal networks in The Netherlands, the UK, the USA, and Germany. Five out of forty analysed cybercriminal networks involved OC affiliates. They were known to the police for their active criminal careers in a range of offline illicit activities such as fraud, drug trafficking, money laundering, and racketeering. In one case, they bought malware to carry out online fraud. In another case, they were service providers to cybercriminal groups, supplying fake documents and money laundering services. A motorcycle club was the provider of money mules, recruited in the physical world. The collaboration between OCGs and cybercriminal groups relied on opportunistic drivers in all the cases reported by the police (Leukfeldt, Lavorgna, et al., 2016). This suggests that even if OC affiliates entered the market with varying levels of continuity and investment, computer-focused crimes only represent one of the many criminal areas in which these networks of individuals commonly operate. Computer-focused crimes could be described as modernisation of the criminal opportunities, in line with the evolution of the modern society.

2.6.2.3.2 Cybercriminal networks and computer crimes

Cybercriminal networks build up on computer-focused crimes. Several studies focused on Dutch cases (Bijlenga & Kleemans, 2017; Leukfeldt et al., 2017a; Leukfeldt, Kleemans, et al., 2016; Leukfeldt, 2014); a few complemented the information on cases from the Netherlands with insights form Germany, the U.K., and the U.S. (Leukfeldt et al., 2017b, 2017c; Leukfeldt, Lavorgna, et al., 2016). These studies analyse the modus operandi and the crime scripts of the most common cybercrimes. Leukfeldt, Kleemans, and Stol (2017b, p. 42) present a generalizable framework for describing the cybercriminal networks. Depending on the computer-focused crime carried out, cybercriminal network range between two poles:

1. Locally rooted networks carrying out low-tech attacks with a high degree of direct offender-victim interaction.
2) International networks carrying out high-tech attacks without such interaction.

In both high-tech and low-tech attacks, phishing is the most reported fraud, with differences in the crime script. The attacks usually start from fake emails aimed at obtaining individuals’ sensitive information.

In the case of low-tech attacks, a phone caller obtains the transaction codes. Once the victim provides the information necessary for monetary transactions, money is transferred to the bank accounts of money mules. Other functions requiring human involvement could also be needed, as postal employees intercepting the bank mail correspondence and stealing credit cards and pin codes. Bank employees could be involved in supplying sensitive data, or increasing the withdrawal limits. Phone companies’ employees also cooperate by providing phone SIM cards copies of existing phone numbers. In these cases, networks are likely to be more extended, with core members coordinating the operations, and involving service providers and lower ranked members.

In the case of high-tech attacks, offenders use malware, which is sent to the victims by various channels and enables criminals to have direct access to their computers. In other cases, fake bank websites are launched to intercept the transactions made by the bank customers. Since higher technical skills are required, smaller groups of core members with higher ICTs skills are more likely to be involved in these cybercrimes.

Similar patterns emerged in the case of Nigerian Yahoooboy (see Section Online recruitment into cybercriminal networks). According to Tade and Aliyu (2011), the networks of Yahoooboy present varying levels of specialisation. Some of them specialise in phishing emails from fake bank accounts to obtain consumers’ sensitive private information. Unemployment is also exploited by posting fake job announcements and asking money from people. Another widespread fraud consists of persuading people to donate money by pretending to be a person in need. Once a particular technique becomes obsolete within a country, others are targeted.

2.6.2.4 Traditional offline activities facilitated by online means

The exploitation of cyberspace as a facilitator of the perpetration of traditional organised criminal offences is a growing concern for law enforcement. Traditional organised criminal offences include a range of illegal activities, as illicit trafficking of drugs, humans, firearms, wildlife, etc. These activities fall within the concept of transit crimes: “criminal groups are primarily involved in international illegal trade, using the same opportunity structure that facilitates legal economic activities.” (E. Kleemans, 2007, p. 176)

The emergence of ICTs had an impact on licit trades, and illicit markets could benefit from the same facilitators. The Internet has a potential “transformative impact” on traditional trafficking activities (Lavorgna, 2015c, p. 366). Online drug markets emerged, and therefore they caught the attention of researchers and enforcement agents.
Transit crimes comprise a variety of illicit activities. Some of them are clearly illegal and are traditionally linked to established organised crime groups: e.g. “hard drugs”, as cocaine and heroin. Other trades overlap with their corresponding legal or regulated markets. Differences among national legislations and loopholes in regulations create a wide “grey area” of legality, which complicates the detection of illicit markets. Despite limitations, organised groups’ activities in cyberspace are a recent phenomenon, which is likely to follow the evolution of the use of ICTs by contemporary society. The exploration of new online information sources provides research opportunities for the study of crime, and new tools and methodologies are being explored and refined (Lavorgna, 2015c; Broadhurst, Grabosky, Alazab, & Chon, 2014).

Studies relying the open sources have limitations. They suffer from rates of “rumour” of content which might indicate illegal activities, but could hardly be proven illegal. According to Lavorgna (2015a, p. 158) “the Internet seems to be boosting certain trafficking flows more than others and, in particular, criminal activities that can more easily play with gaps in legislation or that are usually perceived as ‘less serious’.” This is the case of “soft drugs”, as marijuana, opioids, and psychoactive substances, but also human trafficking, wildlife trafficking, counterfeit medicines, and online gambling websites.

For the purpose of this systematic review, only literature covering two areas of interest is included studies directly addressing traditional OCGs as defined in Section 2.3.2, and studies explicitly targeting the social organisation of transit crimes.

2.6.2.4.1 Drug trafficking online

Online drug trafficking is attracting most of the research efforts. Online drug cryptomarkets are the visible part of the internet-mediated trafficking, directly matching the supply and the demand. Research provided quite comprehensive assessments of the functioning of online drug cryptomarkets. The systematic review focuses only on studies on the involvement of OCGs in online drug trafficking and on the social organisation of online cryptomarkets.

There is a general agreement on the existence of various levels of organization in distribution networks. Broseus et al. (2014) trace the connection between different sellers across different cryptomarkets in Canada through the analysis of the usernames and by monitoring the information provided by the sellers themselves. The relevant hubs of the networks comprise vendors with a diversification of the drug supply, capable of handling worldwide shipping, and vendors highly specialised on one specific drug. The results show the existence of well-structured and organised online networks, which may manage different vendor accounts, operate of several cryptomarkets, and trade several products.

Dolliver (2015) analysed Silk Road 2, a “sequel” of the famous Silk Road. The study examined potential connections between the distribution networks and upper-level retail markets, including drug trafficking organisations with established production and distribution chain. The results failed to uncover a connection between the distribution networks and the established upper-level retail markets. Only occasionally sellers acquired drugs at a prior stage and turned them into retail products (Dolliver, 2015, p. 1121). Similarly, Bijlenga
and Kleemans (2017) reported anecdotal evidence of an online drug seller benefiting from daily supply of drugs. While some organisation exists, most collaborations would rely on opportunistic partnership driven by supply and demand. These are unlikely to be formally organised distribution chains. In some cases, online drug trafficking accounts for only a part of the illegal revenues of the involved actors (Bijlenga & Kleemans, 2017; Leukfeldt, Kleemans, et al., 2016; Leukfeldt et al., 2017b, 2017c; Leukfeldt, 2014). The role of opportunities is important also for offenders affiliated to traditional organised crime groups (Leukfeldt, Lavorgna, et al., 2016). Whereas, they enter the market with varying levels of continuity and investment, the online drug trade only represents one of their many criminal ventures. The research by Lavorgna (2014a) supports this argument. The Internet would have not affected the opportunity structure of the “hard drugs” market. The online grey market of “soft drugs” (e.g., cannabis, NPSs, and synthetic drugs), instead, would host new middle-level criminal networks, managing both international trades and local distribution (Lavorgna, 2014a, p. 266).

There is a general agreement on the irrelevance of the online drug distribution channels for traditional DTOs. This is consistent with the evidence on the use of social networking sites by members of DTOs, which connect these sites to individual and symbolic needs (Dávila, 2016; Nix et al., 2016; Womer & Bunker, 2010). The systematic review could not find any evidence on online drug trafficking by DTOs. The scarcity of “hard drugs”, traditionally traded by established DTOs, corroborates this possibility.

In conclusion, the online trade of drugs is an evidently growing trend. However, the drug trade remains an offline market. The whole supply chain (from the production to the final distribution) relies on consolidated practices, modi operandi, and social relations. The online distribution networks have various levels of organisation and build up on opportunistic demand-supply drivers. The opportunities offered by online cyberspaces better fit less regulated niches (as “soft drugs”, NPSs, and synthetic drugs), where a number of new actors can exploit grey areas of quasi-legality.

2.6.2.4.2 Other online-facilitated transit crimes
The systematic review included additional literature on wildlife trafficking and human trafficking online.
Two studies by Lavorgna (2015b, 2014b) analyse the network structure behind the online wildlife trafficking. The author highlighted how the characteristics of the market would fit the opportunities provided by online cyberspace, facilitating the transit through local intermediaries in the destination countries and the distribution stage (Lavorgna, 2014b). Online advertisements, forums, selling platforms and consumers’ groups allow for promotional campaigns of rare and precious flora and fauna species. Unlike the dark web drug cryptomarkets, the visible “surfing web” plays a main role for reaching a wide public. Four types of offenders emerged with varied levels of organisational complexity, including individual offenders, peer partners, loose networks, and structured networks (Lavorgna, 2015b). The comparison of online wildlife and pet trafficking shows that the nature of the animals trafficked influences the
characteristics of the trafficking networks: the lower level of organisational effort, the lower the structured of the networks involved. For the trafficking of human beings, cyberspace would play a facilitating role mainly at the stage of recruitment of victims (Lavorgna, 2015a). This is an element of originality, is in contrast with other transit crimes which are mostly affected during later stages. The exploitation of online spaces would provide several opportunities for sex trafficking, including victim selection, victim subjection, and deceptive opportunities (Lavorgna, 2015a). Some examples of recruiting channels are fake job announces, fake agencies of dating, marriage, or children adoption (Leng, Khan, & Rahim, 2014). Ibanez and Suthers (2014) relied on these recruiting channels online to analyse the organisation of the US human trafficking market. The network identified through online source was comparable to established circuits along the US federal states. This exploratory study has some limitations. The authors specify that the presence of sex trafficking indicators does not mean trafficking is occurring. However, the comparability with results from other sources highlights the potential exploitation of cyberspace by human trafficking facilitators (Ibanez & Suthers, 2014).

2.6.3 METHODOLOGIES AND SOFTWARE
This systematic review initially identified a number of studies (n=43) focusing on methodological/technical approaches to the study of organised crime and terrorist networks online. This literature failed to meet the inclusion criteria because the discussed methodologies were not applied on actual data, or were tested with the only objective of validating the methodology itself, and could not produce generalizable results. Nevertheless, the systematic review screened these works to provide better understanding of possible promising methodologies.

The forty-three studies present methodologies for the monitoring of information on open sources, mainly public social networking sites (n=23), websites on the dark/deep web (n=10), forums (n=6), and blogs (n=1). Most studies focus on terrorism (n=37), whereas only few on criminal groups (n=6). The proposed methods, however, might often fit both the study of terrorism and criminal groups.

The identified methodologies and software had different purposes. The majority of them aim at analysing the typology of contents published by criminal and terrorist groups, as their share of extremist and violent contents (n=16); some tools focus on the identification of actual offenders or radical members (n=12); other tools aim at reconstructing the social organisation and social networking structure of the groups/networks analysed (n=12); also tools explicitly aiming at predicting terrorist attacks through the analysis of type/frequency of targeted extremist contents shared were identified (n=3). Fehler! Verweisquelle konnte nicht gefunden werden. provides a full list of references of tools.
2.7 Discussion and conclusions

2.7.1 Online activities of terrorist and organised crime networks

2.7.1.1 Terrorist networks

The majority of studies on online activities of terrorist networks focus on communication, social networking, propaganda, recruitment and mobilisation. The reviewed studies suggest that the degree of the online intergroup communication depends on offline factors. These include the sharing of common ideology and the openness of the cultural environment to extremist discourses (Al-Saggaf, 2016; Caiani & Parenti, 2013). Online and offline communications are strongly correlated, and each type of interaction reinforces the probability of the other (Gill & Corner, 2015).

Terrorist networks actively use their official websites, forums and social media platforms to disseminate their messages. The internet offered an opportunity to construct their own media image and counteract the rhetoric of traditional media (Ismail, 2008; Janbek, 2009). The propaganda targeting the society at the global level aims to legitimise the use of violence, engage wider audiences with the radical ideology, spread fear and intimidate the enemy (Awan, 2017; Tsfati & Weimann, 2002). The propaganda towards potential supporters constructs a collective identity that motivates terrorism violence and generates further support (Drissel, 2015; Erez et al., 2011).

The propaganda narratives of oppression and injustice serve as effective recruitment messages. Whereas not all extremist and terrorist websites contain direct calls to join the network, most of them publish mobilising messages and information about joint events and activities (Caiani & Parenti, 2013; Davies et al., 2015). Social media platforms opened possibilities for bottom-up radicalisation of young individuals (Conway & McInerney, 2008). Forum archives show that the active participation of a supporter in online activities could be an initial step to involvement in jihadi terrorist networks (Torres-Soriano, 2016b).

The included studies provide limited empirical evidence about the use of the Internet for training and planning of attacks, despite frequent allegations by popular media and anecdotal literature. The reviewed literature suggests that real-world training and low-tech communication techniques still play a key role in the planning and execution of terrorist attacks (Drozdova & Samoilov, 2010; Gill et al., 2017; Stenersen, 2008). Online training efforts of terrorist networks are limited to publication of training materials (Stenersen, 2008). Their content is providing basic and outdated information, indicating that online training could be a preparatory step (Stenersen, 2008).
2.7.1.2 ORGANISED CRIME NETWORKS

The majority of studies of traditional organised crime groups focus on the online visibility of groups and affiliates, the web content they share, and the dissemination of criminal cultures. These studies examine the web content and social media image of street gangs (Pyrooz, Decker, & Moule, 2015; Moule et al., 2014; Morselli & Décary-Hétu, 2013; Hellemont, 2012; Décary-Hétu & Morselli, 2011) and DTOs (Dávila, 2016; Nix et al., 2016; Womer & Bunker, 2010). The literature adopts two different research approaches: the study of individuals’ personal profiles likely to be members of organised crime groups; and the study of “official” web pages, visited by individuals with varied levels of involvement in illicit organisations, but also by supporters and followers. The studies analyse the use of social networking sites for building self-glorifying portraits, showing strength, wealth, and social status. Social networks are described as “electronic graffiti wall” (Pyrooz, Decker, & Moule, 2015, p. 49). DTOs are more likely to publish content directly showing illegal activities. DTOs attempt to legitimise their illicit activities through narratives on social and economic injustice. No study identified any direct reference to recruitment. The Internet is a facilitator, spreading criminal culture and presenting members of criminal organisations as “aspirational role models” (Dávila, 2016).

The majority of studies of cybercriminal networks focus on the crime script followed for carrying out cybercrimes, and on methods for recruiting members. Most of the studies describe the mechanisms for recruitment and crime scripts of cybercriminal networks. They focus on European cybercriminal networks (Bijlenga & Kleemans, 2017; Leukfeldt et al., 2017b, 2017c, 2017a; Leukfeldt, Lavorgna, et al., 2016; Leukfeldt, Kleemans, et al., 2016; Leukfeldt, 2014) and on the Nigerian Yahooboys (Tade & Aliyu, 2011; Adeniran, 2008). The findings suggest that offline social relations and opportunities are still the leading factors explaining the involvement in criminal groups. In the case of European cybercrime networks, digital forums facilitate recruitment, especially of specific service providers, money mules and other crime enablers. Digital forums are also a channel for expanding individuals’ social networks within auto-selected populations of individuals interested in computer hacking. Most of the literature on transit crimes focuses on drug trafficking. However, only few studies examine their social organisation. The amount of drugs trafficked online is rarely compared to the overall size of the illicit markets. Online drug cryptomarkets are the visible part of the internet-mediated trafficking. While there is literature on their market characteristics, only a few studies on their social organisation could be identified (Dolliver, 2015; Broseus et al., 2014; Lavorgna, 2014a). The lack of publications comparing the online and offline market further limits the evaluation of the role of the Internet in drug trafficking.

The systematic review identified some exploratory analyses of other transit crimes: wildlife trafficking (Lavorgna, 2015a, 2014b) and human trafficking (Lavorgna, 2015a; Ibanez & Suthers, 2014; Olivia Tan Swee Leng et al., 2014b). However, they have limitations, and their descriptive results could not
be generalised. There is a lack of research addressing the social organisation of these crimes.

2.7.2 POPULAR METHODOLOGIES IN RESEARCHING ONLINE ACTIVITIES

2.7.2.1 TERRORIST NETWORKS

The social network analysis is the most popular quantitative method (n=13). Almost all of the studies applying network analysis examine the intra- and intergroup organisation among extremist organisations. Five of the selected studies draw conclusions on the basis of descriptive statistics and four on the basis of inferential statistical methods. Three of the publications use more advanced techniques, including computational models, machine learning tools and sentiment analysis. Almost all qualitative and mixed-method studies carry out structured content analysis. Almost all studies (90%) rely on open source data from different Internet platforms. The sample sizes depend on goals of the study, the type of online source and the possibilities for extracting the information. Only three studies interviewed experts and law enforcement representatives. Overall, open source data could provide significant information not only on the online activities of terrorist networks, but also on their members, modus operandi, rhetoric and intentions. Primary data, such as investigation files and interviews with convicted terrorists could complement the findings from open sources (Torres-Soriano, 2016b; Klausen et al., 2012).

The building of datasets of terrorism cases on the basis of government files and open sources could provide insights into the online activities of terrorist networks. Two included studies applied this approach and provided quantitative data on almost all online activities of terrorists (Gill et al., 2017; Gill & Corner, 2015). The database of Gill et al. (2017) allowed the examination of correlates of terrorists’ decisions to use the Internet. It overcame the limitations of open source research, which cannot prove the affiliation of online radicals and determine the extent of online learning in planning of attacks. To ensure the reliability of their results Gill et al. (2017) used various data sources and applied strict coding procedure.

2.7.2.2 ORGANISED CRIME NETWORKS

Most of the studies on organised crime groups on social media networks rely on primary web data from open sources, and apply mixed qualitative and quantitative analytical methods. Automatic information retrieval software is often used for monitoring blogs, social networking sites and other open sources within a selected time period (Dávila, 2016; Nix et al., 2016; Morselli & Décary-Hétu, 2013; Hellemont, 2012; Décary-Hétu & Morselli, 2011; Womer & Bunker, 2010). This technique could be compared to non-participant observation (Lavorgna, 2015b). Descriptive content analysis is the most common technique, applied to text and multimedia web content for describing the narratives of gangs and DTOs. Statistical findings have the same descriptive objective. One study applies
social network analysis to the Facebook Profiles of (potential) DTO-affiliated
individuals (Nix et al., 2016).
Open sources have advantages, such as the easy access to data and the non-
interaction with the observed group (Lavorgna, 2015b). Their limitations
include the difficulty of proving the affiliation of profiles/pages to a criminal
organisation. Integrating other data sources could mitigate these limitations
and provide insights into the links with the offline activities.
The literature on cybercriminal networks relies on traditional data sources.
Studies on European networks apply secondary text analysis to police files,
judicial records and interviews with public prosecutors, experts, and police
officers (Bijlenga & Kleemans, 2017; Leukfeldt et al., 2017c, 2017b, 2017a;
Leukfeldt, Lavorgna, et al., 2016; Leukfeldt, Kleemans, et al., 2016; Leukfeldt,
2014). Descriptive content analysis of case studies is the preferred analytical
approach. The specific case of Nigerian Yahooboys relies on primary sources
such as semi-structured interviews, participant observation, and focus groups
(Adeniran, 2008; Tade & Aliyu, 2011).
These traditional “offline” sources have provided insights into cybercriminal
networks. Due to the nature of the research topic, the use of open sources
could be used as a complementary source of information, as in the case of the
recruitment on digital forums of members and enablers of European
cybercriminal networks.
The literature on the social organisation of transit crimes is scarce, and shows
a variety of exploratory methodologies applying mixed method approaches to
information available online. Quantitative analytical approaches prevail in the
case of online drug trafficking, enabled by the nature of online cryptomarkets
(Dolliver, 2015; Broseus et al., 2014). One experimental mixed methodology
has been applied to online prostitution-related advertisements to reconstruct
the US human trafficking markets (Ibanez & Suthers, 2014). The remaining
literature relies on traditional data sources, i.e. interviews and case studies
(Lavorgna, 2015a, 2015b, 2014a, 2014b; Olivia Tan Swee Leng et al., 2014b).

2.7.3 SIMILARITIES AND DIFFERENCES BETWEEN
TERRORIST AND CRIMINAL NETWORKS
The different goals of organised crime groups and terrorist networks determine
their distinctive uses of ICTs. Academic research reflects these differences in
the choice of its topics. Online activities of terrorist networks have attracted
more research attention than organised criminal groups’ ICTs use. A possible
explanation is the greater visibility of terrorist networks, who focus on the
online dissemination of their propaganda. The lower visibility of organised
crime groups’ online activities does not necessarily imply a lower exploitation
of the opportunities offered by the cyberspace. In fact, it is a likely result of
their different goals in using ICT. This is why some authors have focused on
how ICTs have facilitated the commission of traditional criminal activities or
the extent to which computer crimes’ perpetration is organised. The research
efforts on terrorism have focused on online propaganda, recruitment and
internal and intergroup communication (Tsfati & Weimann, 2002; Torres-
Soriano, 2016a).
While pursuing different goals through ICTs, both organised crime groups and terrorist networks construct a public image through social media platforms. Both groups use social media networks to counteract the public image projected by authorities and traditional media. Terrorist networks construct their own image by legitimising their activities and focusing on their ideology (Tsafati & Weimann, 2002; Erez et al., 2011). Gangs publish contents that glorifies and portrays individuals’ criminal style and culture, and spread symbols of pride for the organisation.

The Internet serves as a facilitating tool for the recruitment of both organised criminal groups and terrorist networks. Nevertheless, its role should not be overestimated, as studies show that offline interactions and social ties still play a decisive role in the recruitment process. Examination of terrorism cases shows that the radicalisation process includes both virtual and physical communication of like-minded individuals (von Behr et al., 2013; Gill & Corner, 2015). While forums provide the required connections with terrorist networks, discussions on involvement in their activities happen mainly in face-to-face interactions (Stenersen, 2008). Similarly, research on organised crime groups suggests that social networking platforms facilitate recruitment through enlarging the relational opportunities of individuals attracted to criminal subcultures (Pyrooz, Decker, & Moule, 2015).

There is scarce empirical knowledge on the commission of computer crimes by terrorists and organised crime groups. The systematic review failed to identify any empirical study on the use ICTs for e.g. financing and money laundering.

2.7.4 POLICY IMPLICATIONS

The studies on online activities of OCTNs discuss various issues and differ in their relevance to counter-organised crime and counterterrorism policies. Nevertheless, the results of the systematic review point out a few policy implications.

Terrorist networks’ online activities should be considered, analysed and targeted with regard to their offline strategies. While most authors consider the online platforms a force-multiplier for terrorists, empirical research shows that the offline activities of radicals still play an important role in communication and recruitment (Drozdova & Samoilov, 2010; Gill et al., 2017; Qin et al., 2007; von Behr et al., 2013). Drozdova and Samoilov (2010) conclude that hi-tech tools allowed greater efficiency on a large scale across distance, while offline communication provided greater secrecy. Therefore, the aim of counterterrorist authorities should be to target both online and offline communication channels (Drozdova & Samoilov, 2010). Von Behr et al. (2013) conclude that the Internet is only a mode for radicalisation, but not a driver on its own. Similarly, Gill et al. (2017) suggest that an over-concern with the online behaviour of terrorist networks may undermine the detection of key offline components of their strategies. The enforcement focus should be on the overall drivers for radicalisation and terrorist attacks rather than the environment in which the extremist behaviour is forming (Gill et al., 2017).

Online activities of terrorists could provide significant intelligence information to counterterrorism authorities. Authorities should balance between the need to shut down online terrorists’ platforms and the possibility to obtain
information through them. Online communication and propaganda may contain hidden information about terrorists’ organisation, activities, tactics and plans. Analysis of these hidden messages could intercept terrorists’ communication and help preventing terrorist attacks (Qin et al., 2007). Forums and social media platforms should be used exclusively for intelligence-gathering. Authorities could use occasional and temporal shut downs to demonstrate their power and show individuals the high risk of joining the movement (Zelin & Fellow, 2013). Pre-monitoring of the forums could help authorities decide on their strategy towards these platforms (Erez et al., 2011). The officers in charge of conducting online intelligence should receive sufficient training in digital literary skills (von Behr et al., 2013).

The increasing effort of academics to develop methodologies for studying of online terrorist and organised crime platforms could complement the intelligence and enforcement activities of national and international institutions. The publications in the “tools” subsection (see above subsection Fehler! Verweisquelle konnte nicht gefunden werden.) suggest methodologies to monitor and analyse terrorists’ and organised crime’s online activities (Johnson et al., 2016; Marcellino et al., 2017; Xu et al., 2006). Institutions could cooperate with academics to further develop these methods and complement them with primary data from investigations (von Behr et al., 2013).

Experts, policy makers and social media companies should investigate the effects of suspension of social media accounts promoting violence and terrorism on the online behaviour of OCTNs. Social media networks have started to suspend accounts for hate speech and dissemination of violent content (Woolf, 2016). They have noted the difficulties in identifying extremist content and in balancing between freedom of expression and the duty to suspend (Woolf, 2016). The systematic review has not identified any study assessing the impact of these measures. The joint efforts of academia, law enforcement institutions and social media companies could examine how these measures have affected the online activities of OCTNs and whether any valuable intelligence information has been lost.
3 Task 3.2: Innovative study - finding the Dark Web signposts

3.1 The experimental platform

The project partners involved in the technical work of PROTON, Fraunhofer and IBM, decided to design an IT system easily split along organizational lines: Fraunhofer created a data collection infrastructure to extract textual information from data sources, the PROCeeD system. The name PROCeeD is short for PROTON Cybercrime Detector. It is a HTML5 single page application developed with cutting-edge web technology for maximum scalability and usability. IBM focused on providing classification information on such textual data, also adding in additional, partially proprietary, partially public data sources. The common data exchange mechanism was a Postgres database with content from the different data sources, enriched by 52 personality traits. The primary experiments focused on:
- finding concrete Darknet pointers (i.e., addresses of .onion Darknet sites) in these social media data sets
- determining common machine-identifiably personality traits of specific personas in these data sets

3.1.1 SYSTEM ARCHITECTURE

The overall system architecture of PROCeeD is illustrated in Fehler! Verweisquelle konnte nicht gefunden werden.. PROCeeD is designed as a distributed system running on Web. It consists of two main functional blocks: the service server with the database and the collectors that crawls various kinds of information on the Web.

The backend service server consist of:
- a web server that provides multiple RESTful service endpoints and
- a database where relevant data like the configured collectors, the darknet shops etc. are stored.

On top of the managed data in the database, multiple in-database analytic methods have also been designed to provide cyber-ware analysis focusing on OCTN activities. For more information about the backend design, please see Section Fehler! Verweisquelle konnte nicht gefunden werden..
The collectors in the architecture of PROCeeD play a central role for OCTN-related activity analysis. The architecture provides a flexible plugin mechanism to handle multiple collectors/crawlers. At runtime, collectors can be configured via RESTful API and managed by the PROCeeD server. Collectors are responsible to retrieve certain kinds of web information like Twitter or YouTube. This can be a crawler by parsing the web pages of a forum or using APIs provided by websites to retrieve the relevant data. More details can be found in Section Fehler! Verweisquelle konnte nicht gefunden werden.

![PROCeeD Architecture](image)

**Figure 9:** The architecture of the PROCeeD system. It is a web application with two main blocks: the service server with the database and the collectors that crawl various kinds of information on the Web.

The following sub-sections cover more technical details about the design of the three major functional blocks in PROCeeD.

### 3.1.2 BACK-END DESIGN
The backend of PROCeeD consists of two components: the web server to provide RESTful web services and the database system maintaining all OCTN-relevant data.

The web server is implemented with Node.js\(^{10}\), a free and open source server-side JavaScript technology for building scalable web applications. For better software maintainability TypeScript \(^{11}\) is used for trans-compilation into JavaScript programs. The RESTful web services currently defined are summarised in **Fehler! Verweisquelle konnte nicht gefunden werden.**:

**Table 5 Summary of the PROCeeD RESTful web services, which are categorised by the associated HTTP operations like GET, PUT, etc.**

<table>
<thead>
<tr>
<th>HTTP Operation</th>
<th>Path</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GET</td>
<td>/shop</td>
<td>Retrieve all Darknet shops from the database.</td>
</tr>
<tr>
<td></td>
<td>/shop_category</td>
<td>Retrieve the defined Darknet shop categories from the database. Categories are drugs, weapon, illegal documents, etc.</td>
</tr>
<tr>
<td></td>
<td>/twitter/cfg</td>
<td>Retrieve the configuration of the Twitter collector from the database.</td>
</tr>
<tr>
<td></td>
<td>/collector</td>
<td>Retrieve all configured collector information from the database.</td>
</tr>
<tr>
<td>PUT</td>
<td>/shop</td>
<td>Update the information of an existing shop in the database.</td>
</tr>
<tr>
<td></td>
<td>/collector</td>
<td>Update the information of a defined collector. The information is provided as HTTP body.</td>
</tr>
<tr>
<td></td>
<td>/twitter/cfg</td>
<td>Update the configuration of the twitter collector.</td>
</tr>
<tr>
<td>POST</td>
<td>/shop</td>
<td>Create a new Darknet shop in the database. Information needed is name, category, URL, etc.</td>
</tr>
<tr>
<td></td>
<td>/youtube/data</td>
<td>Create a new YouTube entry in the database by providing the name of the video, the author, the recursive comments, etc.</td>
</tr>
</tbody>
</table>

The RESTful services provided by the PROCeeD server are designed to be used by other clients who want to access the data stored in the PROCeeD database. This provides an excellent solution for other work packages in the PROTON project to consume the OCTN-related analysis results for further advanced process.

The database in PROCeeD is a PostgreSQL\(^{12}\) instance with advanced JSON and spatial support. With sophisticated JSON support, the application logics can be simplified for better robustness. This follows the idea of “pushing logics closer to data” to accelerate the analysis. The overhead of transferring the data via

---

\(^{10}\) [https://nodejs.org](https://nodejs.org)

\(^{11}\) [https://www.typescriptlang.org](https://www.typescriptlang.org)

\(^{12}\) [https://www.postgresql.org](https://www.postgresql.org)
network can be saved for better performance. The current version of the database schema is illustrated in Fehler! Verweisquelle konnte nicht gefunden werden. and Fehler! Verweisquelle konnte nicht gefunden werden.. Fehler! Verweisquelle konnte nicht gefunden werden. is about the core table structure in the PROCeeD system including data that is general for all collectors and OCTN-related analysis like the shop category. Fehler! Verweisquelle konnte nicht gefunden werden. is a specific schema for the YouTube collector. It depends on the design of the collector to store the collected data – either in the database of PROCeeD or its own storage system like a file system. However for concurrency reasons, it is better to use a transactional system to manage the parallel access issues.

A specific design to achieve the concept of “push the logic closer to data” in the YouTube schema is the application of a raw data table as depicted in Fehler! Verweisquelle konnte nicht gefunden werden.. This table is only defined with three columns:
- id as the primary key
- the insert_at timestamp and
- a column raw with type JSON

The YouTube collector retrieves the raw data and constructs the needed information into a JSON object. The only task after this step is to issue an HTTP POST request (see the POST category of Fehler! Verweisquelle konnte nicht gefunden werden.) to the PROCeeD server to insert the schema-free JSON object in HTTP body.
Internally in the database system, a trigger is defined on the raw_data table for the monitoring of any insert operations. A part of the trigger definition is presented below:

```sql
create FUNCTION youtube.relationalise() RETURNS TRIGGER LANGUAGE plpgsql as $$
DECLARE
  _video_id BIGINT;
BEGIN
  insert into youtube.video(title, description, url, dislike_cnt, like_cnt, view_cnt)
  VALUES (NEW.raw->'title', NEW.raw->'description', NEW.raw->'url',
          NULLIF(NEW.raw->'dislike_cnt', '')::int,
          NULLIF(NEW.raw->'like_cnt', '')::int,
          NULLIF(NEW.raw->'view_cnt', '')::int)
  RETURNING id into _video_id;

  INSERT INTO youtube.author (author_id, name) VALUES (NEW.raw->
  >>'author_id', NEW.raw->'author') ON CONFLICT DO NOTHING;

  with RECURSIVE t as (SELECT
    nextval('youtube.comment_id_seq'::regclass) id, author,
    author_id, content, _video_id vid, null::bigint prev,
    replies from json_to_recordset(NEW.raw->'comments') as x(author text,
    author_id text, content text, replies json)
    UNION ALL
    select nextval('youtube.comment_id_seq'::regclass) id, x.author,
    x.author_id, x.content, _video_id vid, t.id prev, x.replies
    from t cross JOIN LATERAL json_to_recordset(t.replies) as x(author
    text, author_id text, content text, replies json)
    where t.replies is not null
  ) INSERT into youtube.author (author_id, name)  SELECT DISTINCT
  author_id, author from t on CONFLICT do NOTHING ;

  RETURN NEW;
END;
$$;
```

It uses a recursive query with common table expressions to handle the recursive comment and reply structure in an elegant way. After a successful insertion in the raw_data table, the inserted value is processed and mapped into the relational tables for further analysis.
Figure 11: The database schema of the table structure for the management of the crawled YouTube data.

3.1.3 DATA COLLECTORS

The Data Collectors in the PROCeeD system implement a so-called Web-Crawler\textsuperscript{13} and are responsible to retrieve OCTN-related data from the Web, the social network, etc. PROCeeD provides a flexible design, so that collectors can be registered and managed at runtime without restarting the PROCeeD server. Most important data needed for the registration is the base URL. A set of operations based on the RESTful design is specified in PROCeeD for the interplay between the core system and the collectors. These operations are summarised in Fehler! Verweisquelle konnte nicht gefunden werden. For simplicity, only HTTP GET is used in this design.

Table 6 Summary of the operations specified in the PROCeeD system to ensure correct interplay with collectors

<table>
<thead>
<tr>
<th>Path</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>/init</td>
<td>Initialise the collector</td>
</tr>
<tr>
<td>/start</td>
<td>Start the collector. The collector must work in a synchronised mode, i.e. with HTTP 200 response, it means the collector has been successfully started.</td>
</tr>
<tr>
<td>/stop</td>
<td>Stop the collector. The collector must work in a</td>
</tr>
</tbody>
</table>

\textsuperscript{13} https://en.wikipedia.org/wiki/Web_crawler
The collector configuration is encoded as a JSON object. Depending on the objective of the collector, the configuration can differ. For instance, for the built-in YouTube collector, this is a list of the keywords that will be used to search the videos.

As described before, collectors can be configured at runtime. Third-party collectors can be developed and coupled into the PROCeeD system in a flexible way. The only two prerequisites are: 1) the base URL needs to be provided; 2) the RESTful services summarised in Fehler! Verweisquelle konnte nicht gefunden werden. must be implemented by the collector. Regarding the storage, it can be decided by the collectors themselves. The data can be stored in a file system or in a database system independent of the PROCeeD database. However, for analysis purposes, it is recommend to put them in the PROCeeD database for performance reasons.

Within the work of WP3, data collectors for the social media platforms Twitter, YouTube and Facebook were implemented.

To simply the development of collectors using TypeScript, the attributes and behaviours of a collector is extracted and modelled as a TypeScript interface:

```typescript
interface ICollector {
    init(res :express.Response);
    start(res :express.Response);
    stop(res :express.Response);
    pause(res :express.Response);
    continue(res :express.Response);
    restart(res :express.Response);
    getConfiguration(res :express.Response);
}
```
Based on this interface design, an abstract TypeScript classes is implemented to provide basic and common functionalities needed by all collector in PROCeeD.

```typescript
abstract class AbstractCollector implements ICollector {
    private _name : string;
    private _port : number;
    private app = express();
    constructor(name : string, port : number) {
    }
    private installRESTfulServices(): void {
    }
    protected cfg : any;
    protected state : string; // stopped, running, paused
    abstract init(res : express.Response) : void ;
    abstract start(res : express.Response) : void ;
    abstract stop(res : express.Response) : void ;
    abstract pause(res : express.Response) : void ;
    abstract continue(res : express.Response) : void ;
    abstract restart(res : express.Response) : void ;
    abstract status(res : express.Response) : void ;
    abstract getConfiguration(res: express.Response) : void;
    abstract updateConfiguration(req: express.Request, res: express.Response) : void;
}
```

The main class of the Twitter collector inherits this abstract class and further implements its own Twitter-specific logics.

```typescript
export default class TwitterCollector extends AbstractCollector {
    private twitterAPI = {'filter': 'statuses/filter'};
    private clientTN = new Twitter();
    private clientOC = new Twitter();
    constructor(name: string, port : number) {
    }
    loadConfiguration() {
    }
    init(res: e.Response): void {
    }
    start(res: e.Response): void {
    }
    stop(res: e.Response): void {
    }
    pause(res: e.Response): void {
    }
    continue(res: e.Response): void {
    }
    restart(res: e.Response): void {
    }
    status(res: e.Response): void {
    }
    getConfiguration(res: e.Response): void {
    }
    updateConfiguration(req: e.Request, res: e.Response): void {
    }
}
```
To illustrate this kind of logical inheritance relationships, a UML diagram is provided below in Figure 12:

![UML class diagram of the inheritance relationships of PROCeeD collectors](image)

**Figure 12: The UML class diagram of the inheritance relationships of PROCeeD collectors**

The Twitter data collector is developed using the TypeScript programming language and runs inside a NodeJS environment. The program is trans-compiled at first into JavaScript during the development time before it is loaded into the runtime environment. The Twitter collector uses the Streaming API\(^{14}\) provided by Twitter.

The information collected by the Twitter data collector can be divided into two parts:

- Tweets related to a set of given topics. These topics are used as input for the collector and they are configurable by storing the value in a table inside of the PROCeeD database\(^{15}\). These topics are treated as “hashtags” when using the Twitter Streaming API.
- Dedicated users. All tweets of a dedicated user can be collected using a specific user ID. In the current version, this is still not configurable. For experimental purposes, the Twitter user “sunny_wantsome” is used. This is a hub in Twitter, where plenty of OCTN activities are aggregated.

A snapshot of the database structure is illustrated in Figure 13. This database has been dumped out and provided to IBM for the personality analysis. More details about this analysis can be found in Section 6.1.2.

---


\(^{15}\) This is done only for convenience. For better design, data tables related to individual collectors should be separated from the core PROCeeD database. However, since Twitter collector is a built-in collector, it shares the databases with the core system.
The YouTube data collector is implemented in the Python programming language and uses the Scrapy library (see section 3.1.6) to download and parse YouTube web-pages. It is build up out of three independent crawlers allowing an automated search for videos by given keywords, as well as the evaluation of each video found regarding the video description text and the comments. An overview of the data extraction process is shown in Fehler! Verweisquelle konnte nicht gefunden werden.

The input to the program is a list of keywords which represent the search terms used for the YouTube Search function. Entering a search term manually in the text field of the YouTube search function and pressing the “search” button results in the creation of a link which is automatically called. For example, when invoking a search with the search terms european commission, a link to the HTML page https://www.youtube.com/results?search_query=european+commission is created by YouTube, and the web browser of the caller is redirected to YouTube’s result page. All resulting videos for the according search terms are presented on this result page. The YouTube data collector imitates the user interaction of entering the search terms into the text field and pressing the search button by generating the HTML link to the results page automatically. This result page is then downloaded by the software by sending a HTTP request. The downloaded HTML code of the result page is then parsed to
retrieve the URLs of the result videos, all URLs are then stored in a list. The list of video URLs is then passed to the collectors which extract the video description text for each video from the video URL list, and the video comments, respectively. As the YouTube page for a specific video is implemented as an asynchronous web application which makes use of the asynchronous JavaScript and XML (Ajax 16) technique to load the video descriptions and video comments dynamically, the data collectors needed to be implemented accordingly to be able to load all the data.

3.1.4 **In-Database Analysis**

The data stored in the PROCeeD database can be analysed to detect potential OCTN activities. Currently, ten categories of the YouTube comments have been defined inside of the database. This work is done by analysing the formal methods introduced in (Schultes, Dorner, & Lehner, 2013). Technically, these categories are implemented as database views. One definition of such views is presented below:

```sql
CREATE VIEW offensive_discussion_post AS
SELECT comment.id,
    comment.text,
    comment.reply_to_comment,
    comment.author_id,
    comment.video_id,
    comment.spam_hint,
    comment.offensive_hint,
    comment.emotional_hint,
    comment.emoticon_hint,
    comment.part_of_thread,
    comment.number_of_words,
    comment.timestamp_hint,
    comment.keyword_match,
    comment.title_match
FROM youtube.comment
WHERE (comment.part_of_thread AND (comment.offensive_hint OR comment.spam_hint));
```

Two of these views are illustrated in [Fehler! Verweisquelle konnte nicht gefunden werden..](#). They have similar structures. The major differences are the rule definitions in the `WHERE` clause of the view definition.

---

16 https://en.wikipedia.org/wiki/Ajax_%28programming%29
3.1.5 Deployment

Fraunhofer has setup a public web server to host the PROCeeD application. It is available under the domain name proton.iais.fraunhofer.de. To access the PROCeeD application, the following URL is needed:

https://proton.iais.fraunhofer.de/proceed

Two measures have been conducted to protect the application against unauthorised access to the data:

- Use HTTPS for any human-in-the-middle attack. The SSL certification is based on the root certification of Fraunhofer. This should work for most modern browsers without giving any warning.
- Use HTTP basic authentication on the reverse proxy server to protect the path. The reason to choose the HTTP basic solution is for its simplicity. Similarly, the PROCeeD application itself does not provide any built-in authentication mechanism. Using the security measures on the reverse proxy server should be sufficient for the current release. In case of any vulnerabilities sophisticated security approach can still be applied.

3.1.6 Third-Party Libraries

A set of third party libraries is used to implement the data crawling and basic text analysis tasks. At the current state of implementation, the most important libraries are
• Scrapy\(^1\), a Python based open source web crawling framework.
• TextBlob\(^2\), a Python library for processing textual data.
• pandas\(^3\), a Python based open source data analysis and data visualization library.
• scikit-learn\(^4\), a Python based library for data analysis and machine learning.
• Beautiful Soup\(^5\), a Python based library for parsing HTML-code of web-pages.

### 3.2 Terms and Definitions

**Bitcoin Network:** A peer to peer payment network that does not use darknet technology. All transactions that have ever taken place are publicly available. Participants are pseudo anonymous by the fact that they use cryptographically generated identifiers for identification.

**Darknet:** Overlay technologies that use encryption to prevent an exchange between two or more parties being eavesdropped. Technologies range from use case where the parties are trusted (VPN,s Friend to Friend (F2F) systems) and use cases that protect the parties identity and anonymity (Tor (https://www.torproject.org/) , I2P ("Invisible Internet Project", https://geti2p.net/en/))

**Deep Web:** Parts of the world wide web not indexed by search engines.

Dark Web: Small part of the deep web that intentionally hidden and inaccessible through standard web browsers. The Dark Web is hidden through the use of Darknet technologies that protect identity and anonymity.

**Open Internet:** Internet transactions that are not hidden or encrypted.

**Synonym:** Clearnet

**Onion Routing Addresses:** Addresses that are used in Onion Routing Protocols based anonymous communications systems. In the Tor system an Onion Address looks like a URL with a special “.onion” suffix. These addresses are handled differently by software configured to do so (E.g. Tor Browser) and will not use public (traceable) DNS services to resolve the IP Address associated with the hidden services.

**Open Web:** Parts of the web indexed by search engines and accessible using standard browser technology.

\(^1\) https://scrapy.org/
\(^2\) https://textblob.readthedocs.io/en/dev/
\(^3\) http://pandas.pydata.org/
\(^4\) http://scikit-learn.org/stable/
\(^5\) https://www.crummy.com/software/BeautifulSoup/
Synonym: **Surface Web**

**Signposts:** Searchable artefacts in the surface web that point to something used in the dark web. Examples are onion routing addresses, bitcoin keys and addresses, bitcoin transactions,

**World Wide Web:** All the resources and users on the Internet that are using the Hypertext Transfer Protocol (HTTP).

### 3.3 The Dark Web

Cyberspace can be characterised in a number of ways. For this project we use a simple three layer classification

- Surface Web
- Deep Web
- Dark Web

#### 3.3.1 Surface Web

The surface web is commonly referred to as the part of the Internet indexed by search engines such as Google. It comprises normal HTML data as well as data sources accessible without authentication, e.g., e-Banking websites, or other unspecified access tooling, e.g., specific SQL database access queries.
According to different sources, the surface web comprises between 0.5% and 5% of all data accessible via the protocols underlying the Internet.

3.3.2 Deep Web
In the security world and mainstream media, the term “Deep Web” means that it is not indexed by search engines, and is a place with much more content than the “standard” Internet (World Wide Web) or Surface Web that everyone uses.

3.3.3 Dark Web
The Dark Web is a part of the Deep Web only accessible by special protocols and tools built on top of standard Internet protocols. Examples are TOR and I2P (see below). The common theme for these tools is the use of cryptography and sophisticated data routing technologies to hide the contents and flow of communication over the Internet. They all address a well-known vulnerability of the surface web, namely the ability to reveal a user’s identity and activities by simple traffic review and more sophisticated traffic analysis. The latter is increasingly being used as more and more surface web sites activate the use of the TLS cryptographic protocol effectively hiding the contents of data served to their users. What cannot be encrypted is the Internet packet headers as these are used to route the data from source to destination, thus permitting significant insight into traffic patterns between all types of computers and users of the Internet.

3.3.4 Relevant Technologies

TOR
Tor was originally designed, implemented and deployed in 2004 as a third-generation onion routing project of the U.S. Naval Research Laboratory. Developed primarily to protect government communications, it’s now used every day for many purposes by the military, law enforcement agencies, journalists, activists and a wide variety of other civilians. Tor at its most basic level is a peer-to-peer (P2P) network that allows anonymous, encrypted communication from host to host. It can also serve as a proxy with exit points known as “exit nodes” to allow users to anonymously browse web pages on the World Wide Web. This offers moderate anonymity to anyone looking to hide their identity as well as encrypt communication back to their host computer or device.

I2P
In Dark Web terms, the Invisible Internet Project (I2P) network would be considered the “very Deep Web.” Technically it’s very different from Tor, but the basic principle remains the same: It’s an encrypted P2P network that has
services as well as exit nodes, or out proxies. Whereas Tor has a much larger user base, I2P was designed and optimized for its hidden services. The following list of services is taken from the I2P main site:

- **Email**: integrated web mail interface, plugin for serverless email
- **Web browsing**: anonymous websites, gateways to and from the public Internet
- **Blogging and forums**: blogging and Syndie plugins
- **Website hosting**: integrated anonymous web server
- **Real-time chat**: instant messaging and Internet relay chat (IRC) clients
- **File sharing**: ed2k and Gnutella clients, integrated BitTorrent client
- **Decentralized file storage**: Tahoe-LAFS distributed filesystem plugin

I2P sites are known as “eepSites” with an extension of .i2p, and they offer anonymous and encrypted web hosting not accessible by the standard Internet. Finding these eepSites can prove far more difficult than locating nefarious sites on Tor, simply because the majority of underground markets and sites still remain in the .onion network. Currently I2P seems less used by criminals than Tor, but Tor has had its issues with arrests, snooping and even malware injected at exit nodes, so in the coming years we may see more movement of black markets to anonymous networks such as I2P.

**Bitcoin**

Bitcoin is a type of unregulated digital currency that was first created by Satoshi Nakamoto in 2008. Also known as a “cryptocurrency,” it was launched with the intention to bypass government currency controls and simplify online transactions by getting rid of third-party payment processing intermediaries. Of course, accomplishing this required more than just the money itself. There had to be a secure way to make transactions with the cryptocurrency.

Bitcoin transactions are stored and transferred using a distributed ledger on a peer-to-peer network that is open, public and anonymous. Bitcoin thrives due to anonymity. Anyone can look at the Bitcoin ledger and see every transaction that happened, but the account information is a meaningless sequence of numbers.

Bitcoin is not the only cryptocurrency that offers varying degrees of anonymization. Other include Dodgecoin, Litecoin, Ethereum, and others. More recent alternative offer increasing strong anonymization properties such as ZCash.

**Darknet Markets**

Dark Net Markets are virtual markets that are typically hosted as Tor hidden services. They look and feel like large online retailers providing a service between buyers & sellers transacting in Bitcoin or other cryptocoins. The most famous dark net market was Silk Road 1 which pioneered such an anonymous business model in 2011. More recent markets are known under such names as
DreamMarket, Agora, Valhalla, etc (see 9.1 for a more complete list of markets and offerings in the Darknet).

Dark Web Markets can be characterised with the following attributes:
- They are unregulated
- They are reached via an anonymizing mixer network such as Tor or I2P
- Participants are pseudonymous entities, who communicate privately and securely via public-key cryptography to arrange purchases
- The provide escrow schemes that pay sellers only on receipt of the goods from buyers
- Where sellers are required to post collateral similar to bonds before they are allowed to sell
- They use buyer ratings as a mechanism to build trust in sellers.

### 3.4 Dark web signpost identification

The Dark Web was initially designed to be a place to share information anonymously, addressing the communication needs of whistle blowers, journalists working in countries with suppressive regimes or secret service agents in enemy territory. It soon became -- and by some accounts is now predominantly -- used for criminal activity. It seems to be even possible to order assassinations, either directly (pay for murder and maiming of victims) or indirectly (by placing “crowd-sourced” bets on the day of death of VIPs – bets their murderers will surely win). The by far most significant use of the Darknet is trade in materials and services considered illicit in most jurisdictions, such as drugs, arms, malicious software, or illegal data, such as
stolen credit card or social security numbers. The Dark net markets being used to facilitate this business allow the participation of buyers and sellers in diverse geographic locations further complicating the policing of activities. Again, the References appendix contains a large number of such Pointers to the DarkWeb (section 9.1).

**Darknet markets**
Let us take the example of organised criminal using dark net markets to describe the sort of information that can be considered signposts.

**Bitcoins** are used for payments. Bitcoin public keys are signposts to the bitcoin payment system which is itself pseudonymous but where all transactions are visible. Users of darknet markets are generally cognisant of this sort of signpost and there are a number of schemes in operation that make following these signposts difficult such as **Bitcoin washers**

**Discussions** are on clear net chat forum such as Reddit.

**Social media posts** on platforms such as Twitter that contain links to darknet platforms.

**Market recommendations** use trust based systems built on recommendations. Here partially anonymised users rate their experience of using a particular seller. Enough information needs to be distilled to make the recommendation believable. This potentially leaks information that can be used as signposts to forum etc. For example, recommendations commonly are displayed with the first and last letters of the user name that submitted the recommendation. It has been known that people will use the same username across different web domains.

**Lexical analysis** of the text that participants leave in the form of tweets, recommendations, comments etc.

### 3.5 Project results

In general, we learned during the project that most Darknet market actors are relatively well-represented also in the surface web due to their need to obtain customer leads. As in the surface web, many Darknet users begin their foray by searching for certain terms, combined with the unique Internet identifier of Darknet websites “.onion”: For example, a search on google for “drugs .onion”, i.e., restricting the results to contain links to Darknet servers, returns at the time of writing 44’000 entries with the top results listing HowTo (order) guides as well as directory of Darknet market places for such products.

Another very common starting point is the search engine DuckDuckGo: It regularly crawls the Darknet itself, which is however made difficult by the rather dynamic nature of the Darknet: Due to the illegal nature of many
services, content and addresses change more frequently than in the surface web: For example, during the course of this project, one address for the presently most significant Darknet Market, DreamMarket, became suddenly inaccessible, pointing to a successful law enforcement activity. However, the well-published backup sites of DreamMarket remained active, allowing trading to continue. Also constantly updated are Meta-Site review sites like DarkNetStats (dnstats.net) or Darknetmarkets (https://darknetmarkets.org/) that constantly observe activity, availability and reputation of popular Darknet sites). Other Darknet search engines are TORCH (http://xmh57jrznw6insl.onion and GRAMS: http://grams7enufi7jmdl.onion).

To sum up, Darknet signposts exist aplenty in the clearweb for all illegal commodities, primarily as sellers need to find buyers just as in the surface web. The references presented provide proof to this fact. Another picture presents itself for Darknet content of terrorist organizations: For such organizations, the Darknet presents a perfect technology to distribute digital content, such as instructions for bomb-making, but by their nature, they try to hide this content and not advertise it. The contrary is true for distribution of propaganda material: This is generally made available via surface web channels such as YouTube and Twitter such as to reach a broad an audience as possible. The latter observation also was the rationale for designing the experimental platform utilized in this project.

In order to get some statistics of the use of .onion and .i2p links in social media, we investigated their use in Twitter and in reddit (online and open community site http://www.reddit.com).

3.5.1 TWITTER
In a corpus of approximately 200 Mio Twitter messages captured in January 2017, we found only a minute number of links to .onion addresses (only 43 tweets contained a .onion link in the ‘body’ of the tweet, though we found 1242 as part of a user’s profile) and to .i2p addresses (no links in the ‘body’, 9 as part of user’s profiles).

3.5.2 REDDIT
As entries in reddit communities (called subreddits) are openly available, we used a corpus with comments from January 2015 through June 2017 to check for .onion/.i2p links.

Table 1 and 2 below show the distribution of unique and total number of .onion respectively .i2p links in different communities in reddit (subreddits). Although we find more links here compared to Twitter data, the absolute number is still very low – with again far fewer .iop links than .onion links. The tables also show in which communities these links have been found – besides the ‘onions’ and ‘i2p’ communities these are mostly communities linked to DarkWeb marketplaces.
3.6 Personality as a signpost

One hypothesis for WP3 in the PROTON is project is that where users express themselves in cyberspace, it is possible to lexically analyse the text that they create and from this derive some aspects of their personality. One aspects of platforms in cyberspace is the ability that we have to group people based on
their interests. This can be topics in online forum, people in the case of Twitter, content in the case of images, music and video. A further hypothesis is that we can use our ability to group text from multiple persons based on their interests in order to develop personality traits representative for a group of persons. The aim would be to identify those aspects of personality that are particularly dominant for a specific interest group.

For this approach to work there are a number of assumptions that we need to be aware of:

- That words are written by individuals about themselves or about any topic are chosen and written by themselves
- The diversity of a group will impact the accuracy of an analysis.
- Text that is generated by transcription or translation services can impact the reliability of the results
- That personalities may evolve over time
- Where a person writes fictional text this may reflect the personality of a fictional character not the author

Transcripts of interviews

The personality analysis software that we will use is the IBM Watson™ Personality Insights service. This software has been developed outside of the PROTON project and is available as an application programming interface (API) on IBM’s cloud platform. The API enables applications to derive insights from social media, enterprise data, or other digital communications. The service uses linguistic analytics to infer individuals' intrinsic personality characteristics, including Big Five, Needs, and Values, from digital communications such as email, text messages, tweets, and forum posts. This technology is typically used by online enterprises to provide personalized product recommendations to customers.

The concept is based on the well-accepted theory of psychology, marketing, and other fields, in that human language reflects personality, thinking style, social connections, and emotional states. The frequency with which we use certain categories of words can provide clues to these characteristics. Several researchers found that variations in word usage in writings such as blogs, essays, and tweets can predict aspects of personality (Fast & Funder, 2008; Gill et al., 2009; Golbeck et al., 2011; Hirsh & Peterson, 2009; and Yarkoni, 2010).

Personality Insights analytics are developed based on the psychology of language in combination with data analytics algorithms. The characteristics are described in terms of three models:

- **Big Five**
- **Needs**
- **Values**
3.7 Data input to Personality Insights analytics

To reliably infer an individual's personality, does the text have to be written by the individual or can it be written about the individual by someone else? When one person writes about another person, whose personality is inferred from the text?

IBM's intuition is that writing always reflects the author's personality, regardless of the subject matter. For example, if individual A writes text about individual B, an analysis of the text infers the personality of individual A. Although individual A is writing about individual B, it is individual A who is choosing the words to express things about individual B. However, IBM has done no studies to explicitly test this scenario.

Inferring personality from speech transcripts

Speech transcription engines, such as the IBM Speech to Text service, generate written text from spoken words. Because different transcription engines have different accuracy ranges, customers who transcribe speech to text for use as input to the Personality Insights service need to be aware that the results vary widely depending on the performance of the engine. Specifically, IBM advises clients and business partners to determine the quality of the transcription against two types of errors:

- Drop out: A spoken word is omitted from the transcription.
- Substitution: A spoken word is transcribed incorrectly.

Substitution can be a more serious issue because it can introduce words that were not spoken but that match words that are used to determine personality. Before you use transcribed text, consider manually correcting the text of a test corpus and counting the errors that you find. Then, compare the results of the automatically generated text with the manually corrected version to determine the variance in results due to transcription errors.

Inferring personality from translated text

Language translation services translate text that is written in one language to another language. As with speech transcription, the question arises whether translated text can be used as input to the Personality Insights service. IBM does not recommend that you use text that is obtained from translation services as input to the Personality Insights service. Depending on the translation service, the results of both the translation and the personality inference can vary widely. Words, their senses, and cultural sensitivities tend to get lost in translation, yielding invalid results.

IBM recommends that you use as input only text that is written in languages in which the Personality Insights service is enabled. Language-enabled versions of the service parse the input text in its native language, use native language dictionaries to identify personality characteristics, and use models that are calibrated for the native language to produce statistical results. If you must
analyze translated text, IBM recommends that you first manually translate some text samples by using the services of a human language expert. You can then compare the results that the Personality Insights service obtains from both manually and automatically translated text to understand the variance in results.

IBM will continue to add more languages as business demand increases. IBM understands that the Personality Insights service might not support your native language fast enough for your purposes. IBM is conducting tests to compare results that are obtained from native language enablement with results obtained from language translation services and will report its findings as they become available.

3.8 Big Five

The first personality model, Big Five, was developed by Costa and Norman and is the most widely used model to generally describe how a person engages with the world. The model includes five primary categories, or dimensions:

- Agreeableness
- Conscientiousness
- Extraversion
- Emotional range
- Openness

Each top-level dimension has six subdimensions, or facets, that further characterize an individual according to the dimension. The following sections introduce the dimensions and provide detailed information about their facets. Each section includes three tables that provide the following information:

- The first table, Facets, lists the dimension's facets and provides a brief description of individuals who score highly in each facet.
- The second table, Range of characteristics, presents general descriptions that might apply to individuals whose scores evidence more or less of each facet of the dimension, as well as terms that might describe such individuals.
- The third table, Primary and secondary dimensions, presents information that relates the dimension to other dimensions, describing combinations of personality characteristics. In this table, the rows represent the dimension that is identified as the individual's primary characteristic, and the columns provide terms that might describe people whose scores on the remaining secondary dimensions are higher or lower. The table provides interesting insight into how primary and secondary characteristics might interrelate to represent an individual's composite personality.
### 3.8.1 Agreeableness

**Agreeableness** is a person's tendency to be compassionate and cooperative toward others.

<table>
<thead>
<tr>
<th>Facet</th>
<th>People who score high...</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Altruism / Altruistic</strong></td>
<td>Find that helping others is genuinely rewarding, that doing things for others is a form of self-fulfillment rather than self-sacrifice.</td>
</tr>
<tr>
<td><strong>Cooperation / Accommodating / Compliance</strong></td>
<td>Dislike confrontation. They are perfectly willing to compromise or to deny their own needs to get along with others.</td>
</tr>
<tr>
<td><strong>Modesty / Modest</strong></td>
<td>Are unassuming, rather self-effacing, and humble. However, they do not necessarily lack self-confidence or self-esteem.</td>
</tr>
<tr>
<td><strong>Morality / Uncompromising / Sincerity</strong></td>
<td>See no need for pretense or manipulation when dealing with others and are therefore candid, frank, and genuine.</td>
</tr>
<tr>
<td><strong>Sympathy / Empathetic</strong></td>
<td>Are tender-hearted and compassionate.</td>
</tr>
<tr>
<td><strong>Trust / Trusting of others</strong></td>
<td>Assume that most people are fundamentally fair, honest, and have good intentions. They take people at face value and are willing to forgive and forget.</td>
</tr>
</tbody>
</table>

**Table 9 Personality Trait - Agreeableness – Facets**

<table>
<thead>
<tr>
<th>Description of LOW value</th>
<th>Term</th>
<th>Facet</th>
<th>Term</th>
<th>Description of HIGH value</th>
</tr>
</thead>
<tbody>
<tr>
<td>You are more concerned with taking care of yourself than taking time for others.</td>
<td>Self-focused</td>
<td>Altruism</td>
<td>Altruistic</td>
<td>You feel fulfilled when helping others and will go out of your way to do so.</td>
</tr>
<tr>
<td>You do not shy away from contradicting others.</td>
<td>Contrary</td>
<td>Cooperation</td>
<td>Accommodating</td>
<td>You are easy to please and try to avoid confrontation.</td>
</tr>
<tr>
<td>You hold yourself in high regard and are satisfied with who you are.</td>
<td>Proud</td>
<td>Modesty</td>
<td>Modest</td>
<td>You are uncomfortable being the center of attention.</td>
</tr>
<tr>
<td>You are comfortable using every trick in the book to get what you want.</td>
<td>Compromising</td>
<td>Morality</td>
<td>Un-compromising</td>
<td>You think it is wrong to take advantage of others to get ahead.</td>
</tr>
<tr>
<td>You think people should generally</td>
<td>Hard-hearted</td>
<td>Sympathy</td>
<td>Empathetic</td>
<td>You feel what others feel and are</td>
</tr>
</tbody>
</table>
rely more on themselves than on others. compassionate toward them.

You are wary of other people's intentions and do not trust easily. Cautious of others Trusting of others You believe the best of others and trust people easily.

Table 10 Personality Trait - Agreeableness - Range of characteristics

<table>
<thead>
<tr>
<th></th>
<th>Conscientiousness</th>
<th>Extraversion</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Helpful, cooperative, considerate, respectful, polite</td>
<td>Unpretentious, self-effacing</td>
</tr>
<tr>
<td>Low</td>
<td>Strict, rigid, stern</td>
<td>Inconsiderate, impolite, distrustful, uncooperative, thoughtless</td>
</tr>
</tbody>
</table>

Table 11 Personality Trait - Agreeableness - Primary and secondary dimensions – part 1

<table>
<thead>
<tr>
<th></th>
<th>Emotional range</th>
<th>Openness</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Sentimental, affectionate, sensitive, soft, passionate</td>
<td>Generous, pleasant, tolerant, peaceful, flexible</td>
</tr>
<tr>
<td>Low</td>
<td>Strict, rigid, stern</td>
<td>Inconsiderate, impolite, distrustful, uncooperative, thoughtless</td>
</tr>
</tbody>
</table>

Table 12 Personality Trait - Agreeableness - Primary and secondary dimensions – part 2

3.8.2 Conscientiousness

Conscientiousness is a person’s tendency to act in an organized or thoughtful way.
<table>
<thead>
<tr>
<th>Facet</th>
<th>People who score high...</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Achievement striving / Driven</strong></td>
<td>Try hard to achieve excellence. Their drive to be recognized as successful keeps them on track as they work hard to accomplish their goals.</td>
</tr>
<tr>
<td><strong>Cautiousness / Deliberate / Deliberateness</strong></td>
<td>Are disposed to think through possibilities carefully before acting.</td>
</tr>
<tr>
<td><strong>Dutifulness / Dutiful / Sense of responsibility</strong></td>
<td>Have a strong sense of duty and obligation.</td>
</tr>
<tr>
<td><strong>Orderliness / Organized</strong></td>
<td>Are well-organized, tidy, and neat.</td>
</tr>
<tr>
<td><strong>Self-discipline / Persistent</strong></td>
<td>Have the self-discipline, or &quot;will-power,&quot; to persist at difficult or unpleasant tasks until they are completed.</td>
</tr>
<tr>
<td><strong>Self-efficacy / Self-assured / Sense of competence</strong></td>
<td>Are confident in their ability to accomplish things.</td>
</tr>
</tbody>
</table>

**Table 13 Personality Trait - Conscientiousness – Facets**

<table>
<thead>
<tr>
<th>Description of LOW value</th>
<th>Description of HIGH value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
<td></td>
</tr>
<tr>
<td>You are content with your level of accomplishment and do not feel the need to set ambitious goals.</td>
<td>You set high goals for yourself and work hard to achieve them.</td>
</tr>
<tr>
<td>You would rather take action immediately than spend time deliberating making a decision.</td>
<td>You carefully think through decisions before making them.</td>
</tr>
<tr>
<td>You do what you want, disregarding rules and obligations.</td>
<td>You take rules and obligations seriously, even when they are inconvenient.</td>
</tr>
<tr>
<td>You do not make a lot of time for organization in your daily life.</td>
<td>You feel a strong need for structure in your life.</td>
</tr>
<tr>
<td>You have a hard time sticking with difficult tasks for a long period of time.</td>
<td>You can tackle and stick with tough tasks.</td>
</tr>
<tr>
<td>You frequently doubt your ability to achieve your goals.</td>
<td>You feel you have the ability to succeed in the</td>
</tr>
</tbody>
</table>
Deliverable D3.1

Table 14 Personality Trait - Conscientiousness - Range of characteristics

<table>
<thead>
<tr>
<th>Agreeableness</th>
<th>Extraversion</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>High</td>
<td>Dependable, responsible, reliable, mannerly, considerate</td>
</tr>
<tr>
<td>Low</td>
<td>Unpretentious, self-effacing</td>
</tr>
</tbody>
</table>

Table 15 Personality Trait - Conscientiousness - Primary and secondary dimensions – part 1

Table 16 Personality Trait - Conscientiousness - Range of characteristics

<table>
<thead>
<tr>
<th>Emotional range</th>
<th>Openness</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>High</td>
<td>Particular, high-strung</td>
</tr>
<tr>
<td>Low</td>
<td>Scatterbrained, inconsistent, erratic, forgetful, impulsive</td>
</tr>
</tbody>
</table>

Table 17 Personality Trait - Conscientiousness - Primary and secondary dimensions – part 2

3.8.3 Extraversion
Extraversion is a person's tendency to seek stimulation in the company of others.
**Activity level / Energetic**

Lead fast-paced and busy lives. They do things and move about quickly, energetically, and vigorously, and they are involved in many activities.

**Assertiveness / Assertive**

Like to take charge and direct the activities of others. They tend to be leaders in groups.

**Cheerfulness / Cheerful / Positive emotions**

Experience a range of positive feelings, including happiness, enthusiasm, optimism, and joy.

**Excitement-seeking**

Are easily bored without high levels of stimulation.

**Friendliness / Outgoing / Warmth**

Genuinely like other people and openly demonstrate positive feelings toward others.

**Gregariousness / Sociable**

Find the company of others pleasantly stimulating and rewarding. They enjoy the excitement of crowds.

<table>
<thead>
<tr>
<th>Table 18 Personality Trait - Extraversion – Facets</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description of LOW value</strong></td>
</tr>
<tr>
<td><strong>Description</strong></td>
</tr>
<tr>
<td>You appreciate a relaxed pace in life.</td>
</tr>
<tr>
<td>You prefer to listen than to talk, especially in group settings.</td>
</tr>
<tr>
<td>You are generally serious and do not joke much.</td>
</tr>
<tr>
<td>You prefer activities that are quiet, calm, and safe.</td>
</tr>
<tr>
<td>You are a private person and do not let many people in.</td>
</tr>
<tr>
<td>You have a strong desire to have time to yourself.</td>
</tr>
</tbody>
</table>
Table 19 Personality Trait - Extraversion - Range of characteristics

<table>
<thead>
<tr>
<th>Agreeableness</th>
<th>Conscientiousness</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Social, energetic, enthusiastic, communicative, vibrant</td>
<td>Opinionated, forceful, domineering, boastful, bossy</td>
</tr>
<tr>
<td>Unaggressive, humble, submissive, timid, compliant</td>
<td>Skeptical, wary of others, seclusive, uncommunicative, unsociable</td>
</tr>
</tbody>
</table>

Table 20 Personality Trait - Extraversion - Primary and secondary dimensions – part 1

<table>
<thead>
<tr>
<th>Emotional range</th>
<th>Openness</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Explosive, wordy, extravagant, volatile, flirtatious</td>
<td>Confident, bold, assured, uninhibited, courageous</td>
</tr>
<tr>
<td>Guarded, pessimistic, secretive, cowardly</td>
<td>Tranquil, sedate, placid, impartial, unassuming</td>
</tr>
</tbody>
</table>

Table 21 Personality Trait - Extraversion - Primary and secondary dimensions – part 2

3.8.4 Emotional Range

**Emotional range**, also referred to as Neuroticism or Natural reactions, is the extent to which a person's emotions are sensitive to the individual's environment.

<table>
<thead>
<tr>
<th>Facet</th>
<th>People who score high...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger / Fiery</td>
<td>Have a tendency to feel angry.</td>
</tr>
<tr>
<td>Anxiety / Prone to worry</td>
<td>Often feel like something unpleasant, threatening, or dangerous is about to happen. The &quot;fight-or-flight&quot; system of their brains is too easily and too often engaged.</td>
</tr>
</tbody>
</table>
### Deliverable D3.1

<table>
<thead>
<tr>
<th><strong>Depression / Melancholy / Moodiness</strong></th>
<th>Tend to react more readily to life's ups and downs.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Immoderation / Self-indulgence</strong></td>
<td>Feel strong cravings and urges that they have difficulty resisting, even though they know that they are likely to regret them later. They tend to be oriented toward short-term pleasures and rewards rather than long-term consequences.</td>
</tr>
<tr>
<td><strong>Self-consciousness</strong></td>
<td>Are sensitive about what others think of them. Their concerns about rejection and ridicule cause them to feel shy and uncomfortable around others; they are easily embarrassed.</td>
</tr>
<tr>
<td><strong>Vulnerability / Susceptible to stress / Sensitivity to stress</strong></td>
<td>Have difficulty coping with stress. They experience panic, confusion, and helplessness when under pressure or when facing emergency situations.</td>
</tr>
</tbody>
</table>

**Table 22 Personality Trait - Emotional range – Facets**

<table>
<thead>
<tr>
<th>Description of LOW value</th>
<th>Term</th>
<th>Facet</th>
<th>Description of HIGH value</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>It takes a lot to get you angry.</td>
<td>Mild-tempered</td>
<td>Anger</td>
<td>You have a fiery temper, especially when things do not go your way.</td>
<td>Fiery</td>
</tr>
<tr>
<td>You tend to feel calm and self-assured.</td>
<td>Self-assured</td>
<td>Anxiety</td>
<td>You tend to worry about things that might happen.</td>
<td>Prone to worry</td>
</tr>
<tr>
<td>You are generally comfortable with yourself as you are.</td>
<td>Content</td>
<td>Depression</td>
<td>You think quite often about the things you are unhappy about.</td>
<td>Melancholy</td>
</tr>
<tr>
<td>You have control over your desires, which are not particularly intense.</td>
<td>Self-controlled</td>
<td>Immoderation</td>
<td>You feel your desires strongly and are easily tempted by them.</td>
<td>Hedonistic</td>
</tr>
<tr>
<td>You are hard to embarrass and are self-confident most of the time.</td>
<td>Confident</td>
<td>Self-consciousness</td>
<td>You are sensitive about what others might be thinking of you.</td>
<td>Self-conscious</td>
</tr>
<tr>
<td>You handle unexpected events calmly and effectively.</td>
<td>Calm under pressure</td>
<td>Vulnerability</td>
<td>You are easily overwhelmed in stressful situations.</td>
<td>Susceptible to stress</td>
</tr>
</tbody>
</table>
Table 23 Personality Trait - Emotional range - Range of characteristics

<table>
<thead>
<tr>
<th>Agreeableness</th>
<th>Conscientiousness</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Emotion, gullible,</td>
<td>Temperamental, irritable,</td>
</tr>
<tr>
<td>affectionate,</td>
<td>quarrelsome, impatient, grumpy</td>
</tr>
<tr>
<td>sensitive, soft</td>
<td>Particular, high-strung</td>
</tr>
<tr>
<td></td>
<td>Compulsive, nosy, self-indulgent, forgetful, impulsive</td>
</tr>
<tr>
<td>Patient, relaxed,</td>
<td>Unemotional, insensitive,</td>
</tr>
<tr>
<td>undemanding, down-to-earth</td>
<td>unaffectionate, passionless</td>
</tr>
<tr>
<td></td>
<td>Rational, objective, steady, logical,</td>
</tr>
<tr>
<td></td>
<td>decisive</td>
</tr>
<tr>
<td></td>
<td>Informal, low-key</td>
</tr>
</tbody>
</table>

Table 24 Personality Trait - Emotional range - Primary and secondary dimensions – part 1

<table>
<thead>
<tr>
<th>Extraversion</th>
<th>Openness</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Excitable,</td>
<td>Guarded, fretful, insecure,</td>
</tr>
<tr>
<td>wordy,</td>
<td>secretive</td>
</tr>
<tr>
<td>flirtatious,</td>
<td>Excitable, passionate,</td>
</tr>
<tr>
<td>explosive,</td>
<td>sensual</td>
</tr>
<tr>
<td>extravagant</td>
<td>Easily rattled, easily irked, apprehensive</td>
</tr>
<tr>
<td>Unself-</td>
<td>Unassuming, unexcitable, placid, tranquil</td>
</tr>
<tr>
<td>conscious,</td>
<td>Heartfelt, versatile, creative, intellectual,</td>
</tr>
<tr>
<td>wearless,</td>
<td>insightful</td>
</tr>
<tr>
<td>indefatigible</td>
<td>Imperturbable, insensitive</td>
</tr>
</tbody>
</table>

Table 25 Personality Trait - Emotional range - Primary and secondary dimensions – part 2

3.8.5 OPENNESS

Openness, or Open to Experience, is the extent to which a person is open to experiencing a variety of activities.

<table>
<thead>
<tr>
<th>Facet</th>
<th>People who score high...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adventurousness /</td>
<td>Are eager to try new activities and experience</td>
</tr>
<tr>
<td>Willingness to</td>
<td>different things. They find</td>
</tr>
<tr>
<td>experiment</td>
<td>familiarity and routine boring.</td>
</tr>
<tr>
<td>Artistic interests</td>
<td>Love beauty, both in art and in nature. They</td>
</tr>
<tr>
<td></td>
<td>become easily involved and</td>
</tr>
<tr>
<td></td>
<td>absorbed in artistic and natural events. With</td>
</tr>
<tr>
<td></td>
<td>intellect, this facet is one of the</td>
</tr>
</tbody>
</table>
two most important, central aspects of this characteristic.

<table>
<thead>
<tr>
<th>Emotionality / Emotionally aware / Depth of emotions</th>
<th>Have good access to and awareness of their own feelings.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imagination</td>
<td>View the real world as often too plain and ordinary. They use fantasy not as an escape but as a way of creating for themselves a richer and more interesting inner-world.</td>
</tr>
<tr>
<td>Intellect / Intellectual curiosity</td>
<td>Are intellectually curious and tend to think in symbols and abstractions. With artistic interests, this facet is one of the two most important, central aspects of this characteristic.</td>
</tr>
<tr>
<td>Liberalism / Authority challenging / Tolerance for diversity</td>
<td>Have a readiness to challenge authority, convention, and traditional values.</td>
</tr>
</tbody>
</table>

**Table 26 Personality Trait - Openness – Facets**

<table>
<thead>
<tr>
<th>Description of LOW value</th>
<th>Term</th>
<th>Facet</th>
<th>Description of HIGH value</th>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>You enjoy familiar routines and prefer not to deviate from them.</td>
<td>Consistent</td>
<td>Adventurousness</td>
<td>You are eager to experience new things.</td>
<td>Adventurous</td>
<td></td>
</tr>
<tr>
<td>You are less concerned with artistic or creative activities than most people.</td>
<td>Unconcerned with art</td>
<td>Artistic interests</td>
<td>You enjoy beauty and seek out creative experiences.</td>
<td>Appreciative of art</td>
<td></td>
</tr>
<tr>
<td>You do not frequently think about or openly express your emotions.</td>
<td>Dispassionate</td>
<td>Emotionality</td>
<td>You are aware of your feelings and how to express them.</td>
<td>Emotionally aware</td>
<td></td>
</tr>
<tr>
<td>You prefer facts over fantasy.</td>
<td>Down-to-earth</td>
<td>Imagination</td>
<td>You have a wild imagination.</td>
<td>Imaginative</td>
<td></td>
</tr>
<tr>
<td>You prefer dealing with the world as it is, rarely considering abstract ideas.</td>
<td>Concrete</td>
<td>Intellect</td>
<td>You are open to and intrigued by new ideas and love to explore them.</td>
<td>Philosophical</td>
<td></td>
</tr>
<tr>
<td>You prefer following with tradition to maintain a sense of</td>
<td>Respectful of authority</td>
<td>Liberalism</td>
<td>You prefer to challenge authority and traditional values</td>
<td>Authority-challenging</td>
<td></td>
</tr>
</tbody>
</table>
stability. to help bring about change.

Table 27 Personality Trait - Openness - Range of characteristics

<table>
<thead>
<tr>
<th>Agreeableness</th>
<th>Conscientiousness</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>High</td>
<td>Idealistic,</td>
</tr>
<tr>
<td></td>
<td>diplomatic, deep</td>
</tr>
<tr>
<td></td>
<td>tactful,</td>
</tr>
<tr>
<td></td>
<td>genial</td>
</tr>
<tr>
<td>Low</td>
<td>Simple,</td>
</tr>
<tr>
<td></td>
<td>dependent</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 28 Personality Trait - Openness - Primary and secondary dimensions – part 1

<table>
<thead>
<tr>
<th>Extraversion</th>
<th>Emotional range</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>High</td>
<td>Worldly,</td>
</tr>
<tr>
<td></td>
<td>theatrical,</td>
</tr>
<tr>
<td></td>
<td>eloquent,</td>
</tr>
<tr>
<td></td>
<td>inquisitive,</td>
</tr>
<tr>
<td></td>
<td>intense</td>
</tr>
<tr>
<td>Low</td>
<td>Verbose,</td>
</tr>
<tr>
<td></td>
<td>unscrupulous,</td>
</tr>
<tr>
<td></td>
<td>pompous</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 29 Openness - Primary and secondary dimensions – part 2

3.9 Needs

The second model, Needs, describes at a high level which aspects of a product are likely to resonate with the author of the text. The model includes twelve categories of needs based on Kotler’s and Ford’s work in marketing. The following table describes the twelve needs that are evaluated by the Personality Insights service.

<table>
<thead>
<tr>
<th>Need</th>
<th>People who score high...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excitement</td>
<td>Want to get out there and live life, have upbeat emotions, and want to have fun.</td>
</tr>
<tr>
<td>Harmony</td>
<td>Appreciate other people, their viewpoints, and their feelings.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Curiosity</td>
<td>Have a desire to discover, find out, and grow.</td>
</tr>
<tr>
<td>Ideal</td>
<td>Desire perfection and a sense of community.</td>
</tr>
<tr>
<td>Closeness</td>
<td>Relish being connected to family and setting up a home.</td>
</tr>
<tr>
<td>Self-expression</td>
<td>Enjoy discovering and asserting their own identities.</td>
</tr>
<tr>
<td>Liberty</td>
<td>Have a desire for fashion and new things, as well as the need for escape.</td>
</tr>
<tr>
<td>Love</td>
<td>Enjoy social contact, whether one-to-one or one-to-many. Any brand that is involved in bringing people together taps this need.</td>
</tr>
<tr>
<td>Practicality</td>
<td>Have a desire to get the job done, a desire for skill and efficiency, which can include physical expression and experience.</td>
</tr>
<tr>
<td>Stability</td>
<td>Seek equivalence in the physical world. They favor the sensible, the tried and tested.</td>
</tr>
<tr>
<td>Challenge</td>
<td>Have an urge to achieve, to succeed, and to take on challenges.</td>
</tr>
<tr>
<td>Structure</td>
<td>Exhibit groundedness and a desire to hold things together. They need things to be well organized and under control.</td>
</tr>
</tbody>
</table>

**Table 30 Personality Trait - Needs**

### 3.10 Values

The third model, Values, describes motivating factors that influence the author's decision-making. The model includes five dimensions of human values based on Schwartz's work in psychology. The following table describes the five values that are inferred by the Personality Insights service.

<table>
<thead>
<tr>
<th>Value</th>
<th>People who score high...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-transcendence / Helping others</td>
<td>Show concern for the welfare and interests of others.</td>
</tr>
<tr>
<td>Hedonism / Taking pleasure in life</td>
<td>Seek pleasure and sensuous gratification for themselves.</td>
</tr>
<tr>
<td>Self-enhancement / Achieving success</td>
<td>Seek personal success for themselves.</td>
</tr>
<tr>
<td>Open to change / Excitement</td>
<td>Emphasize independent action, thought, and feeling, as well as a readiness for new experiences.</td>
</tr>
</tbody>
</table>

**Table 31 Personality Trait - Values**
3.10.1 How Personality Characteristics are Inferred

The Personality Insights service infers personality characteristics from textual information based on an open-vocabulary approach. This method reflects the latest trend in the research about personality inference (Arnoux et al., 2017, Schwartz et al., 2013, and Plank & Hovy, 2015).

The service first tokenizes the input text to develop a representation in an n-dimensional space. The service uses an open-source word-embedding technique called GloVe External link icon to obtain a vector representation for the words in the input text (Pennington et al., 2014). It then feeds this representation to a machine-learning algorithm that infers a personality profile with Big Five, Needs, and Values characteristics. To train the algorithm, the service uses scores obtained from surveys conducted among thousands of users along with data from their Twitter feeds.

IBM developed the models for all supported languages in an identical way. The models were developed independent of user demographics such as age, gender, or culture. In the future, IBM might develop models that are specific to different demographic categories.

Earlier versions of the service used the Linguistic Inquiry and Word Count (LIWC) psycholinguistic dictionary with its machine-learning model. However, the open-vocabulary approach just described outperforms the LIWC-based model. For more information about the service's precision for each language in terms of average Mean Absolute Error (MAE) and correlation, see How precise is the service. For guidance about providing input text to achieve the most accurate results, see Providing sufficient input.

IBM conducted a validation study to understand the accuracy of the service's approach to inferring a personality profile. IBM collected survey responses and Twitter feeds from between 1500 and 2000 participants for all characteristics and languages. To establish ground truth, participants took four sets of standard psychometric tests:

- 50-item Big Five derived from the International Personality Item Pool (IPIP)
- 120-item Facet derived from the IPIP Neuroticism, Extraversion & Openness (IPIP-NEO)
- 52-item fundamental Needs developed by IBM
- 26-item basic Values developed by Schwartz

IBM then compared the scores that were derived by its models with the survey-based scores for the Twitter users. Based on these results, IBM determined the average Mean Absolute Error and average correlation between
the inferred and actual scores for the different categories of personality characteristics. Per-language average MAE and correlation provides the statistical values for each supported language.

**Average Mean Absolute Error**

Mean Absolute Error (MAE) is a metric that is used to measure the difference between actual and predicted values. For the Personality Insights service, the actual value, or ground truth, is the personality score that was obtained by administering a personality survey. The predicted value is the score that the service's models produce.

IBM computed the MAE by taking the average of the absolute value of the difference between the actual and predicted scores. IBM used the absolute value because predicting more or less of the actual value is irrelevant; as long as there is a difference, the model is penalized by the magnitude of the error. The lower the MAE, the better the performance of the model. IBM uses a scale of 0 to 1 for MAE, where 0 means no error (the predicted value is the exact same as the actual value), and 1 means maximum error.

**Average correlation**

Average correlation is a statistical term that measures the interdependence of two variables. With this metric, IBM measured the correlation between inferred and actual scores for the different categories of personality characteristics. If the predicted score closely tracks the actual results, the correlation score is high; otherwise, the score is low.

IBM measures correlation on scale of -1 to 1: 1 indicates a perfect direct (increasing) linear relationship, and -1 indicates a perfect inverse (decreasing) linear relationship. In all other cases, the value lies between these extremes. If the variables are independent (they have no relationship at all), the correlation is 0.

IBM looks for numbers closer to 1 for best predictions. But personalities are difficult to predict based solely on text, and it is rare to see correlations exceed 0.4 for these types of psychological models. In research literature for this domain, correlations above 0.2 are considered acceptable.

**Per-language average MAE and correlation**

The following table shows per-language average MAE and correlation results for the Personality Insights service. The results place the service at the cutting edge of personality inference from textual data as indicated by Schwartz et al. (2013) and Plank and Hovy (2015).

<table>
<thead>
<tr>
<th>Language</th>
<th>Big Five dimensions</th>
<th>Big Five facets</th>
<th>Needs</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average MAE</td>
<td>0.12</td>
<td>0.12</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td>Average correlation</td>
<td>0.33</td>
<td>0.28</td>
<td>0.22</td>
<td>0.24</td>
</tr>
<tr>
<td>Spanish</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### 3.11 Analysis of criminals in the Dark Web

This section details our approach for creating an agent persona for a dark net criminal. It describes an approach to persona development that can be used for the basis for agent based modelling. We base our approach on the work of Adlin et al. (2007) who defined a five step model used in generating security personas. This work builds on the concept of using design personas to create personas that can be used for modelling various security interactions. We extend this concept to create personas for modelling cyber related agents enriching the model with attributes from computer generated personality insights.

The model uses five steps based on data gathered in the real world as follows:

1. Define the user population and gather background data on target users

   - Identify the target user population

   ![](image_url)

   *Figure 16: Agent Persona Creation*

   1. Define the user population and gather background data on target users
Create data collection plan

2. Collect user data related to the target population.

3. Transform data gathered into a fictional user or persona
   - Write-up the details into a persona narrative or description
   - Create a foundation document based on the narrative for each persona
   - Create lexical analysis
   - Create photos and names for each persona.
   - Cross-check each persona with "real world" users.
   - Cross check with other data sources that the persona is realistic

3.11.1 Case Study: Proton Sellers on Darknet Markets

The Proton project structure and timetable leaves little room for interaction between WP3 and other work packages hence the need to find a format for WP3 output that is consumable for the remaining work packages. We have designed the agent persona approach with the background that there are differences in researcher profiles between WP3 (mostly computer science/cyber security researchers) and other work packages (mostly social science/criminology researchers). We believe that a persona based approach is a good tool for bridging these specialist areas. Our initial case study is in the domain of organised crime and the darknet. More specifically sellers using darknet markets.

Step One: Defining a User Population and data collection plan

The first step is to decide which user population to select for the study. There are a number of ways that organised crime use the darknet. The inspiration from selecting sellers on darknet market places came from the book xxx – in which it the author made the point that the type of person that sells drugs on an online darknet market place is very different from the seller of drugs in the physical world. This insight opened is interesting in that it gives the opportunity of comparing what should be two very different sets of users using IBM’s personality insights tool.

After defining the target users we create a data collection plan for this set of users. If the user group is relatively unknown this is generally an iterative process.

The phenomenon of darknet market sellers is relatively new. This means that academic literature on this subject is very scarce. In addition to this, the fact that our target user group was very careful to remain anonymous also meant that direct data collection was not possible. The data sources left to us were essentially recent books, online discussion forums and electronic news articles. The first step in data collection was to review these sources in order to build up a picture of our target group. It became clear from this review that our target
users were shy, educated and technically very savvy. When it came to collecting text for lexical analysis we managed to locate a number of anonymous interview transcripts.

A typical data collection plan include
Background research
   Use Search engines to create an initial overview
Collect current books and magazine articles,
   Understand where the target users ‘hang out’ in cyberspace
Academic publications
Interviews with SME’s
   Select the target online media
      • Cyber bulletin boards/chat forum
      • Interview transcripts
      • Chat forums
      • Media comments
Infrastructure plan for collecting data

Step 2 Collect User Data related to target population
The methodology for this process was developed out of our initial attempts at collecting cyber data. We collected hundreds of millions of social media data items looking for generic ways to filter and cluster data. Two things quickly became clear:
   1.) It became clear that the amount of data required some form of coarse grained filtering in order to be able to analyse the data in a reasonable amount of time.
   2.) In order extract to cohort text for lexical analysis we needed a finer grained filtering.

In order to help with coarse grained filtering and fined grained clustering it was helpful to have an initial version of the persona that detailed characteristics such as the form and locations of cyber activity.

Our data collection plan evolved to the something like the following:

**Background research**
Our background research is in part documented in the reference section of the document. We used this initial research to create an initial version of persona prior to the foundation document described in the next section.

**Collect current books and magazine articles,**
Understand where the target users ‘hang out’ in cyberspace

**Academic publications**
Formal review of academic literature as carried out in Task 3.1.

**Interviews with SME’s**
We interviewed a number of law enforcement organisations

**Cyber footprint**
Selecting the target online media based on the background research. This may include:
- Cyber bulletin boards/chat forum
- Interview transcripts
- Chat forums
- Media comments

**Infrastructure plan for collecting data**
- Based on our problems with platforms such as Twitter we evolved some of our tools to analyse Reddit (Forum) and Youtube (Video platform that allows comments)

Our initial data collection experiments were generic in the sense that we trying to identify the signposts between different cyber domains. It was only after we created an initial draft persona that we were able to focus at the appropriate level.

**Step three: Transform data into a persona**
This step is used to create what Adlin et al. call a foundation document. The aim of this document is communicating the persona to all of the stakeholders in a project. The PROTON project is interesting in that the project stakeholders are very diverse. On the one side cyber-crime is very much the domain of computer scientists and cyber security practitioners. It became clear during the first year of the project that the cyberspace ecosystem, its agents and the technology appear very different to other crime domains.

**Stakeholders**
There are two sets of stakeholders for the PROTON project: The first are the participants that will implement and evaluate the agent based modelling experiments in subsequent years of the project. The second are the SME organisations interesting in results of the project.

Our agent persona concept extends the approach of Adlin et al to add concepts of the threat agent library (TAL) defined in the TARA cyber risk assessment methodology. In addition to the initial foundation document we add a number modelling attributes with simple metrics.

**Initial foundation document**
An initial foundation document sets the baseline around which to develop the persona. It is a simple description together with an image that should be immediately identifiable with all stakeholders in the project. It is against this basis that more specific attribute categories are added at later stages.

The following table shows the initial persona for our darknet drugs seller.
It may require a number of iterations to come up with a persona that is specific enough for the purpose. For example darknet markets are not the only platform for selling drugs on the internet. If the persona was to include persons behind such platforms this would include a wide range of organisations and motivations.

As mentioned previously, if the persona is relatively new it may be necessary to develop the foundation document before it is clear what data is to be collected.

### 3.11.2 Cyber Behaviour

The cyber behaviour attributes uses simple metric scales derived from the TARA approach.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Media Usage</td>
<td></td>
</tr>
<tr>
<td>External Country</td>
<td>5</td>
</tr>
<tr>
<td>Darknet Usage</td>
<td></td>
</tr>
<tr>
<td>Extra Legal Medium</td>
<td>4</td>
</tr>
<tr>
<td>Clearweb Usage</td>
<td></td>
</tr>
<tr>
<td>Organisation</td>
<td>4</td>
</tr>
<tr>
<td>Cryptocurrency User</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

Scale: 1 to 5 (1=Low, 5 = High)

To make the scale simpler to user different levels are allocated specific meanings, for example:

1. Code of Conduct
2. Legal
3. Extra-legal minor
4. Extra Legal medium
5. Extra-legal major
The selection of attribute metrics does involve some research together with SME input. The advantage of collecting these metrics in an agent persona is that the metrics are embedded in a rich context that make facilitate sharing and library creation.

### 3.11.3 LEXICAL ANALYSIS

The following analysis is the output from the Watson Personality Insights solution based on the transcript of an interview with the former owner of the silk road dark web market place. The interview was given anonymously before he was identified and convicted.

![Personality Insights from single person](image)

**Figure 17: Personality Insight from single person**

Obviously it is challenging to carry out individual analytics for known users. It is therefore fortunate that a number of persons have given interviews anonymously. The text from these interviews we can combine to generate a cohort perspective. The following is the high level output from an analysis on two such interviews.
Figure 18: Aggregate Personality Insight form multiple persons

The following diagram is a sunburst visualization of the cohort text showing the individual dimensions that the model provides.
This task has shown that signposts between the clear and the darknet do indeed exist. It has developed an approach to create personas that help focus the data collection activities.
4 Task 3.3: Innovative study - online visibility and social media impact of gangs

The work in this task aims at measuring the social media presence of European gangs, and the analysis of the content these groups produce in the web, with a focus on social media. In the initial planning phase of WP3, a collaborative online-search during a workshop took place in order to find relevant target groups (i.e. gangs) which are suitable for the studies intended. It was decided to put the focus within this task on outlaw motorcycle gangs. The reasons for choosing these groups were:

- They maintain a relatively large presence in social media (see section 2.6.2.1.2), in particular on Facebook
- They show frequent activity on social media, in particular on Facebook
- They are associated or have connections to organized crime

We followed the approach of analysing the content of these social media precenses automatically, by the use of the data analysis and collection system. Four different outlaw motorcycle clubs were examined for the studies conducted in this task:

- **Hells Angels MC** – probably one of the most famous internationally acting outlaw motorcycle club with branches in 56 countries worldwide (40 countries in Europe)
- **Bandidos MC** – international acting outlaw motorcycle club with branches in 32 countries\(^22\) (14 countries in Europe)
- **Mongols MC** – international acting outlaw motorcycle club with branches in 12 countries\(^23\) (3 in Europe)
- **Osmanen Germania** – Turkish-national outlaw motorcycle club with the main focus of operations in Germany. Compared to the other clubs with a long tradition, this club is quite young, approximately 3 years.

The selection of these clubs as the area of study is based on their activity in Europe, their different sizes and their time of activity to evaluate if this has an effect on the social media presence. The social media platform Facebook was chosen as the primary target for data collection, as all of the clubs (and the local subgroups) listed can be found there, and the accessability of the data through a web-interface which makes an automated collection and analysis of the data possible.


\(^{23}\) [https://www.mongolsmc.com/chapters/](https://www.mongolsmc.com/chapters/)
Within the following sections, the approach for the measurement of the presence and the analysis of the online content will be described.

4.1 Measuring the social media presence of gangs in Europe

To measure the social media online presence of outlaw motor cycle gangs, one approach is the comparison of the total number of existing subgroups to the number of subgroups (charters) maintaining a social media presence to get an understanding how common social media usage is amongst the various subgroups.

Regarding the example of the *Hells Angels* MC, the information on the existing chapters can be collected directly online, as this group maintains an official web presence\(^\text{24}\). On this web presence, all different chapters within the world are listed (c.f. Figure 20).

\(^{24}\) [http://hells-angels.com](http://hells-angels.com)
Figure 20: The web presence of the Hells Angels MC showing the various chapters over Europe (top), and within the particular countries (bottom).

The information available on this web presence can be extracted and analysed automatically, which enables an observation over time. Within this task, we implemented a web crawler which is capable of parsing the websites listed in Figure 20) and extracting the name of the cities where a presence of the club exists (i.e. Hells Angels MC Berlin Central). From this information, a list of available clubs can be generated and serve as a foundation for an online search within social media. Within the technical system (section 3.1.3), the data collectors for crawling data from the social media platform Facebook were used to automatically search for a presence of each of the clubs from the list. The other motorcycle clubs examined in this task do not maintain a web presence which makes it possible to extract a list of all charters. For these clubs, the facebook presences had to be determined using a google search and the analysis of the search results to extract the relevant URLs referring to the Facebook pages of these clubs.

<table>
<thead>
<tr>
<th>Club</th>
<th>Number of Facebook presences found</th>
<th>Number of charters in Europe</th>
<th>Number of charters in Europe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hells Angels MC</td>
<td>38</td>
<td>267</td>
<td>15%</td>
</tr>
<tr>
<td>Bandidos MC</td>
<td>20</td>
<td>81</td>
<td>25%</td>
</tr>
<tr>
<td>Mongols MC</td>
<td>11</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Osmanen Germania</td>
<td>11</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

For the four clubs from the target group, we were able to identify 89 Facebook presences for individual chapters. The presences are all set up as a Facebook
Page, which is according to the terms of service\textsuperscript{25} the appropriate type for Facebook presences sharing content for profit or not using an individual’s name.

\section*{4.2 Analysis of gang-related online contents}

\subsection*{4.2.1 Approach}

Using the data Facebook crawling system, for each of the clubs found on Facebook, the social graph is queried and analysed. Information on the number of the Facebook friends of the club and their available data will be collected and categorized, i.e. the “likes” of all Facebook friends of a club. Furthermore, it will be investigated if texts of sufficient quality can be extracted for the analysis of personality traits. The list of friends from each club will be analysed for the connections to other outlaw motorcycle clubs.

The primary way for apps to \textbf{read and write} to the Facebook social graph is the Graph API\textsuperscript{26}.

The Graph API provides a way to get data out of, and put data into, Facebook’s platform. It’s a low-level HTTP-based API that can be used to programatically query data, post new stories, manage ads, upload photos, and perform a variety of other tasks that an app might implement.

The Graph API is named after the idea of a 'social graph' - a representation of the information on Facebook composed of:

- **nodes** - basically "things" such as a User, a Photo, a Page, a Comment
- **edges** - the connections between those "things", such as a Page’s Photos, or a Photo’s Comments
- **fields** - info about those "things", such as a person’s birthday, or the name of a Page

Almost all requests are passed to the API at graph.facebook.com - the single exception is video uploads which use graph-video.facebook.com.

The easiest way to understand the Graph API is to use it with the Graph API Explorer, a low-level tool you can use to query, add and remove data. It’s a very handy resource to have at your fingertips while you integrate with Facebook\textsuperscript{27}.

By default, not all fields in a node or edge are returned when a query is made. Fields or edges can be choosen and are returned with the \texttt{fields} query parameter. This is useful for making API calls more efficient and fast.

\begin{itemize}
\item \textsuperscript{25} https://www.facebook.com/legal/terms
\item \textsuperscript{26} https://developers.facebook.com/docs/graph-api
\item \textsuperscript{27} https://developers.facebook.com/tools/explorer
For example, the following Graph API call using your own user access token https://graph.facebook.com/me?fields=id,name,picture will only return the id, name, and picture in your profile. A Common Scenario for using the Graph API to query data are is to query if two people are Facebook friends. This can be determined without having to parse their entire list of friends using the API-call.

- `/{user-a-id}/friends/{user-b-id}`

### 4.2.2 STRUCTURE OF THE DATA

The data for each of the specific subgroups was acquired through the Facebook API. The facebook pages hereby represent the the basic nodes. Furthermore, the connected nodes “photos” and “videos” were analysed. The following list shows the nodes and the according fields accessed by the technical platform in order to acquire the data.

- **Page ID**
  - About
  - Artists_we_like
  - Bio
  - Category
  - Description
  - Fan_count
  - General_info
  - Impressum
  - Instagram_business_account
  - Location
  - Hometown
  - Influences
  - Members
  - Mission
  - likes

- **Photos**
  - Event
  - Created_time
  - Album
  - Likes
  - comments
  - Comments

The data was choosen in order to get a better understanding about the information the specific subgroups are sharing with a public audience. The acquisition of textual data was one of the intended goals which led to the
selection of the fields “description”, “general_info”, “bio” and “mission”. Through the selection of the fields “influences” or “artists_we_like”, it was envisioned to gain knowledge if the subgroups maintain a connection to people known to the public or if they in general identify with other people, ideologies or philosophies.

4.2.3 RESULTS OF THE DATA ANALYSIS

The data crawled for the 80 subgroups showed that not all of the fields were either accessible through the API due to privacy settings, or not given by the administrators of the subgroup pages. For the fields “artists we like”, “bio”, “general_info”, “instagram_business_account”, “hometown”, “influences” and “members”, no data could be obtained. The “mission” field contained data for only 2 subgroup pages, saying “Motorcycle Club” in both cases.

From the 80 pages crawled, 44 of them contained information within the “location” field, given information on the geographical location of the subgroup, most of them including postal address and latitude/longitude. From the Mongols MC club, only the subgroups from the USA and Thailand gave full information, the other subgroup pages didn’t provide any useful information here. The page of Mongols MC Sweden contained explicit language in the “city” information, and the code “13!12” within the ZIP code information (which is a common synonym for the abbreviation ACAB28).

The page of the “Hells Angels Motorcycle Club Stuttgart” contained extensive information within the “impressum” field, only two Bandidos MC pages contained information regarding the club’s address or copyright information regarding the page content.

The field “fan_count” represents the number of Facebook users who like the particular page, and thus gives a rough estimate on the range of particular page. As it can be seen within Figure 21, most of the pages examined have a fan count value between 0 and 2000.

---

For the field “description”, 11 from the 80 datasets contained textual information. The information was given in the languages English, German, French, Italian, Swedish and Arabic, which could be detected automatically with the technical platform. The following table gives an overview of the languages and word counts found in the description fields.

<table>
<thead>
<tr>
<th>Language</th>
<th>Word count</th>
</tr>
</thead>
<tbody>
<tr>
<td>de</td>
<td>28</td>
</tr>
<tr>
<td>sv</td>
<td>41</td>
</tr>
<tr>
<td>fr</td>
<td>22</td>
</tr>
<tr>
<td>it</td>
<td>509</td>
</tr>
<tr>
<td>en</td>
<td>194</td>
</tr>
<tr>
<td>en</td>
<td>47</td>
</tr>
<tr>
<td>en</td>
<td>6</td>
</tr>
<tr>
<td>en</td>
<td>3</td>
</tr>
<tr>
<td>en</td>
<td>7</td>
</tr>
<tr>
<td>de</td>
<td>6</td>
</tr>
<tr>
<td>ar</td>
<td>2</td>
</tr>
</tbody>
</table>
The field “category” was filled with information by 80 subgroups. Different categories were chosen for the presence by the individual subgroups:

- Community (24)
- Organization (13)
- Country Club / Clubhouse (7)
- Nonprofit Organization (6)
- Community Organization (6)
- Local Business (4)
- Pub (3)
- Private Members Club (3)
- Motorcycle Dealership (3)
- Social Club (2)
- Sports Team (2)
- Political Organization (1)
- Personal Blog (1)
- Cafe (1)
- City (1)
- Lounge (1)
- Hotel & Lodging (1)
- Non-Governmental Organization (NGO) (1)

Using the technical platform, we were furthermore able to access the “photos” node for each individual subgroup page. The data available to the public which could be accessed through the Facebook API was the photos themselves, as well as the creation time and date of these photos, which means the time when the photos were added. The information on the creation time could for example be used to get information on the activity of a specific subgroup. It was possible to crawl 229 photos from 66 different subgroups. The photos were posted within the years 2013 to 2018. The following table shows the number of photos posted from all pages examined per year.

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of photos</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>5</td>
</tr>
<tr>
<td>2014</td>
<td>27</td>
</tr>
<tr>
<td>2015</td>
<td>72</td>
</tr>
<tr>
<td>2016</td>
<td>63</td>
</tr>
<tr>
<td>2017</td>
<td>55</td>
</tr>
<tr>
<td>2018</td>
<td>7</td>
</tr>
</tbody>
</table>

The content of the photos crawled could be analysed using common image analysis services like IBM Watson Visual Recognition\(^29\) or the Google Vision API\(^30\). These systems are able to determine a set of labels classifying the

\(^29\) https://www.ibm.com/watson/services/visual-recognition/
\(^30\) https://cloud.google.com/vision/?hl=de
image content. The following figure shows the analysis result for a given sample image:

<table>
<thead>
<tr>
<th>Category</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vehicle</td>
<td>89%</td>
</tr>
<tr>
<td>Mode Of Transport</td>
<td>88%</td>
</tr>
<tr>
<td>Black And White</td>
<td>87%</td>
</tr>
<tr>
<td>Product</td>
<td>81%</td>
</tr>
<tr>
<td>Car</td>
<td>69%</td>
</tr>
<tr>
<td>Monochrome Photography</td>
<td>66%</td>
</tr>
<tr>
<td>Monochrome</td>
<td>55%</td>
</tr>
</tbody>
</table>

Using this approach, the content of a specific subgroup’s photo gallery could be categorized into a set of classes. As this might arise issues regarding privacy, we did neither store the images nor conducted any image recognition analysis. However, with the technical system implemented, it would be possible to do this, under the condition the legal conditions are clarified.

### 4.3 Summary

Within this task, we have shown how Outlaw Motorcycle Gangs, as an exemplary group for criminal gangs in general, use the Facebook social network to distribute various kind of information and to show their presence in general. It turned out that there is only a limited amount of data available to the public. However, there are possibilities to extract and analyse this data in an automated or semi-automated way, which has been demonstrated in this work. The legal setting, as well as the terms of usage of the social network platform used as information source has always kept in mind when performing this kind of investigations.
5 Task 3.4: Innovative study - Radicalisation in cyberspace and radical social media networks

T3.4, led by HUJI, improves on the existing knowledge relating to the link between the internet and radicalization outcomes. Specifically, this task examined different types of social media activities and behaviours that may be indicators of online radicalization and recruitment. Moreover, the task sought to identify what patterns of social media activity and behaviour may indicate differences between non-violent radicals and those that go on to commit acts of violence.

This final report presents the analysis of 42 lone wolf terrorist profiles and how they compare with the profiles of 42 known radicals who have not committed acts of terrorism. Each lone wolf was matched with a known radical. The analysis examines online activity over a 100-day period prior to the attack. The mixed qualitative-quantitative analysis identifies that frequency, post type, radicalness of post and reach and reaction to posts all indicate differences between the two groups. Additionally, trend analyses identify that about 2 months prior to attack there is an identifiable shift in online activity and behaviour. The findings have practical applications for policy makers in that they identify additional parameters beyond content as possible indicators of the move from radicalization of belief to radicalization of action.

5.1 Introduction

Whilst social media has certainly become a tool for radicalization, security and intelligence forces have also found it to be a useful tool in combating radicalization and identifying potentially violent individuals before they offend. Indeed, social media has become a battleground against terrorism that crosses the borders of the digital and real worlds, leading to a host of new challenges, such as the balance between efficiency and human rights. While security services have certainly used social media as a source of intelligence for successful preventions, over-policing has lead to many false arrests, whilst simultaneously missing threats that actually lead to violence. It has been suggested that more evidence based approaches would help to improve effectiveness overall. While many advanced systems may be in use, neither human or machine actually know what types of activities or online behaviors
may differentiate potential threats from the large pool of non-violent individuals who may be espousing radical but protected views.

One of the most pressing challenges for counter-terrorism and anti-radicalization policies, strategies and tools is the balance between security needs and human rights. Many policies come under attack for impinging on the human rights of often completely innocent individuals or individuals whose activities are legally protected. The misidentification of specific factors can lead to a failure to identify those most at risk, whilst impinging on the rights of those who would have never turned to violence in the first place (Bertelsen, 2016). The ultimate goal of all policies and tools should therefore be how to limit false arrests and other rights impingements whilst maximizing efficiency in identifying and stopping actual threats. In general, only once a better understanding of specific factors is developed can the development of more effective and scientifically based policies and interventions be achieved (Hawkins, Catalano & Arthur, 2002; Blum & Ireland, 2004; Heath-Kelley, 2013; Monahan, 2012; Horgan & Taylor, 2011; Bhui et-al, 2012; Cherney, 2016).

According to Weine et-al (2017), the entire countering extremist violence (CVE) field is lacking from evidence based approached. In this context, risk assessment tools and online identification algorithms may claim to be advanced but don't actually know what factors to be looking for. In recent years, while intelligence and security services have successfully leveraged social media to stop many potential threats, this has led to many false positive identification, failure to identifies (Grothoff & Porup, 2016; Parra-Arnau, Claude Castelluccia, 2015), claims of false arrests and privacy protection and other legal violations (Hirschauge and Shezaf, 2017; Werbin, 2009; Rosenzweig & Jonas, 2005). In order for more effective solutions to be developed, more evidence is needed. Since only a very small percentage of radicals will ever turn to violence, the challenge is how to identify and single out such individuals from the non-violent radical milieu. The best approach for achieving an identification of terrorist specific risk factors is by comparing them with non-violent radicals (Frielich et-al, 2015; LaFree and Frielich, 2017). This study seeks to advance the evidence based knowledge on the types of activities that may differentiate these two groups. Such evidence can be used to reduce the number of false arrests whilst improving detection capabilities.

5.2 Theoretical framework

According to the policies and definitions of the EU, there is a fundamental difference between two potential outcomes of radicalization. Whereas radicalization is itself "the phenomenon of people embracing opinions, views and ideas which could lead to acts of terrorism" (EU, 2005), recruitment is where an individual has somehow been convinced to act on these opinions and ideas and provide actual support for, facilitate or engage in violent actions (EU, 2014). In this regard, the EU defines recruitment as meaning “to solicit
another person to commit or participate in the commission of a terrorist offence, or to join an association or group, for the purpose of contributing to the commission of one or more terrorist offences by the association or the group” (EU, 2005) and both “Recruitment to carry out terrorist offences” as well as actual “recruitment into a terrorist group” (EU, 2014). It follows that recruitment, is a natural developmental outcome of radicalization for some small percentage of radicalized individuals. Perhaps the greatest challenge for security services is to determine and identify who is on the path from radicalization to recruitment and to separate them out from the larger pool of non-violent radicals.

Whilst there are many competing theoretical frameworks for examining radicalization, the two pyramid model approach closely resembles the distinctions made in policy. According to the two pyramid framework, there are different levels of radical belief and radical action respectively. According to this model, those with radical beliefs who are situated at the top of the pyramid, and who have a sense of individual moral obligation towards a radical doctrine, are at the greatest risk of moving to radical action. However, there are also different degrees of radical action. Whereas some actions may be legal, such as protests, activists may feel unsatisfied and move to become terrorists (McCauley & Moskalenko, 2008, 2014, 2017). Other models follow a similar approach, distinguishing between cognitive and behavioural expressions of radicalization (Neumann, 2013; Khalil, 2014; Hafez & Mullins, 2015). These distinctions mirror the different outcome of 'inert' radicalization and actual recruitment or involvement in a terrorist offence.

![Figure 1. Opinion radicalization pyramid.](image1)

![Figure 2. Action radicalization pyramid.](image2)

The two pyramid radicalization model

While the two pyramid model provides the framework for differentiating between the different possible outcomes, it does not provide a framework for specifying and examining the mechanisms through which individuals shift between the different stages or outcomes, or for how to identify differences between individuals from the different categories. To this end, this study takes
a social learning approach to online radicalization. One of the advantages of social learning approaches is that social learning theory's components are easily converted into testable variables (Pauwells & Schils, 2016). It is perhaps for this reason that social learning remains one of the most popular and most proven of the criminological theories. Additionally, the application of a social learning approach to the online learning of deviant behaviours more generally already has demonstrable utility and success (Holt & Bossler, 2014). This is because online radicalization isn't the product of simply watching a few videos, rather it is one potential outcome from having been immersed in radical content and engaging frequently with like-minded individuals (Neuman, 2013; Keene, 2011).

According to social learning theory, more frequent and intense interactions over longer durations increase the likelihood that shared behaviours will be adopted. Many believe that the passive/active nature of the internet amplifies this process. There is strong evidence that in general, differential associations play an important role in the learning of deviant behaviours online (Warr, 2002; McCuddy & Vogel, 2015a, 2015b; Pauwells & Schils, 2016). Differential associations have been found to be an important risk factor for involvement in a range of cybercrimes (Holt, Burrus & Bossler, 2010; Chua & Holt, 2016; Higgins et-al, 2006; Morris & Higgins, 2010). Increased frequency of interactions with differential associations can lead to an "online disinhibition" that increases the likelihood for online learning to be translated into real-world violence (Davies et-al, 2015; Ducol et-al, 2016; Suler, 2004). This "digital drift", where online learning crosses over to real world offending (Goldsmith & Brewer, 2015), can be identified in a range of behaviours, such as pedophilia (Beech et-al, 2008), school shootings (Smithey, 2016; Oksanen, Hawdon & Rasenen, 2014), and lone wolf terrorism (Akers & Silverman, 2004; Goldsmith & Brewer, 2015). Indeed, studies have found that up to 38% of paedophilic offenders were part of online paedophile networks (D’Ovidio et-al, 2009; Alexy, Burgess, & Baker, 2005; Seto & Eke, 2005; Mancini et-al, 2014) while school shooters (Oksanen, Hawdon & Rasenen, 2014) and lone wolf terrorists (Weimann, 2012) are also active in radical online networks. Social learning provides an approach to understanding how the development of the radical beliefs and actions occur in online radicalization (Akers & Silverman, 2004; Friedburg & Crane, 2011; Armstrong & Matusitz, 2013; Akins & Winfree, 2017; Pauwels & Schills, 2016; Peterson & Densley, 2017).

The current study applies and tests the integrated framework, combining the two pyramid approach and social learning theory by examining a range of social media level metrics derived from social learning variables. The study examines the way in which different types of differential associations are observable in the social media profiles of lone wolf terrorists and their potential influence on behavior. Individual components of differential associations, namely frequency and intensity are also examined by analyzing the frequency and patterns of behaviors and the type of material and content associated with different radicalization outcomes. The study also analyzes different types of online behaviors and interaction types and how these may be connected to
imitation and expression. Differential reinforcements are also examined in the context of analyzing the reach and reactions of the online activities of radicals.

5.3 Data and methodology

This study examined the profiles of 42 lone wolf terrorists who committed attacks in Israel between 2015-2017. While there were additional profiles identified, missing data and coding issues meant that only 42 could be included in the final analysis. Using Facebook's advanced graph search functions, each profile was matched with a non-violent radical from the same network and based on age, gender and place of residence. Where possible, matches were made based on the additional criteria of educational institution, workplace and/or organizational affiliation. For most profiles, only 1-2 potential matches of open profiles were identified. Where 2 or more potential matches were identified, the researchers attempted to select the profile which appeared to be qualitatively more similar. Each matched profile was backdated to be coded from the same date as the radical action group profile to which they were matched. This enabled the analysis to control for confounding environmental and temporal factors. A team of Arabic speaking researchers carried out the searches, data extraction and content analysis.

Following the matching process, for each activity the measures on 4 types of metrics were manually extracted and coded using a data extraction tool, a coding key, and a quantitative content analysis tool developed by the research team:

- Date and time
- Activity type: 10 types of activities including text posts, images, videos, shares etc.
- Level of radicalness: 0-4 scale based off of a quantitative content analysis tool
- Reach and reactions: Likes, comments and shares were coded separately

The proportion of fully open and public profiles identified is somewhat less than the average number of open profiles on Facebook among the general population (20%). However, it should be considered that terrorist profiles are often systematically removed from Facebook following attacks either at the request of authorities, or due to Facebook or even family members' intervention. While Facebook has stepped up its efforts in this regard recently, it seems that the more serious attacks are more likely to have their profiles removed more quickly. While even fully open profiles cannot provide data on every type of action taken on the platform, they do provide full information of all activities that occurred on the individual's own profile/page. To the best of the researchers' knowledge, this is the largest and richest data set of this kind. Future analyses will ideally be based on larger datasets.
5.4 Results

Of the 42 radical action group members, all of them were male. The age range of the group members was diverse, with the youngest attacker being 15 and the oldest 57. However, the mean age was 21 and this is representative of the group. The split in radical action type was, stabbing/bladed weapon attack (45%), driving/vehicular attack (19%), firearms (9.5%), combined attack (9.5%) explosives (4.8%), other (7.1%). The majority of the attackers showed indications of radicalization (92.6%), with only three attackers not displaying such signs, however, only 11 attackers (26.2%) had clear indications of intentions to attack prior to attack.

5.4.1 Differential Associations

It is known that for many social media users, their closest online associations are also offline associations, often close family and friends. This effect has been noted by many studies and shows the important cross-over that exists between the offline and online worlds (Subrahmanyam et-al, 2008; Hampton et-al, 2012). In this regard, social learning theory posits that those associations which have greater 'priority' and 'duration' will have a greater influence on behavior (Akers, 2008). Family members and close friends therefore have an especially strong influence on beliefs and behaviors, including through online social network structures (Pauwells & Schills, 2016). While the influence of wider network structures are certainly important in online radicalization, close connections have greater priority and the duration of the relationships also increase the intensity of the effect they have on behavior. Those who have radical friends are at a high risk for engaging in radical actions themselves (Pauwells & Schills, 2016; Pauwells et-al, 2014; Jasko et-al, 2016; LaFree et-al, 2017).

In this study, exactly half (50%) of the actions group attackers had connections with family members who had carried out attacks prior to their own and with whom they were online association. However, 19 of them (45.24%) had online associations with friends who had carried out attacks prior to their own. In only two cases did the attacker have differential associations from both categories. As such, for only 3 attackers the research team was unable to identify online associations with family or peers who had previously attacked. In half of the 39 cases for which differential associations were identifiable, the attackers shared at least one picture, video or post pertaining to the attacks of their friends or family members in the 100 days prior to attack. These attacks appear to have also occurred in close temporal proximity to each-other. Of the 42 attackers, 4 of them were from the same wider network of friends, who are also connected by another 6 of those in the matched comparison group. It should also be noted that close differential associations with both family and friends online are also reflected in offline
spatial characteristics. That is, differential associations generally live in the same locales and neighborhoods.

This is quite an important finding which likely has implications in other settings, such as the EU. In the case of the EU, recent lone wolves and foreign fighters have shown that small networks produce multiple actors. These actors will work together or separately, often with little temporal proximity between them. In Europe, like in Israel, lone wolf attackers are often connected by being friends from the same small neighborhoods. A small number of specific neighborhoods can produce a disproportionately large number of radicals (Marone, 2017; Van Vlierden, 2016; Linus & Magnus, 2017). Overall, differential associations have been found to have high correlations with differentiating between non-violent and violent radicals (Pauwells & Schills, 2016; Pauwells et-al, 2014; Jasko et-al, 2016; LaFree et-al, 2017). This finding appears to be in line with what is known about the effects of deviant online peers with respect to other forms of cyber related deviance (Holt & Burruss, 2010). Lone wolves are likely to be connected to each other through various direct and indirect online ties (Klausen et-al, 2016) and this appears to be supported by the current findings.

5.4.2 FREQUENCY

According to social learning theory, frequency is one of the most important key functionaries of differential associations, together with duration, intensity and priority. Generally speaking, frequency refers to interactions with individual associations, however it can also relate to frequency or amount of time spent within a network. Since differential associations includes media consumption, frequency also refers to the how much time is spent consuming media (Hawdon, 2012; Pauwells & Schills, 2016). Increased frequency of interactions increases the speed at which shared beliefs and behaviors may be adopted. More frequent involvement in online networks can lead to polarization of beliefs, which in turn can lead to radicalization.

In this analysis it was identified that the radical action group posted more frequently than the radical beliefs group on a per day basis over the 100-day period. However, it is important to note that the radical belief group also posted more frequently than the average user (Hampton et-al, 2012). This does not necessarily mean that sheer frequency is a sole determinant of the potential negative effect of differential associations. For example, in Ellis et-al's (2015) study of Somalian immigrants in North America found that the highest frequency users were the least likely to have radical beliefs. There is probably a tipping point at which frequency can act to either protect or radicalize but this will be dependent on other determinants such as patterns of usage as will be discussed below. However, others have found that increased internet and social media usage may be associated with anti-social and violent behaviors (Fisoun et-al, 2012). In this regard, increased frequency of
Facebook use has been found to be associated with increased loneliness and lower social well-being (Pittman & Reich, 2016).

Table 32 Average posts/day for all group members (N=84)

<table>
<thead>
<tr>
<th>Posts/day Radical Actions</th>
<th>Posts/day Radical Beliefs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean= 0.61619*</td>
<td>Mean= 0.411428571</td>
</tr>
<tr>
<td>SE= 0.141319</td>
<td>SE= 0.5509</td>
</tr>
</tbody>
</table>

Following attacks in Europe and elsewhere, it is often revealed that attackers were active users of social media (Weimann, 2012, 2016; Pantucci, 2011). While there are thousands of radicals online, it seems that those who go on to attack are generally speaking, more active in their posting activities. Important differences in usage of frequency may also effect the extent to which individuals may fall under active surveillance. Previous studies have found that the most active of radical social media users were more likely to be caught (Meloy & Gill, 2016). Additionally, the most active radical on Twitter were ISIS supporters and recruiters, essentially non-violent roles (Berger & Morgan, 2015). While the results of the current analysis don't show statistically significant differences in posts per day between the violent and non-violent radical groups, patterns in frequency may be able to differentiate between the two groups (see section 1.4.6).

5.4.3 RADICALNESS

According to social learning theory, the intensity of differential associations, which includes engagement with online materials and content items, is believed to influence the likelihood that an individual will adopt the shared behaviors (Pauwells & Schills, 2016). Studies have found that exposure to hate content on social media is connected to decreased social trust (Nasi et-al, 2015). Exposure to radical social media content is also correlated with increased propensity for justifying radical violence (Pedersen et-al, 2017; Ducol et-al, 2016), and for engaging in violence more generally (Phillips, 2017).

This study used a quantitative content approach for this part of the study (as in Ducol et-al, 2016; Jensen & LaFree, 2016; Smith 2004; Smith et-al ,2008). The tool was developed by the research team and based on expertise in the field of online radicalization. The tool consists of 3 levels of radicalness and codes a 1 for non-radical or otherwise innocuous content. Radical 3 posts had to have used direct or specific calls for violence or provide specific religious justifications and instructions or incitement to carry out radical violence or include specific displaying of support for terrorists and prescribed terrorist organizations. These also included pictures of death and extreme violence, including pictures of dead terrorists and victims. Low level radical content, or Radical 2 posts, are those posts where the content depicts low-level violence
or depicts weapons, especially in the form of the user displaying weapons and there being a radical context to the display. Radical 2 posts also included but were not limited to general Islamist imagery, general updates regarding terrorism, specifically radical quotes and religious quotes clearly displaying radical intentions.

Low level radical content appears to be the most common type of material posted by both the radical beliefs and radical actions groups. However, there are clear differences in the proportion of other types of posts.

![Table 33 Proportion of posts/level](image)

<table>
<thead>
<tr>
<th>Radicalness</th>
<th>Radical actions</th>
<th>Radical beliefs</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>21.76%**</td>
<td>13.46%</td>
</tr>
<tr>
<td>2</td>
<td>61.2%</td>
<td>49.76%</td>
</tr>
<tr>
<td>1</td>
<td>13.65%</td>
<td>32.68%**</td>
</tr>
</tbody>
</table>

**p<0.05

The radical beliefs group was significantly more likely to post completely innocuous content; especially personal content such as pictures with family, with these types of posts representing almost double the proportion compared to the radical actions groups. The radical actions group generally displayed some element of radicalism with each post, with relatively few posts not being somehow related to radical ideology (13.65%). This group was more likely to post more intense and radical content. However, radical belief group members also were active in the posting and sharing of content depicting deceased 'martyrs', especially those with whom they had direct connections.

5.4.4 POST TYPE

As is known, there are many different types of online activities, just as there are offline. While communications may be made face-to-face, one-on-one, in group settings, over the phone, over the internet etc., online activities are just as diverse. Facebook provides users the option to create their own context, as well as share others' and provide reactions to others'. These types of activities can be categorized as active and passive. It is believed that the active/passive nature of the internet and social media make its effects on social learning greater than only passive forms of association and consumption such as television (Bandura, 1978; Pauwells and Schills, 2016). In this analysis, only active forms of activity were analyzed and these included text, image and video posts, as well as text, image, video and link shares.

The findings of this analysis may be in line with what some have theorized regarding social media being an outlet to voice grievances and that by access to such an outlet provides a sort of protective factor (Helms, Erin York, Peter Chalk, 2012). This means that for some, being part of a radical online network
may be a protective factor, whilst for others it is a risk factor and perhaps an important source in their radicalization (Özdemir & Kardas, 2014, 2018). This also needs to be better understood however in the context of the online-offline interaction effect. It is believed that for most users, their most active online associations are also offline associations, with such associations regularly being family members and close friends. For others, there may be little connection between their online and offline networks (Subrahmanyam et al., 2008).

The results of this analysis also provide evidence for the effects of imitation. By having such a high volume of shared material and a lack of original content, the radical actions group displays significantly more imitation than the radical beliefs group (Type 7 and Type 8 activities). This is even more interesting and important in light of recent reports regarding trends in social media usage more generally. According to recent reports, Facebook users have overall decreased in the proportion of original content being created, especially in the form of original posts, with text posts being replaced by images (Thelwall & Vis, 2017). Rather, sharing of images and videos has increased at the expense of original content creation (Edgerly et al., 2016).

<table>
<thead>
<tr>
<th>Type</th>
<th>Radical action</th>
<th>Radical belief</th>
</tr>
</thead>
<tbody>
<tr>
<td>1- Original text post</td>
<td>16.23%</td>
<td>29.15%***</td>
</tr>
<tr>
<td>2- Original image post/upload</td>
<td>42.21%</td>
<td>43.42%</td>
</tr>
<tr>
<td>3- Original video post/upload</td>
<td>1.81%</td>
<td>4.10%**</td>
</tr>
<tr>
<td>4- Link to post</td>
<td>0.87%</td>
<td>1.74%</td>
</tr>
<tr>
<td>5 –Shared text post</td>
<td>0.45%</td>
<td>0.28%</td>
</tr>
<tr>
<td>6 –Shared image</td>
<td>8.03%</td>
<td>4.34%</td>
</tr>
<tr>
<td>7- Shared video</td>
<td>9.05%*</td>
<td>4.39%</td>
</tr>
<tr>
<td>8- Shared link</td>
<td>2.44%**,</td>
<td>0.51%</td>
</tr>
<tr>
<td>9- Text based image</td>
<td>7.14%*</td>
<td>3.41%</td>
</tr>
<tr>
<td>10- Text within image</td>
<td>7.28%</td>
<td>4.39%</td>
</tr>
</tbody>
</table>

* p<0.1, ** p<0.05, *** p<0.01

Overall, for both groups the largest proportion of activities is the posting of images. One interesting observation here is that the overwhelming majority are original uploads as opposed to shares. However, the radical actions group has a much larger proportion of shares. Given that many of the images were evidently not created by the user, one possible explanation is that such images were received on mobile devices from another platform (e.g WhatsApp, Telegram etc.) and then uploaded from the user's device. Increased levels of sharing are however more characteristic of the radical action than the radical beliefs group. Sharing may indicate a degree of imitation, especially in the
form of images. Additionally, higher levels of sharing may be indicative of a larger proportion of time spent browsing and engaging in passive consumption. Additionally, the radical actions group was more likely to post test contained within images, which present problems for automatic detection systems generally.

It was observed that in addition to creating more original text posts, the radicals were far more likely to add their own caption to pictures, whereas the terrorist users were more likely to simply post the image without any added comments of their own. In this context, it has been found that frequent updates of this kind are linked to emotional support and stability (Hampton et al., 2012). However, other studies have found that sharing of images is more likely to improve emotional well-being than text based posts (Pittman & Reich, 2016).

5.4.5 DIFFERENTIAL REINFORCEMENT

According to social learning perspectives, the way in which beliefs and behaviors are reinforced will also determine the rate at which they are adopted. There are essentially two types of differential reinforcement, positive and negative (Pauwells & Schills, 2016). A positive reinforcement in the case of radicalization may read along the lines of 'I agree, we need to kill them', or 'if you act you will be rewarded'. A form of negative reinforcement may be along the lines of 'If you perform this action you will remove your sins'. Differential reinforcement may be especially strong in online social media networks of likeminded individuals. One reason for this is that dissent is abhorred in these insular and reinforcing community structures (Geeraerts, 2012; Hawdon, 2012). Indeed, while the researchers were unable to code every single comment, the original categories of positive and negative reinforcement were abandoned on account of no dissenting comments having been identified.

As can be seen in the table below, the radical actions group received a significantly larger number of likes per post and shares per post, although not comments per post. The average number of comments per post were almost identical for the two groups. The radical actions group had a larger amount of shares per post, however it cannot be discounted that post-mortem sharing accounts for a large proportion of this difference.

<table>
<thead>
<tr>
<th>Type</th>
<th>Radical actions</th>
<th>Radical beliefs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likes</td>
<td>71.59*</td>
<td>41.99</td>
</tr>
<tr>
<td>Comments</td>
<td>10.69</td>
<td>9.53</td>
</tr>
<tr>
<td>Shares</td>
<td>1.36***</td>
<td>0.16</td>
</tr>
</tbody>
</table>

*p<0.1, ***p<0.01
The posts that appeared to attract the most likes were innocuous and were often associated with the changing of a profile picture, updating of a personal status, or a birthday. The researchers attempted to control for the confounding factor of post-mortem comments by excluding those comments that were made after the date of attack. Unfortunately, this was not possible for 'likes'. It should be considered that in at least a few of the cases, a significant proportion of the likes were likely made by supporters of the terrorist following his/her actions. The findings are interesting in the context of what is known about average users' emotional responses to post reactions. Studies have found that receiving comments has a real impact on social support and connectedness (Scissors, Burke & Wengrovitz, 2016). Similar findings have been found with receiving of likes as well, with a greater number of likes per post correlating with increase sense of social support (Wohn, Carr & Hayes, 2016). The support offered by wider online networks is believed to contribute to online radicalization and this may be measured in the reach and response generated by posts.

5.4.6 PATTERNS OF ACTIVITY

Activity patterns need to be looked at closely as spikes in frequency in the days and hours leading up to an attack were only observably identified in one case of radical actions (T42). Rather, spikes in activity may be hardly identifiable based on sheer magnitude, since average posts per day are generally less than 1/day. Micro-spikes may be identifiable in the weeks and even months leading up to a possible attack. There were two identifiable outliers in the radical actions group data, with T42 and T30 displaying clear spikes in activity but also significantly divergent and more frequent activity overall. T42 was by far the most active, with T30's most active day still representing only half as many activities as T42’s most active day. In the case of the former, the researchers noted that by analyzing the time stamps of each post, T42 apparently did not sleep on this day. This was also identifiable for T30 although their posts were more spread out. While actively less frequent, the spaces of time in between posts were not long enough to have provided sufficient time for sleeping or working. As such, it seems quite likely that T30 spent the time in between actively posting with passive consumption and other forms of activity. As more evidence is gathered it may be identified that different trajectories exist and that T30 and T42 represent a possible alternative trajectory from what is identifiable among the other 40 radical actions group users.

Figure 1 Radical actions group activity frequency
After having removed these outliers, it appears that the radical actions group and radical beliefs group have some differences in the patterns of activity and non-activity that they display.

*Figure 2- Radical actions group activity frequency (without T30 & T42)*

*Figure 3- Radical beliefs group frequency*
Since each match is based on the same timeframe of the radical action group member to whom they are matched, more activity closer to the time of their association's attack may simply reflect their own response to the event. As mentioned above, radical beliefs group users regularly post images, videos and status updates pertaining to the attacks of their associations. Although there may be some differences that are observable to the naked eye, additional analyses were needed in order to draw proper comparisons between the two groups. One way to analyze for trends in cumulative time series like these is with the Mann-Kendall trend tests, which is an adaptation of a Whitney-Kendall U test. In these tests, the summary number of observations across the date span indicates how many times the sample acted on each day. Using a Mann-Kendall (MK) and Heterogeneity test method for this type of data has also been employed by Perrot (2014) in his work for the French Interior Ministry and intelligence services in developing evidence based crime prediction tools (also Perrot and Achi, 2015). More recently, Li et-al (2016) demonstrated the usefulness of such an approach for identifying trends in terrorism attack frequency in 16 middle eastern countries between 1970-2014. Using data from the Global Terrorism Database (GTD), this approach successfully identified shifting points and directions of trends for 10 of the 16 countries. In a recent paper, Malakar et-al (2018) demonstrate how a Mann-Kendall based method can be used to identify trends in specific key words used and trending on the social media platform Twitter.

In the MK analysis a trend was identified among the radical actions group but not among the radical beliefs comparison group. Among the radical beliefs comparison group Kendall's tau was not significant (p=0.150), indicating that there is no identifiable trends or pattern in the data. However, a clear and identifiable trend was found for the actions group (p < 0.0001).
Figure 4- Mann-Kendall trend tests

<table>
<thead>
<tr>
<th>Radical actions</th>
<th>Radical beliefs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kendall’s tau</td>
<td>0.707</td>
</tr>
<tr>
<td>S</td>
<td>3480.000</td>
</tr>
<tr>
<td>Var (S)</td>
<td>112680.667</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Following this, a series of heterogeneity tests were carried out in order to identify where changes in activity may be taking place for the radical actions group. Pettitt's test identified a significant mark at \( t=54 \) (\( K=2160.000, p < 0.0001 \)). A Standard Normal Homogeneity Test (SNHT) also identified \( t=54 \) was consistently identified as a significant time at which the pattern of activity experienced an identifiable change (\( T0=50.938, p < 0.0001 \)). Buishand’s test also identified \( t \) as \( 54 \) (\( Q=35.750, p < 0.0001 \)). The overall \( R \) was 39.372 (\( p < 0.0001 \)) and Von Neumann’s \( N \) was .458 (\( p < 0.0001 \)).

Figure 5- Radical action groups trend analysis

According to these results, there is a clear and identifiable change in activities, represented by a decreasing trend in activity at the \( t=54 \) (54 days prior to attack) interval.

A second series of the same analyses were repeated after having removed the two outliers. In the first instance the MK trend test identifies that there is still an identifiable trend (Kendall's tau = -0.341, S = -1654.000, Var (S) = 112399.333, \( p < 0.0001 \)). However, \( t \) moved up by ten days to \( t=64 \) in Pettitt's test (\( K=1470, p < 0.0001 \)) and Buishand's test (\( Q=21.137, t=64, p=<0.0001 \)). In the SNHT test however, although significant (\( T0= 42.152, p < 0.0001 \)), \( t \) was identified as 1. The downward trend is replaced by a slight upward trend until \( t \), followed by a downward trend (\( R=21.137, p < 0.001 \)). Von Neumman’s test was also significant (\( N=1.119, p <0.0001 \)).
As a final exercise to explore the issue further, the radical action and radical belief group data sets were combined. Not only did all tests identify a trend but \( t_{64} \) and \( t_{1} \) remained as the points of change. The overall findings are that within the range of \( t_{54} \) and \( t_{64} \) a change in patterns of posting activity is identifiable. Additionally, \( t_{1} \) may indicate a part of departure or it may signify a lack of activity in the days immediately preceding an attack.

### 5.5 Discussion

The findings of this study indicate that a range of social media level metrics may be better predictors than content alone for identifying the possible move from radicalization of belief to radicalization of action. While none of these parameters should be used on their own as the sole basis upon which to assess risk of the move from radicalization of belief to action, the identification of additional factors presents a first step towards developing a more evidence based approach. While the majority of radical group member showed indications of radicalization (77%), only 23% indicated intentions prior to attack. As such, there is great importance of the identification of a range of alternative metric upon which to differentiate between the two groups. Overall, all the parameters explored show some promise of being good predictors of online radicalization to varying degrees. Frequency and timing of activity shows clear differences between the two groups. Additionally, type of activity and radicalness of each activity are also quite different between the two groups. As noted above, radicals are more likely to pepper their radical posts with personal and otherwise innocuous content and are were also more likely to post original content, especially in the form of text. The response and reaction to posts was also greater among the radical action group.
It can be summarized from these results that radical action group members post more frequently than radicals, post overall more threatening content and are less likely to post innocuous content. They are more likely to have a higher proportion of shared content relative to radicals who may have a higher proportion of original content. Radical action group members are perhaps more likely to have wider networks and their posts attract larger reactions. These more qualitative understands are important because as the final tests found, while the radical group activity has no identifiable trend on its own, it easily fits in to the trend of the radical action group data set. In this regards, an identifiable change in post frequency behavior is observable in the 54-64 day period before an attack, followed by a trailing off period in activity that is identifiable by a post or more on the day before or of the attack.

Perhaps these findings lend support to the idea that the most ideal assessment and detection tools are those that require human input. This also does not mean that looking at keywords, phrases and other content should be abandoned. Rather, automated systems should better incorporate qualitative approaches together with statistical methods and human operation and oversight.

5.6 Summary

This analysis has successfully demonstrated both the usefulness and applicability of the social learning two pyramid integrated model. This study goes beyond the state of the art by identifying a range of new variables which can differentiate between the radical belief and radical action groups. Future analyses should combine these metrics into the development of a single predictive model. Future analyses should also look at what type of posts attract what type of responses and how that effects later posting behaviors. In other words, do differential associations effect future behavior which in turn can predict behavioral outcomes? Many other questions like this can and should be explored further. For example, what type of activity at what level of radicalness is more associated with differing radicalization outcomes? In order to answer such questions, more data is needed, especially from non-violent radical comparisons.

Additionally, future analyses should use longer frames of time, especially given that radicalization processes can take place over both short and long spans of time. Radicalization processes take place over different spans of time, from many years to sometimes only a few months or even weeks. However, the 100-day framework is a useful approach given that security services are likely to already have a long list of radicals under some form of surveillance or
monitoring. In this case, the 100-day window approach may be useful for examining known radicals already believed to be at a high risk of offending.
6 Task 3.5: Innovative study - terrorist-related contents in cyberspace

The goal of Task 3.5 is the identification and collection of terrorist-related online content and the assessment of the mechanisms of online propaganda. The online activities of the so-called Islamic State (IS) were chosen as the primary study object. The reasons for this choice are the relevance of the topic in general, the availability of real-world data samples within social media platforms, and the availability of related research work. The propaganda content produced by the IS gained a global attention, mainly because of the usage of social media and the professional quality of the content [2]. Several media-centres were or are still operated by IS [4]. Besides that, a multitude of research dealing with the analysis and assessment of IS propaganda has been published since the first propaganda videos emerged around 2014, which allowed us to build our work on a solid scientific base.

The collection of the relevant content from social media platforms is performed with the technical platform developed within WP3 (see section 3.1). The approach to collect the data uses textual information available from the data sources. For Task 3.5, data from the social media platforms Twitter and YouTube was collected and stored within the system’s database. All personal data is stored in an anonymized form, to ensure that no backtracking of users can be conducted. This is implemented by storing SHA-256 hashes of the content creators (i.e. the creators of tweets, video comments or YouTube user names).

The data collected from social media platforms is characterized by the fact that it is usually created by a multitude of different authors with an unknown background, attitude and role. Besides the data from social media platforms, textual data from several issues of the IS propaganda magazine Rumiyah [32] was collected for an analysis. The data from this magazine is likely to be created by only a few people with a known background, attitude and role and allows the comparison and validation of the data collected from social media.

Details on the data acquisition from the described data sources are given within the following paragraphs.

The online magazine Rumiyah is published on a regular base in multiple language and mainly comprises content which can be seen as propaganda. The magazine is published in the portable document format (PDF) and can be accessed at https://en.wikipedia.org/wiki/Rumiyah_%28magazine%29

---


32 https://en.wikipedia.org/wiki/Rumiyah_%28magazine%29
found online using a simple google search. The English versions of the issues 5, 7, 8, 10, 11 and 12 could be found and downloaded. From the PDF files, plain textual data was extracted for further processing and analysis. The goal from analysing the data from the magazine was to see if specific information about the creators of the propaganda content, i.e. the authors of the texts, could be extracted. A detailed description of the analysis can be found within section 6.1.4.

Although being primary a video platform, YouTube exposes also textual data in the form of video descriptions and user comments. It has been shown in Literature that this textual information could be used for a semantic analysis of the video content, without the need to perform a difficult and computationally expensive analysis of the video content through methods of image- and audio processing. For example, the work of Schultes et. al. [3] shows how user comments in YouTube videos can be used to classify the video content. In WP3, we followed a similar approach and used textual data from YouTube to identify and analyse propaganda content.

The datasets we collected from YouTube consist of information on the video and the associated comments. For the videos, the data fields collected are:

- Video title
- Description text
- Like count
- Dislike count
- View count

Furthermore, the user comments for each video are collected and stored in the database. For each comment, the following data is stored:

- Comment author name (anonymized)
- Comment text
- Information if it is a reply to another comment

The collected textual content i.e. the video descriptions for each video and the accompanying user comments, the Tweets and the text fragments from the propaganda magazine can be accessed through the implemented technical platform (see section 3.1) and were used for further in-depth examinations using statistical analysis methods and natural language processing, as well as an analysis of the personality structures of the actors involved. In the following sections, the methods to identify terrorist related content, and the methods to assess the mechanics of online propaganda are described in detail.

## 6.1 Identification and collection of terrorist-related online contents

Although social media platforms like Twitter, Facebook and YouTube undertake great efforts to detect and eliminate such unwanted content from their
networks, there is still a lot of this content traceable in the web. One aspect of the work within WP3 is the automated detection of the relevant data sources by a computer program. This allows the collection of much more data than it would be doable for human investigators with reasonable effort. The downside of collecting large amounts of data automatically is the collection of data which is misclassified as relevant content. It is very likely in completely automated data collection that the collected data contains this kind of content to a certain percentage. To overcome this problem, the data collection and classification system needs to be precise enough to keep the number of outliers small, compared to the relevant data.

### 6.1.1 Radicalization signals in social media

A common approach to identify IS-related social media content by text fragments is the analysis of the text for the occurrence of certain keywords which are statistically overrepresented in text-corpora with known IS-content. A multitude of literature deals with machine-based or manual corpus analysis to identify relevant keywords, the work described in [1] and [4] served hereby as the foundation for the keyword list used within this task of WP3.

Table 36: List of keywords used to identify radical content

<table>
<thead>
<tr>
<th>Arabic</th>
<th>English translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>راية</td>
<td>Banner of Punishment</td>
</tr>
<tr>
<td>اخلافة</td>
<td>Succession</td>
</tr>
<tr>
<td>الاسلامية الخلافة</td>
<td>Islamic Caliphate</td>
</tr>
<tr>
<td>أكبر الله</td>
<td>Allah is the greatest</td>
</tr>
<tr>
<td>ردة</td>
<td>Repetition</td>
</tr>
<tr>
<td>عقاب</td>
<td>Punishment</td>
</tr>
<tr>
<td>عقاب شرک</td>
<td>Trap of Punishment</td>
</tr>
<tr>
<td>عقاب کافر ر</td>
<td>Infidel punishment</td>
</tr>
<tr>
<td>الجهاد</td>
<td>Jihad</td>
</tr>
<tr>
<td>مجازوین</td>
<td>Mujahedeen</td>
</tr>
</tbody>
</table>

### 6.1.2 Analysis of Twitter data

In one stream of investigations we followed the tweets of a group of 'ISIS hunters' that mention twitter accounts that they associate with ISIS activities and propaganda. From these tweets we extract user accounts mentioned and subsequently check their status (active/suspended). The timeframe observed here is February 2017 to August 2017.
This stream of tweets, subsequently called *daesh* stream, is selected by querying Twitter regularly for tweets of the authors:

- @sunny_wantsome
- @CtrlSec
- @CtrlSec0
- @CtrlSec1
- @CtrlSec2
- @CtrlSec9
- @CtrlSecAR and
- @CtrlSec_FR

This stream contains about 1500 tweets per day, like:

Over the time February to August 2017 we find 33272 unique Twitter users mentioned in here – out of which at the end of the period 25185 (more than 75%) are now suspended by Twitter.
This gives the impression that this group of Twitter users are actually finding a large amount of other Twitter users that are deemed for suspension by Twitter.

The next figure shows a timeline of a small selection of accounts mentioned in the *daesh stream*. Each horizontal line corresponds thereby to a specific account mentioned. Blue markers mark the time when the account is mentioned in the *daesh stream*. Green/red markers mark the time and result when the account was checked for its active/suspended status.

![Timeline for specific accounts mentioned in the daesh stream](image)

*Figure 22 Timeline for specific accounts mentioned in the daesh stream - for details see text.*

The regular structure of the vertical lines of marks corresponds to the running of the automated checking process across the mentioned accounts.

One can see different behaviours in the timelines, i.e., for specific accounts mentioned in the daesh stream - better visible in the following enlarged view.
For example, for the uppermost and lowest 2 accounts we find several mentions in the daesh stream while they are active (green markers), but no more mentions when the accounts are suspended (red markers) – suggesting that the goal of mentioning was reached and noted.
In other cases, like the third from the top, continuous mention does not seem to lead to suspension by Twitter.
For the 3rd and 4th account from below, only one/few mentions occur, but no further mentions although the account stays active.
Few cases have also been observed where suspended accounts have been reactivated later by Twitter (not shown here).
In summary, we observe that a group of Twitter users acting as ‘ISIS hunters’ are trying to bring attention to Twitter of suspicious, ISIS-related Twitter accounts – with a great deal of success as we find that more than 75% of the mentioned accounts are indeed suspended at the end of our observation period.
The mentioned accounts therefore represent an interesting pool of Twitter accounts for potential future investigations.

6.1.3 ANALYSIS OF YOUTUBE DATA

The collection of IS-related content on YouTube was realized using the dedicated web-crawlers which are part of the technical system described in section 3.1.3. The list of keywords was derived from literature ([1], [4]). Several tests showed that the language used for the search terms greatly affects the results delivered by YouTube. The implemented data collectors are therefore able to process keywords in different languages. To be able to deal with Arabic language, which mainly means the processing of the Arabic characters, the processing of Unicode strings needed to be taken into account.
The web-crawlers are able to search for YouTube videos by one or more given keywords. The crawler simply automates the process which is usually
performed by a human (see Figure 24) by sending a dedicated HTTP request to YouTube. The HTTP code returned from YouTube, which represents the search result presented to a user, is then analysed and the URLs to the resulting videos are extracted (c.f. Figure 14). All the data is then stored in a database in an anonymized form.

Figure 24: Example of a YouTube search. The search term is entered into the text field on the top of the web page; a list of videos with according links is presented.

We were able to collect data from about 5,000 videos, containing around 60,000 comments from ca. 50,000 commenters. The resulting text fragments are then further examined for the occurrence of keywords indicating radicalization, and the amount of words within the text. The number of words in a text is a significant parameter for the extraction of personality traits, which is described in detail within section 6.2. Within the set of video descriptions and comment texts, the proportion of elements containing keywords indicating radicalization is quite low. Roughly 10% of the video descriptions in the dataset contain the keywords, so do roughly 1.7% of the comments (c.f. Figure 25).
Figure 25: Proportion of video descriptions with matching radicalization indicating keywords (left), and video comments (right)

A large number of video description are containing only a few words, only around 200 video descriptions contain more than 500 words, the same holds true for the video comments, as the histograms in Figure 26 show. The graphs in Figure 26 show that most of the textual data accompanying YouTube videos is quite small w.r.t. the number of words.
6.1.4 Analysis of texts from the Rumiyah magazine

From the PDF files containing the issues of the magazine, the plain texts were extracted using the Python-based pdfminer\textsuperscript{33} library. The extracted texts needed some manual processing and correction steps, as the software was not able to extract the text completely in the correct order, presumably due to the fact that the original text from the magazine contained a lot of images with captions. The texts from all issues available were assessed for the suitability for a personality profile analysis. To get a realistic impression of the personality of the text author, it is necessary that the text is mainly written by this author. For this reason, we focussed on text containing interviews and eyewitness accounts, as chances are higher that these texts are more likely to be written by a single person and are subject to less editorial revisions than other types of texts. We were able to identify 3 interviews and one report from a Mujahedeen fighter from the issues available:

- Interview with the Amir of hisbah in Sinai (issue 5)
- Interview with the Amir of the soldiers of the Khilafah in East Asia (issue 10)
- Interview with the military commander in the city of Raqqa
- A Mujahid’s memories from the battle of Mosul (issue 12)

For the interviews, the text of the interviewee was separated from the text of the questioner. Only within the interview in issue 10, the name of the interviewee together with a photo is shown (see Figure 27). According to [5], the person interviewed could be identified Isnilon Hapilon, leader of the Abu Sayyaf Group.

\textsuperscript{33} https://pypi.python.org/pypi/pdfminer/
All the texts selected from the available issues were analysed using the Watson Personality Insights service (see section 3.7). As a result, a personality profile of each interviewee based on the text could be calculated. A discussion of the results and the meaning for the assessment of the mechanics of online propaganda is given in section 6.2.2.

6.2 Assessing mechanisms of online propaganda

Our approach for the assessment of the mechanics of online propaganda is the derivation of the personality profiles of the actors involved, i.e. the propaganda content creators, and the propaganda content “consumers”. We made the results of the personality insights analysis available online in form of a HTML / Javascript based visualization system. In addition to this document, this will support the distribution of the results to the PROTON consortium’s psychology experts, as it allows for an interactive exploration of the results. The system will be based on simple HTML pages and can be distributed easily to the partners. To access the data, only a web browser is needed. For a maximum data protection, the visualization system will send to the partners on request, avoiding any online hosting of the data.
6.2.1 Personality traits clustering for social media data

We analysed the textual content of ca. 5000 YouTube videos which were crawled by our technical system (see section 6.1.3). The text fragments were filtered according to their length and the occurrence of radicalisation keywords. All comments and video description suitable for a personality traits extraction were loaded into the WPI system, resulting in a set of 52 personality traits for each text. The 52-dimensional data samples are then analysed for clusters forming in the data. As the numbers of potential clusters are not known in advance, we choose the Mean-Shift clustering algorithm (Cheng, 1995) for this task. The cluster centers derived from this process are interpreted as the main personality traits of the actors involved and can then be mapped to specific parameters of the agent models. In Figure 28, a schematic overview of the computational pipeline is shown.

We examined the possibilities of a dimensional reduction of the 52-dimensional personality traits data samples through PCA in order to improve the clustering results, a common approach proposed in machine learning literature (Domingos, 2012). As it is shown Table 1, the first 12 principal components...
describe most of the variance of the personality traits in the datasets for video comments and video descriptions.

Table 37: Dimensionality reduction through PCA. Roughly 90% of the variance in both comments and descriptions can be explained by the first 12 principal components.

<table>
<thead>
<tr>
<th>video descriptions</th>
<th>variance</th>
<th>summed variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>component 0</td>
<td>0.402</td>
<td>0.402</td>
</tr>
<tr>
<td>component 1</td>
<td>0.2122</td>
<td>0.6142</td>
</tr>
<tr>
<td>component 2</td>
<td>0.1032</td>
<td>0.7174</td>
</tr>
<tr>
<td>component 3</td>
<td>0.0532</td>
<td>0.7706</td>
</tr>
<tr>
<td>component 4</td>
<td>0.0438</td>
<td>0.8144</td>
</tr>
<tr>
<td>component 5</td>
<td>0.0323</td>
<td>0.8467</td>
</tr>
<tr>
<td>component 6</td>
<td>0.0209</td>
<td>0.8676</td>
</tr>
<tr>
<td>component 7</td>
<td>0.0194</td>
<td>0.887</td>
</tr>
<tr>
<td>component 8</td>
<td>0.0169</td>
<td>0.904</td>
</tr>
<tr>
<td>component 9</td>
<td>0.0148</td>
<td>0.9187</td>
</tr>
<tr>
<td>component 10</td>
<td>0.0126</td>
<td>0.9313</td>
</tr>
<tr>
<td>component 11</td>
<td>0.0103</td>
<td>0.9416</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>video comments</th>
<th>variance</th>
<th>summed variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>component 0</td>
<td>0.2654</td>
<td>0.2654</td>
</tr>
<tr>
<td>component 1</td>
<td>0.1575</td>
<td>0.4229</td>
</tr>
<tr>
<td>component 2</td>
<td>0.1251</td>
<td>0.548</td>
</tr>
<tr>
<td>component 3</td>
<td>0.0828</td>
<td>0.6308</td>
</tr>
<tr>
<td>component 4</td>
<td>0.0579</td>
<td>0.6887</td>
</tr>
<tr>
<td>component 5</td>
<td>0.0498</td>
<td>0.7385</td>
</tr>
<tr>
<td>component 6</td>
<td>0.0354</td>
<td>0.7739</td>
</tr>
<tr>
<td>component 7</td>
<td>0.0322</td>
<td>0.8062</td>
</tr>
<tr>
<td>component 8</td>
<td>0.0267</td>
<td>0.8328</td>
</tr>
<tr>
<td>component 9</td>
<td>0.0241</td>
<td>0.857</td>
</tr>
<tr>
<td>component 10</td>
<td>0.0235</td>
<td>0.8805</td>
</tr>
<tr>
<td>component 11</td>
<td>0.0178</td>
<td>0.8982</td>
</tr>
</tbody>
</table>

For the personality clusters derived from the comment data, the number of clusters varies between the PCA transformed data (6 clusters), and the raw data (8 clusters). This is not the case for the video description data, where there is only a slight variation within the number of data samples within each cluster (c.f. Figure 29)
To measure the quality of the calculated clusters, we calculated the silhouette coefficients for each data sample and visualized them in silhouette coefficient plots. Within Figure 30 and Figure 31 the plots for both PCA transformed and raw data sets are shown. For the video comments, the clustering for the largest cluster can be considered as sub-optimal, as there are some silhouettes ranging in the negative area. This is an indication that the parameters for the mean shift clustering algorithm might need an adjustment. In contrast, the cluster quality for the video description data can be considered as quite good.
The following series of images show the personality clusters derived from the processed data sets for both video descriptions and video comments. The personality traits derived from the video descriptions can be interpreted as the main personality structures of the video content providers, as the text probably originates from the person uploading the video, or at least from the person providing the video.
Figure 32: Personality clusters derived from the WPI analysis of the video description texts. Cluster 0 is shown on the left, cluster 1 is shown on the right.
Figure 33: Personality clusters derived from the WPI analysis of the video comments texts. From top left to bottom right: Cluster 0 – Cluster 7.

6.2.2 PERSONALITY PROFILES FROM PROPAGANDA TEXT

Compared to the aggregated extraction process of the main personality patterns which was conducted for mass data from social media (see section 6.2.1), no aggregation and clustering process needed to be conducted, as the texts originate with a high probability from one author. Within Figure 34, the extracted personality profiles from the four texts are shown.
Figure 34: Sunburst visualizations for the results of the WPI analysis from the different texts from the Rumyiah magazine. Top left: Interview from issue 5. Top right: Interview from issue 10. Bottom left: Interview from issue 12. Bottom right: Narrative "A Mujahid Memories"

6.3 Summary

In this task, several approaches for the identification and assessment of online terrorist content were investigated and realized. Our main contribution for assessing the mechanics of online propaganda is the derivation of personality
profiles of the propaganda content providers and the propaganda content consumers. We generated a data collection which can be used for further investigations. The personality profiles generated will be discussed with the consortium’s experts in order to provide the foundation for the work conducted in WP4.

7 Task 3.6: Policy makers’ contribution

The goal of T3.6 is the connection of the activities conducted within WP3 with the relevant policy makers in the field of cyberspace and OCTN. This connection is achieved through the inclusion and consideration of the policy maker’s comments and views into the work of WP3. These contributions were collected through a draft report which was delivered in month 11 and reviewed by the policy makers. The results of these reviews were discussed during the first consortium meeting in October 2017. Besides this, a meeting between the partners EUROPOL and Fraunhofer took place in August 2017.

Contributions w.r.t. WP3 were given by the following policy makers:
- European Police Office EUROPOL (EUROPOL)
- United Nations Office on Drugs and Crime (UNODC)
- Brottsförebyggande rådet (Brå)
- European Crime Prevention Network (EUCPN)
- Ministry of Security and Justice / WODC (WODC)

The individual contributions of each policy maker are summarized within the following subsections.

7.1 EUROPOL

1. NSM activities and behaviours as predictors of terrorism support and involvement. WP3, T3.4 (HUJI)
   a. Little importance is given to the online activity. This may be due to the specific context of Israel, which is not easily comparable with the increasingly relevant role of online radicalisation / online behaviour of
radicalised individuals in the EU. Decision to get involved and decision to carry out an attack should be distinguished.

b. Interesting to be able to determine shifts from radicalisation of belief to action.

2. Organized Crime and Terrorism in Cyberspace. WP3, T3.2, T3.3, T3.5 (Fraunhofer, IBM)
   a. The collection and analysis of various social media and internet data sources should be wider: Not only Twitter and Youtube but also Telegram, Baaz, VK. Justpaste.it and 100s of other platforms should also be considered.
   b. Has the difference between interactive platforms and content hosting platforms been taken into consideration?
   c. Liaising with research centres or law enforcement agencies should be taken into serious consideration to keep the collection plan relevant. Contacts of potentially relevant stakeholders in Germany have been provided to the researcher in early September. We recommend trying to liaise with them as soon as possible.
   d. Which are the areas of concern? Europe? Or other regions?
   e. Which online behaviour is the subject of study? What are the criteria to assess the selected material?
   f. If material is in Arabic language, the use of Google translator may not be the best solution for translation purposes. We recommend liaising with native speakers in the academic context that can provide more reliable translations and interpretations – especially for a language like Arabic that heavily relies on specific dialectical variations.
   g. Which is the expected margin of error that the researcher will assess as valid?
   h. How to assess the anomalies?
   i. Is there a training of the system planned? For how long? And how?
   j. In regards with Agent Personas methodology, how will it help to classify the material and determine the impact of one piece rather than another?
   k. How the comparison between ISIS recruiter versus a recruiter of the "Libyan Islamic Fighting group" was chosen?
   l. Will the tool itself aim to detect suspicious online behaviours?

7.2 UNODC

WP3, T3.4: to which extent the limitations of the study (only lone wolves) affect the applicability of the findings to ALL (potential) terrorist? I had a conversation with Dr. Simon Perry and Prof. Badi Hasisi about this in Jerusalem and do not think there is a need to further elaborate on this as they will remember what my argument is.
I also echo the comments made by my EUROPOL colleagues about the need to perhaps also look at other social media (Instagram etc.) though I am aware there may be time and budget constraints.

7.3 EUCPN

**WP3, T3.4: NSM activities and behaviours as predictors of terrorism support and involvement:**
How are ‘lone wolf terrorists’ selected and defined? This has a number of implications for the research, for example: if you define lone wolf sensu stricto, then he or she has no social ties, therefore he or she cannot score on the indicator of having social media and offline friends, disrupting the social learning theory. Moreover, in defining ‘lone’ are only social and family ties of importance or can motivation be a group defining feature in this research? Fluid network structures can be a good example here. The hacktivist group Anonymous has a similar working pattern. Same issue with defining Radicalization: when is something called radicalization and when are real intentions indicated?

**WP3, T3.2, T3.3, T 3.5 Organized Crime and Terrorism in Cyberspace:**
Personas modelling: what are the structural differences between an ISIS recruiter and a recruiter of Libyan Islamic Fighting Group? They have more in common than differences,
Only Twitter and Youtube as social media and internet data... Other sources?

7.4 WODC

**WP3, T3.4 – NSM activities**

1 Theoretical Framework
Referring to “By comparing violent and non-violent radicals, this study aims to identify social media level metrics which can be used as predictors to identify shifts from radicalisation of belief to action”: Relevant question. Issue seems to be though how to assess whether offline factors are indeed the deciding factors of influence that have someone cross the line to action-prone/willing to take action (is there such a thing as pure online radicalization towards violence?)
7.5 Actions resulting from the Policy Maker’s contributions

The valuable feedback we received from the policy makers helped us to adjust the work we conducted so far and pointed us into new directions to look for. We were able to realize at least some of the suggestions within the available time and with the available resources.

We addressed the issue of automatic translation, mentioned by the partner Europol, by completely avoiding any translation. The data analysis works in the native languages, especially in the context of T3.5, the processing of Arabic language was implemented through implementing the ability to handle Unicode strings within the whole calculation chain, including the WPI software provided by the partner IBM.

We were also investigating the possibilities to acquire datasets from other data sources. The extraction of data from the propaganda magazine “Rumyiah” was conducted in order to get ground truth data for the validation of the outcome of our system. Besides that, we investigated the technical requirements to automatically extract data from the sources:

- Google drive
- Archive.org
- WordPress
- Tumblr
- Telegram
- Steam

A detailed description of the outcome of this research is listed in Annex A of this document.
8 Agents and Interactions

8.1 Agent Persona Library

We propose a methodology that we have adapted from an established cyber security risk modelling scheme based on agent personas as the mechanism for the transfer of our output. We intend to deliver a number of agent models packaged into a tool called the agent persona library (APL). The concept of an agent persona library is an adaptation of the Threat Agent Library (TAL) tool in the TARA approach to cyber security risk modelling.

An agent persona can be described as a set of attributes that describe a specific type of actor for modelling purposes. For example, in the case of modelling terrorist recruitment processes, one may create an agent persona for the terrorist recruiter, one for an ISIS sympathiser and a further agent persona for the target victim. The proposed concept allows simple refinement of personas for modelling more specific actor roles such as ISIS recruiter verses a recruiter of the “Libyan Islamic Fighting group”.

The methodology is based on a combination of:
- The Persona concept originating in the usability design space (see section 3)
- The Threat Agent Library used in the Threat Assessment & Remediation Analysis (TARA) cyber risk modelling methodology (see section 8.4)
- Models used for behavioural/personality analysis (see section 6)
Agent Personas are characterised by a set of attributes. Some attributes are descriptive only and are used to provide context. Some attributes are metrics and can be used for modelling. Attributes are collected in different ways.

- Some attributes are generated by subject matter experts and domain specialists
- Some attributes are generated from the linguistic analysis of written text.
- Some attributes are generated from the analysis of observed cyber behaviour

**8.1.1 AGENT PERSONAS AND DATA PRIVACY**

Agent Personas are not individual people, they are artificial constructs in the form of stereotypes or genotypes that are composite representations of a typical actor in a modelling scenario. Personas consist of aggregated attributes and not specific identities and as such do not represent a problem with data privacy laws. Care still needs to be taken that attributes do not constitute pseudo identifiers, primarily by ensuring that there is sufficient aggregation.

Note: Public data can be collected in an anonymous form.  
Note: Interview transcripts can be used such that only anonymous personality traits are used.  
Notes: Data from individuals are aggregated into agent personas.
8.1.2 Agent Persona Template

An Agent Persona template comprises of a number of sections:

1. An image and description for the persona. This section is designed to give depth and context to a Persona profile in order to better guide the selection of attribute values.

2. An attribute section containing a variation of metrics derived from an extended TARA scheme. This section is configured for each agent persona. One would need to create a set of default personas for each of the agent profiles that PROTON wishes to model. The default values could be modified for a specific use case or scenario to be analysed. An extension to this approach could be a common dataset containing a larger set of personas based on real world cases.

3. A section on observed cyber behaviour

4. A section on Personality Traits

**Agent Title**  Terrorist Recruiter

**Image**

**Description** Male in their early 20s from a middle to upper class background. A 2nd generation immigrant with a bachelors education. ...

**Attribute Metrics**  
- Impact Radius  
- Code of Conduct  
- Organisational Size  
- Visibility  
- Payment  
- Age

**Cyber Behaviour Metrics**  
- Social Media Usage  
- Darknet Usage  
- Clearweb Usage  
- Cryptocurrency Use  
- Messenger Use

**Personality Traits Metrics**  
- Agreeableness  
- Conscientiousness  
- Extraversion  
- Emotional range  
- Openness
8.1.3 COLLECTING DATA FOR PERSONAS

The studies that WP3 is undertaking will be used to collect data for personas according to the following table:

<table>
<thead>
<tr>
<th>Attribute Metrics</th>
<th>Cyber Behaviour Metrics</th>
<th>Personality Traits Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Subject Matter</td>
<td>• Social Media Analytics</td>
<td>• Linguistic Analysis</td>
</tr>
<tr>
<td>• Expert Input from consortium members</td>
<td>• Web Crawling</td>
<td>• Watson Cognitive Personality Traits</td>
</tr>
<tr>
<td>• Previous Studies</td>
<td>• Previous Studies</td>
<td></td>
</tr>
</tbody>
</table>

8.2 Benefits to this approach

We believe that the approach of using agent personas to package modelling attributes offers a number of advantages:

- Personas allow a rich way to embellish a set of attributes required for agent modelling.
- It will allow us to present our output in a form most likely to be consumable by other project partners who otherwise might not understand cyber related attributes.
- They offer a rich narrative for SMEs to adapt/tailor input.
- They offer a means to combine different types of data and data collection methodologies.
- The multi-dimensional attributes of Personas could help improve the effectiveness of crime prevention agencies at identifying at risk individuals based on.
- They can be expanded on during this and be used in later projects as the understanding of genotypes of actors in OCTNs grows and the OCTN actors evolve and change.
8.3 Adapting the TARA methodology

The TARA methodology is shown in Figure 36: TARA Scheme overview

The methodology includes three primary activities:
- Cyber Threat Susceptibility Analysis (CTSA),
- Cyber Risk Remediation Analysis (CRRA), and
- Data and Tools development.

These activities support three workflows:
- TARA assessments,
- catalogue development, and
- toolset development.

A TARA assessment is a sponsor-directed workflow to evaluate selected cyber assets using information about known adversarial Tactics, Techniques, and Procedures (TTPs) and Countermeasures (CMs) stored in catalogues. Each TARA assessment involves a three (3) step process:
- establish assessment scope in terms of the cyber assets and range of TTPs to evaluate,
- apply CTSA to assess a cyber asset's susceptibility to attack over the range of TTPs, and
- conduct CRRA to determine the set of CMs that will effectively reduce or eliminate the cyber asset's susceptibility to attack.
A TARA assessment delivers recommendations that help program managers make informed decisions on how to make systems more resilient and less vulnerable once deployed.

Taking into account the desired output of the PROTON project, a derived scheme would be an assessment that delivers recommendations that help program managers make informed decisions on how to make people and society more resilient against the recruitment efforts of terrorism and organised crime, and less vulnerable on-line.

<table>
<thead>
<tr>
<th>TARA</th>
<th>PROTON</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crown Jewel Analysis</strong></td>
<td>Target Victim Group</td>
</tr>
<tr>
<td><strong>System Architecture and Design Specifications</strong></td>
<td>Context and Behavioural Models</td>
</tr>
<tr>
<td><strong>Threat Agent Library</strong></td>
<td>Agent Persona Library</td>
</tr>
<tr>
<td><strong>Common Exposure Library</strong></td>
<td>Common Adversary Behavioural Library</td>
</tr>
<tr>
<td><strong>Design and operational mitigations, tools and techniques</strong></td>
<td>Behavioural mitigation library, tools and techniques</td>
</tr>
</tbody>
</table>
8.4 TARA Model

Threat Assessment & Remediation Analysis (TARA) is an engineering methodology to identify, prioritize, and respond to cyber threats through the application of countermeasures that reduce susceptibility to cyber-attack. TARA is a system level engineering practice within the MITRE Mission Assurance Engineering (MAE) portfolio [Fehler! Verweisquelle konnte nicht gefunden werden.].

8.4.1 TARA Profiles

The TARA methodology defines a set of attacker profiles with attributes. These attacker profiles could be used as the basis for a set of personas. The profiles that TARA defines are:

- Reckless Employee
- Employee Untrained
- Info Partner
- Anarchist
- Civil Activist
- Competitor
- Corrupt Government Official
- Data Miner
- Employee Disgruntled
- Government Cyber warrior
- Government Spy
- Internal Spy
- Irrational Individual
- Legal Adversary
- Mobster
- Radical Activist
- Sesationalist
- Terrorist
- Thief
- Vandal
- Vendor
8.4.2 TARA Attributes

TARA has the concept of attributes. We would propose to extend the TARA attributes and turn some attribute sections into simple metric scales - from 1-5. This will allow a basic framework for generating attacker metrics.

<table>
<thead>
<tr>
<th>TARA Attributes</th>
<th>Suggested Extension to TARA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal</td>
<td>1. Internal</td>
</tr>
<tr>
<td></td>
<td>2. External Vicinity</td>
</tr>
<tr>
<td></td>
<td>3. External District</td>
</tr>
<tr>
<td></td>
<td>4. External City</td>
</tr>
<tr>
<td></td>
<td>5. External Country</td>
</tr>
<tr>
<td>Code of Conduct</td>
<td>1. Code of Conduct</td>
</tr>
<tr>
<td>Legal</td>
<td>2. Legal</td>
</tr>
<tr>
<td>Extra-legal, minor</td>
<td>3. Extra-legal minor</td>
</tr>
<tr>
<td>Extra-legal major</td>
<td>4. Extra Legal medium</td>
</tr>
<tr>
<td></td>
<td>5. Extra-legal major</td>
</tr>
<tr>
<td>Individual</td>
<td>1. Individual</td>
</tr>
<tr>
<td>Club</td>
<td>2. Club</td>
</tr>
<tr>
<td>Contest</td>
<td>3. Contest</td>
</tr>
<tr>
<td>Team</td>
<td>3. Team</td>
</tr>
<tr>
<td>Organisation</td>
<td>4. Organization</td>
</tr>
</tbody>
</table>

---

**TARA Attributes**

**Suggested Extension to TARA**

**Internal**

- 1. Internal
- 2. External Vicinity
- 3. External District
- 4. External City
- 5. External Country

**Code of Conduct**

1. Code of Conduct

**Legal**

1. Legal

**Extra-legal, minor**

1. Extra-legal minor

**Extra-legal major**

1. Extra-legal major

---

**Individual**

1. Individual

**Club**

1. Club

**Contest**

1. Contest

**Team**

1. Team

**Organisation**

1. Organization
Table 38 TARA Attributes

8.5 Adapting the TARA methodology

The TARA methodology is shown in Figure 36: TARA Scheme overview

Figure 37: TARA Scheme overview

The methodology includes three primary activities:
- Cyber Threat Susceptibility Analysis (CTSA),
- Cyber Risk Remediation Analysis (CRRA), and
- Data and Tools development.
These activities support three workflows:
- TARA assessments,
- catalogue development, and
- toolset development.

A TARA assessment is a sponsor-directed workflow to evaluate selected cyber assets using information about known adversarial Tactics, Techniques, and Procedures (TTPs) and Countermeasures (CMs) stored in catalogues. Each TARA assessment involves a three (3) step process:
- establish assessment scope in terms of the cyber assets and range of TTPs to evaluate,
- apply CTSA to assess a cyber asset's susceptibility to attack over the range of TTPs, and
- conduct CRRA to determine the set of CMs that will effectively reduce or eliminate the cyber asset's susceptibility to attack.

A TARA assessment delivers recommendations that help program managers make informed decisions on how to make systems more resilient and less vulnerable once deployed.

Taking into account the desired output of the PROTON project, a derived scheme would be an assessment that delivers recommendations that help program managers make informed decisions on how to make people and society more resilient against the recruitment efforts of terrorism and organised crime, and less vulnerable on-line.

```
Target Victim Group Analysis

Scoping
- Context, Behavioral Models,...

Cyber Threat Susceptibility Assessment (CTSA)

Ranked List of Threats

Cyber Risk Remediation Analysis (CRRA)

TTP/CM Mapings

X

Data and Tools Development

List of Mitigations

TARA

Cyber operational mitigations, best practices, support techniques

TAL: Threat Agent Library
CEL: Common Exposure Library
MOL: Methods and Objectives Library
```
8.6 Agent Persona Attributes

8.6.1 TARA RELATED

Attributes derived from the TARA model.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Description</th>
<th>Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impact Radius</td>
<td>The interaction range of a persona. Do they interact only within their local neighbourhood or internationally.</td>
<td>1 Street&lt;br&gt;2 Vicinity&lt;br&gt;3 District&lt;br&gt;4 City&lt;br&gt;5 Country</td>
</tr>
<tr>
<td>Code of Conduct</td>
<td>To what extent are they ready to break the law</td>
<td>1. Legal&lt;br&gt;2. Extra-legal minor&lt;br&gt;3. Extra Legal medium&lt;br&gt;4. Extra-legal major</td>
</tr>
<tr>
<td>Organisational Size</td>
<td>The size of the organisation in which they operate</td>
<td>1. Individual&lt;br&gt;2. Friends&lt;br&gt;3. Family&lt;br&gt;3. Organization&lt;br&gt;5. Government</td>
</tr>
<tr>
<td>Visibility</td>
<td>Do they operate in openly in public or very secretive and clandestine</td>
<td>1 Overt&lt;br&gt;2. Covert&lt;br&gt;3 Clandestine</td>
</tr>
<tr>
<td>Payment</td>
<td>What money form do they prefer</td>
<td>0 None&lt;br&gt;1 Natural&lt;br&gt;2 Cash&lt;br&gt;3 Credit Card&lt;br&gt;4 Crypto Currency&lt;br&gt;5 Any</td>
</tr>
</tbody>
</table>

Table 39 Persona Attributes –TARA related
8.6.2 LINGUISTIC ANALYSIS

8.6.2.1 BIG 5 PERSONALITY TRAITS MODEL
Attributes provided by the IBM Watson personality Traits interface.
Input Required: 1200 words of input text

<table>
<thead>
<tr>
<th>Attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agreeableness</td>
</tr>
<tr>
<td>Conscientiousness</td>
</tr>
<tr>
<td>Extraversion</td>
</tr>
<tr>
<td>Emotional range</td>
</tr>
<tr>
<td>Openness</td>
</tr>
</tbody>
</table>

*Table 40 Persona Attributes – BIG 5 Personality Traits*

8.6.2.2 PERSONAL NEEDS MODEL
Attributes provided by the IBM Watson personality Traits interface.
Input Required: 1200 words of input text

<table>
<thead>
<tr>
<th>Attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excitement</td>
</tr>
<tr>
<td>Harmony</td>
</tr>
<tr>
<td>Curiosity</td>
</tr>
<tr>
<td>Ideal</td>
</tr>
<tr>
<td>Closeness</td>
</tr>
<tr>
<td>Self-expression</td>
</tr>
<tr>
<td>Liberty</td>
</tr>
<tr>
<td>Love</td>
</tr>
<tr>
<td>Stability</td>
</tr>
<tr>
<td>Challenge</td>
</tr>
<tr>
<td>Structure</td>
</tr>
</tbody>
</table>

*Table 41 Persona Attributes: Personal Needs Model*

8.6.2.3 AGENT PERSONAL VALUES MODEL
Attributes provided by the IBM Watson personality Traits interface.
Input Required: 1200 words of input text
### Table 42 Persona Attributes: Personal Values Model

<table>
<thead>
<tr>
<th>Attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-transcendence</td>
</tr>
<tr>
<td>Helping others</td>
</tr>
<tr>
<td>Conservation / Tradition</td>
</tr>
<tr>
<td>Hedonism / Taking pleasure in life</td>
</tr>
<tr>
<td>Self-enhancement / Achieving success</td>
</tr>
<tr>
<td>Open to change / Excitement</td>
</tr>
</tbody>
</table>
9 References

References for T3.1


References for T3.2, T3.3, T3.4, T3.5


Khalil, J. (2014). Radical beliefs and violent actions are not synonymous: How to place the key disjuncture between attitudes and behaviors at the heart of our research into political violence. *Studies in Conflict & Terrorism, 37*(2), 198-211.


Plank, B., & Hovy, D. (2015). Personality Traits on Twitter - or - How to Get 1,500 Personality Tests in a Week. 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA 2015).
Rowe, M., & Saif, H. (2016). Mining Pro-ISIS Radicalisation Signals from Social Media Users. ICWSM.
Walli, J. (2015). Analyzing ISIL’s propaganda. Linnaeus University, Faculty of Arts and Humanities, Department of Media and Journalism.
9.1 Security Advisory Board Review

<table>
<thead>
<tr>
<th>Security Advisory Board Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comments</td>
</tr>
<tr>
<td>Recommended Distribution</td>
</tr>
<tr>
<td>Date</td>
</tr>
</tbody>
</table>

9.2 Pointers to the DarkWeb

9.2.1 Hidden Service Lists and Search Engines

- [http://3g2u4lh4p6u4k4cm.onion/](http://3g2u4lh4p6u4k4cm.onion/) – DuckDuckGo Search Engine
- [http://e266al3z4vuorbyg.onion/bookmarks.php](http://e266al3z4vuorbyg.onion/bookmarks.php) – Dark Nexus
- [http://5plvrsqydw2sgce.onion/](http://5plvrsqydw2sgce.onion/) – Seeks Search
- [http://2vlpqcppilhmd5r2.onion/](http://2vlpqcppilhmd5r2.onion/) – Gateway to Freenet
- [http://kpynvyym6xqi7zw2.onion/links.html](http://kpynvyym6xqi7zw2.onion/links.html) – ParaZite
- [http://kpvz7ki2v5agwt35.onion](http://kpvz7ki2v5agwt35.onion) – The Hidden Wiki
- [http://idnxncnne4q76t6q.onion/](http://idnxncnne4q76t6q.onion/) – Tor Project: Anonymity Online
- [http://torlinkbgs6aabns.onion/](http://torlinkbgs6aabns.onion/) – TorLinks
- [http://xdaqknwjc7aayz8h.onion/](http://xdaqknwjc7aayz8h.onion/) – Anonet Webproxy
- [http://3fyb44wdhnd2ghh1.onion/](http://3fyb44wdhnd2ghh1.onion/) – All You’re Base
- [http://j6im4v42ur6dpic3.onion/](http://j6im4v42ur6dpic3.onion/) – TorProject Archive
- [http://p3igkncorhackjtib.onion/](http://p3igkncorhackjtib.onion/) – TorProject Media
- [http://kbhodhnfxl3clb4.onion](http://kbhodhnfxl3clb4.onion) – Tor Search
- [http://cipollatnumrrhad.onion/](http://cipollatnumrrhad.onion/) – Cipolla 2.0 (Italian)
- [http://torlerutegdqgqqa.onion](http://torlerutegdqgqqa.onion) – Torle Search
- [http://dppmfxaacgccuzpc.onion](http://dppmfxaacgccuzpc.onion) – TorDir – One of the oldest link lists on Tor
9.2.2 MarketPlace Financial

- [http://torbrokerge7zxgq.onion/](http://torbrokerge7zxgq.onion/) – TorBroker – Trade securities anonymously with bitcoin, currently supports nearly 1000 stocks and ETFs
- [http://fogcore5n3ov3tui.onion/](http://fogcore5n3ov3tui.onion/) – Bitcoin Fog – Bitcoin Laundry
- [http://easycoinsayj7p5l.onion/](http://easycoinsayj7p5l.onion/) – EasyCoin – Bitcoin Wallet with free Bitcoin Mixer
- [http://jzn5w5pac26sgef4.onion/](http://jzn5w5pac26sgef4.onion/) – WeBuyBitcoins – Sell your Bitcoins for Cash (USD), ACH, WU/MG, LR, PayPal and more
- [http://qc7iionwpv77qigb.onion/](http://qc7iionwpv77qigb.onion/) – Western Union Exploit
- [http://3dba5t4pygdahedms.onion/](http://3dba5t4pygdahedms.onion/) – ccPal Store
- [http://xgi4drqavpm7eecl.onion/](http://xgi4drqavpm7eecl.onion/) – Counterfeit USD
- [http://nr6juudpp4as4qjq.onion/pptobtc.html](http://nr6juudpp4as4qjq.onion/pptobtc.html) – PayPal to BitCoins
- [http://nr6juudpp4as4qjg.onion/doublecoins.html](http://nr6juudpp4as4qjg.onion/doublecoins.html) – Double Your BitCoins
- [http://lw4ipk5choakk5ze.onion/raw/4588/](http://lw4ipk5choakk5ze.onion/raw/4588/) – High Quality tutorials
- [http://ploashqjmntbulrb.onion/](http://ploashqjmntbulrb.onion/) – PayPalace – one of the oldest PP account seller in Onionland
- [http://usjudr3c6ez6tesi.onion](http://usjudr3c6ez6tesi.onion) – Counterfeit USD/EUR
- [http://apple44jaj53hty.onion](http://apple44jaj53hty.onion) – Stolen Mac Store iPhone 5s inside!
- [http://cthulhuap7ch47k.onion](http://cthulhuap7ch47k.onion) – C'thulhu Resume – Assasination Service, When you have to have it done clean!

9.2.3 MarketPlace Commercial Services

- [http://6w6vcynl6dumnn67c.onion/](http://6w6vcynl6dumnn67c.onion/) – Tor Market Board – Anonymous Marketplace Forums
- [http://5vmv7m7c6bqf1ftp.onion/](http://5vmv7m7c6bqf1ftp.onion/) – Discounted electronics goods
- [http://lw4ipk5choakk5ze.onion/raw/evbLevqkDSVKifzy8ZAo/](http://lw4ipk5choakk5ze.onion/raw/evbLevqkDSVKifzy8ZAo/) – Unfriendlysolution – Legit hitman service
- [http://nr6juudpp4as4qjq.onion/torgirls.html](http://nr6juudpp4as4qjq.onion/torgirls.html) – Tor Girls
• http://tuu66yxvrrn3of7l.onion/ – UK Guns and Ammo
• http://nr6juudpp4as4qjq.onion/torguns.htm – Used Tor Guns
• http://ucx7bkbi2dtia36r.onion/ – Amazon Business
• http://nr6juudpp4as4qjq.onion/tor.html – Tor Technology
• http://hbetsipq5yhhrsd.onion/ – Hidden BetCoin
• http://cstoreav7i44h2lr.onion/ – CStore Carded Store
• http://tfwdi3izigxllure.onion/ – Apples 4 Bitcoin
• http://e2qizoerj4d6ldif.onion/ – Carded Store
• http://ucx7bkbi2dtia36r.onion/ – Amazon Business
• http://nr6juudpp4as4qjq.onion/tor.html – Tor Technology
• http://b4vqxw2j36wf2bqa.onion/ – Advantage Products
• http://ypb4oezhfhk24hxmb.onion/ – Hitman Network
• http://mts7hqqqeqqjuc5e.onion/ – Marianic Technology Services
• http://mobil7rab6nuf7vx.onion/ – Mobile Store
• http://54flq67kqr5wvjgf.onion/ – MSR Shop
• http://yth5q7zdmqlycbcz.onion/ – Old Man Fixer’s Fixing Services
• http://matrixtxri745d bw.onion/neo/uploads/MATRIXtxri745dwONION_130827231336iPA_pc.png – PC Shop
• http://storeqso3o5mfxl.onion/ – Samsung StorE
• http://sheep5u64f457aw.onion/ – Sheep Marketplace
• http://nr6juudpp4as4qjq.onion/betcoin.htm – Tor BetCoin
• http://giziroidqwmegq4p5b.onion/ – Tor Web Developer
• http://vfqnd6mieccqyiit.onion/ – UK Passports
• http://en35tuzqmnn4lofbk.onion/ – US Fake ID Store
• http://xfnwyiq7olypdq5r.onion/ – USA Citizenship
• http://uybu3melulmoljnd.onion/ – iLike Help Guy
• http://dbmv53j45pcv534x.onion/ – Network Consulting and Software Development
• http://lw4ipk5choakk5ze.onion/raw/4585/ – Quick Solution (Hitman)
• http://nr6juudpp4as4qjq.onion/tynermsr.htm – Tyner MSR Store
• http://lw4ipk5choakk5ze.onion/raw/5432/ – Unfriendlysolution – Legit Hitman Service (New link with current PGP key)
• http://abbujjh5vqtq77wq.onion/ – Onion Identity Services – Selling Passports and ID-Cards for Bitcoins
• http://2kka4f23pcxggkpv.onion/ – EuroGuns – Your #1 european arms dealer
• http://xyt4inljwmm3gfb.onion/ – Buy cheap iPhones
• http://oh2s6ywmaxtqcj25.onion/store.html – Android store
• http://euhesnelvagifk4.onion/ – Stolen Mac Store iPhone 5s inside!
• http://iliike77nuoh3ummf.onion – iLike Help Guy I’d like to help you in ANY project
• http://iacgg9y2j2nfudy7.onion – Assasination Service, The best place to put your problems is in a grave.
• http://3pkm74bnuxdbaxl.onion – Cheap stolen android phones – #1 store
• http://uizxxzdowrh6rrys.onion/ – One of only two original apple discount stores on Tor. Verified & reliable, since 2011.
9.2.4 MARKETPLACE DRUGS

- http://atlmlxbk2mbupwgr.onion/ – Atlantis Marketplace Forums
- http://atlantisrky4es5q.onion/ – Atlantis Marketplace
- http://dkn255hz262ypmii.onion/ – Silk Road Forums
- http://4yjes6zfucnh7vcj.onion/ – Drug Market
- http://k4btcoezc5tlyyaf.onion/ – Kamagra for BitCoins
- http://smoker32pk4qt3mx.onion/ – Smokeables – Finest Organic Cannabis shipped from the USA
- http://fzqnrlcvhkgbdw5.onion/ – CannabisUK – UK Wholesale Cannabis Supplier
- http://s5q54hfww56ov2xc.onion/ – BitPharma – EU vendor for cocaine, speed, mdma, psychedelics and subscriptions
- http://ll6lardicrvrljvq.onion/ – Brainmagic – Best psychedelics on the darknet
- http://74ypjqjwf6oejmax.onion/ – Beneath VT – Exploring Virginia Tech’s Steam Tunnels and Beyond
- http://76qugh5bey5gum7l.onion/ – Deep Web Radio
- http://shopsat2dotfotbs.onion/ – TorShops – Get your own .onion store with full bitcoin integration

9.2.5 BLOGS

- http://74ypjqjwf6oejmax.onion/ – Beneath VT – Exploring Virginia Tech’s Steam Tunnels and Beyond
- http://76qugh5bey5gum7l.onion/ – Deep Web Radio
- http://edramalpl7oq5npk.onion/Main_Page – Encyclopedia Dramatica
9.2.6 FORUMS AND CHANS

- http://3fyb44wfhnd2ghhi.onion/ib/ – Onii-Chan
- http://bx7zrcsebkma7lfs.onion – Jisko
- http://npdaaf3s3xfrmlm.onion/ – Twitter clone
- http://xdaqknwjc7aaytzh.onion/20/http/1.4.7.9/forummain.htm – Read only access to the Freenet FMS forums via the Anonet Webproxy
- http://rsc2ch4h3rogiggi.onion/ – StaTorsNet
- http://hbjw7wjeoltskhol.onion – The BEST tor social network! File sharing, messaging and much more. Use a fake email to register.
- http://khqtqnhwvd476kef.onion/ – BANANARUSSA: Brazilian imageboard, with cp and jb boards. Everything is fine. The language is Brazilian Portuguese.
- http://jahfuffnmfytotlv.onion/ – The Open Mic
- http://ocu3errhpxppmwr.onion – The anonymous GTS forum
- http://germanyhusicaysx.onion/search/recent/ – German anti censorship community
- http://nifgk5szbodg7qbo.onion – TCF – Tor Carding Forms + Market
- http://t4is3dhdc2jd4ytw.onion/ – OnionForum 3.0 – New Onionforum for general talk, now with marketplace
• http://zw3crggtadila2sq.onion/imageboard/ – TorChan – One of the oldest chans on Tor
• Email and Messaging
  • http://bitmailendavkbec.onion – swiss email
  • http://365u4txyqfy72nul.onion/ – Anonymous E-mail service. You can only communicate with other users currently using this service. So tell all your friends about it!
  • http://sms4tor3vcr2geip.onion/ – SMS4TOR – Self destructing messages
  • http://notestjxctkwbk6z.onion/ – NoteBin – Create encrypted self-destructing notes
  • http://torbox3uiot6wchz.onion/ – [TorBox] The Tor Mail Box
  • http://u6lyst27lmelm6oy.onion/index.php – Blue matrix chat NOT UP ALL THE TIME so chek often to see when it is
  • http://u4uoz3aphqbd7c754.onion/ – Hell Online
  • http://f3ljvgyyujmnfhvi.onion/ – New email host of DW. The URSSMail: anonymous and, most important, SECURE! Located in 3 different servers from across the globe, we will not be taken down! Even if the FBI reach us, our users will be totally secured! All interactions made by our servers are encrypted by an AES 256 (Serpent-Twofish-AES) key, including the content of those messages. Not the admins, or even the FBI can access the content of your mailbox.

9.2.7 POLITICAL
• http://faerieuaahqvzgby.onion/ – Fairie Underground
• http://2r2tz6wzgh7gaji7.onion/ – Kavkaz Center
• http://tnysbtbxsf356hiy.onion/ – The New Yorker Strongbox
• http://duskgytldkxiuq6.onion/ – Example rendezvous points page
• http://rrcc5uuudhh4oz3c.onion/ – The Intel Exchange Forum :: Information and discussion on various topics, ranging from Illegal Activities and Alternative Energy, to Conspiracy Theories and Hacking. Same people from SnapBBS on a fully secure, moderated and categorized forum.
• http://opnju4nyz7wbypme.onion/weblog/index.html – A7B blog :: a blog dedicated to the restoration of a limited constitutional republic in the USA
• http://assmkedzgorodn7o.onion/ – Anonymous, safe, secure, crowdfunded assassinations.
• http://duskgytldkxiuq6.onion/comsense.html – Commo Sense by Thomas Paine
• http://nwycvryrozllb42g.onion/ – Destination Unknown
• http://zbnnr7qzaxlk5tms.onion/ – Wiki Leaks

9.2.8 HACKING
• http://salted7fpnlaguiq.onion/ – SALT
• http://yj5rbziqtulgidy.onion/ – Itanimulli
• http://bbxdfsru7lmmbj32.onion/marketplace/ – Delta Initiative
• http://2ogmrlf2dthnwkez.onion – Rent-A-Hacker
• http://5xki35vc4g5ts6gc.onion – GTF Greek Tor Forum For Greek Speaking Users
• http://doxinumfxyytnh.onion/ – There are no limits on what kind of info you can post
• http://qw7rh24kakedgupf.onion.lu – Hidden TeamViewer software for anonymous access to any PC

9.2.9 WAREZ
• http://2gxzwznj52jutais.onion/ – The Nowhere Server (restored from backup after FH)
• http://jntlesnev5o7zyysa.onion/ – The Pirate Bay – Torrents
• http://am4wuhz3zifexz5u.onion/ – Tor Library – library of books and other media files
• http://uj3wazyk5u4hnvtk.onion/ – The Pirate Bay – Torrents (official .onion)
• http://doxbindtelxceher.onion/ – DOXBIN
• http://wuvdsbmbwyjzsgs.onion/ – Music Downloads
• http://lolicon575r3tm5.onion/ – Lolicore and Speedcore Music
• http://xfmr77i3lxicuc.onion/ – ebooks
• http://vt27twhtksyvirk.onion/ – lol 20th Century Western Music Recordings and Scores
• http://2ygbaezjdmacnro.onion/ – Pony at Noisebridge
• http://xfmr77i3lxicuc.onion/ – Imperial Library of Trantor
• http://c3jemx2ube5v5zpg.onion/ – Jotunbane’s Reading Club
• http://wecttp234pgcxonh.onion/ – 

9.2.10 DRUGS NON-COMMERCIAL
• http://bvnyausygbt5j6iu.onion/viewforum.php?f=4 – A NEW SITE WITH SELLERS THAT ARE NOT SCAM

9.2.11 EROTIC 18+
• http://tklxss3rdzdjppnl.onion/sharepass/ – SharePass – Password sharing community
• http://k4jmdecppnsfe43c.onion/ – Girls Released – Some nice model pics
• http://54dgeda4lk6iypui.onion/ – Gallery – Met-Art, FTVX etc sets
• http://pinkmethylnlenz.onion/ – The Pink Meth (mirror)
• http://2f0qjzhou2bh7yevom.onion/klixen/ – Klixen
• http://orsxvca7gswuwe07.onion/ – EroDir – Lots and lots of Hentai
• http://mmqgh3rgeswrlqzdr.onion/ – VOR-COM

9.2.12 NON-ENGLISH
• http://germanymusicaysx.onion – Deutschland im Deep Web – German forum
• http://ffi5v46ttwggx3fby.onion/ – Das ist Deutschland hier 2.0 – German Board
Annex A – Methodology

This systematic review develops on the methodological guidelines of the Campbell Collaboration, which publishes high-level systematic reviews in several fields, including “Crime and Justice”.

- <http://paisleli66axejos.onion/> – PAIS
- <http://hyjmkmb3lfymiprp.onion/hen/papieze/> – Dziecięca pedofilia
- <http://runionv62ul3roit.onion/> – Russian Onion Union
- <http://s6cco2jylmxqcdeh.onion/> – ñitos bumps
- <http://5xki35vc4g5ts6gc.onion> – GTF Greek Tor Forum . For greek speaking users
- <http://cipollatnumrrahd.onion/index.php> – Cipolla 2.0 – Italian Community
- <http://runionv62ul3roit.onion> – Russian community: market and anonymous talks about security, guns etc.
- <http://amberoadychffmyw.onion> – Amberoad – russian anonymous market
- <http://r2d2akbw3jpt4zbf.onion> – R2D2 – russian anonymous market (drugs only)
- <http://o2tu5zjxjlibrary.onion/> – Biblioteca Alexandrina
- <http://zgjiyrtyam276uogb.onion/> – Thorlauta
- <http://5xki35vc4g5ts6gc.onion> – GTF Greek Tor Forum For Greek Speaking Users
- <http://wfi2sm5thn4piicq.onion> – Server per blog: posso creare aree individuali per singolo utente con controllo totale dell0utente sul proprio blog
- <http://s6cco2jylmxqcdeh.onion> – Cebolla-Chan – el tor chan en espanol – mexico espana argentina
- <http://hw5cve7w5ehyk1vr.onion> – Redes de Computadores – instala��es de software e configura��o
- <http://opzqhu5iftmfoor3.onion/> – Quelques livres...
Scope
This systematic review aims at identifying the online activities of organised crime and terrorist networks. The focus is on empirical studies that provided evidence-based findings about the commonly reported online activities described in Section 2.2.

Data sources

Electronic databases search
The electronic database search is the main literature source, including academic and grey literature. Several databases have been explored before choosing the included ones. Both social sciences and computing databases have been tested. After this preliminary research, databases related to social sciences prevailed, due to the observed major classification of relevant studies as social sciences. A general Computing Database from ProQuest served as a complementary source for avoiding biases. Two meetings with librarians helped assessing (a) the validity of the research terms and the queries and (b) the inclusion of all relevant databases.

Table 43 reports the list of databases (n=6), sub-databases (n=11), and the search strategy selected according to the principle of “best choice available”. When available, the best choice selected was “title, abstract and keywords”. When unavailable, the preferred choice –in order by relevance– was: full text search; title and abstract; abstract. No time span has been set up. Both published and unpublished literature has been included. The language selected was English language. All criteria mentioned apply both to organised crime and to terrorism searches. All databases have been searched through their electronic access.

Table 43 – List of databases and search strategies

<table>
<thead>
<tr>
<th>Database</th>
<th>Sub-database</th>
<th>Search Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBSCO</td>
<td>Criminal Justice Abstracts</td>
<td>Abstract or Author-Supplied Abstract</td>
</tr>
<tr>
<td></td>
<td>Full text</td>
<td></td>
</tr>
<tr>
<td>Open Grey</td>
<td></td>
<td>Full Text</td>
</tr>
<tr>
<td>ProQuest</td>
<td>Social Sciences Premium</td>
<td>Abstract</td>
</tr>
<tr>
<td></td>
<td>NJCRS</td>
<td>Abstract</td>
</tr>
<tr>
<td></td>
<td>PsycInfo</td>
<td>Abstract</td>
</tr>
<tr>
<td></td>
<td>Abi/Inform</td>
<td>Abstract</td>
</tr>
<tr>
<td></td>
<td>International Bibliography of the Social Sciences</td>
<td>Abstract</td>
</tr>
<tr>
<td></td>
<td>Public Health Database</td>
<td>Abstract</td>
</tr>
<tr>
<td></td>
<td>Military Database</td>
<td>Abstract</td>
</tr>
<tr>
<td></td>
<td>EconLit</td>
<td>Abstract</td>
</tr>
<tr>
<td></td>
<td>PsycArticles</td>
<td>Abstract</td>
</tr>
<tr>
<td></td>
<td>Computing Database</td>
<td>Abstract</td>
</tr>
<tr>
<td>PubMed</td>
<td></td>
<td>Title and Abstract</td>
</tr>
<tr>
<td>Scopus</td>
<td></td>
<td>Title, Abstract and Keywords</td>
</tr>
<tr>
<td>Web of Science</td>
<td></td>
<td>Title, Abstract and Keywords</td>
</tr>
</tbody>
</table>
Table 45 reports the queries used for each database. The search terms fall into three main categories (Table 44). The structure of the research queries has two levels, connected through the Boolean Operator “AND”:

1) The first part of the query, composed of keywords related to the concept of **online activities** (Category A), represents the common element of the two searches, guaranteeing their symmetry. This part of the query is generic and inclusive of concepts relating to potential online activities of organised crime and terrorist networks. Keywords related to **Category A** are separated by the Boolean Operator “OR”.

2) The second part of the query differentiates between **organised crime** (Category B) and **terrorism** (Category C). In both cases, the level of complexity is broad enough to include all relevant actors according to the theoretical definitions of organized crime and terrorism adopted by this systematic review. The keywords of this second part of the query are separated by the Boolean Operator “OR”.

*Table 44 – Search categories and related search terms*
<table>
<thead>
<tr>
<th>Category A</th>
<th>Category B</th>
<th>Category C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activities in cyberspace</td>
<td>Terrorism</td>
<td>Organised crime</td>
</tr>
<tr>
<td>cyberspace</td>
<td>organised crime</td>
<td>terroris*</td>
</tr>
<tr>
<td>deep web</td>
<td>organized crime</td>
<td>terror</td>
</tr>
<tr>
<td>dark web</td>
<td>criminal organisation</td>
<td>terrorist group*</td>
</tr>
<tr>
<td>dark net</td>
<td>criminal organization</td>
<td>terrorist network*</td>
</tr>
<tr>
<td>online activit*</td>
<td>criminal association</td>
<td>international terroris*</td>
</tr>
<tr>
<td>internet activit*</td>
<td>crime facilitat*</td>
<td>domestic terroris*</td>
</tr>
<tr>
<td>online presence</td>
<td>crim* network</td>
<td>religious terroris*</td>
</tr>
<tr>
<td>internet presence</td>
<td>criminal group*</td>
<td>freedom fighter*</td>
</tr>
<tr>
<td>online visibili*</td>
<td>crime facilitat*</td>
<td>lone wolf</td>
</tr>
<tr>
<td>internet visibili*</td>
<td>Mafia</td>
<td>lone wolves</td>
</tr>
<tr>
<td>Cybercrime*</td>
<td>mafios*</td>
<td>lone actor</td>
</tr>
<tr>
<td>social network</td>
<td>italian mafia</td>
<td>suicide bomb*</td>
</tr>
<tr>
<td>social networking site</td>
<td>russian mafia</td>
<td>suicide attack*</td>
</tr>
<tr>
<td>social media</td>
<td>japanese mafia</td>
<td>bombing</td>
</tr>
<tr>
<td>online platform</td>
<td>chinese mafia</td>
<td>bomber*</td>
</tr>
<tr>
<td>twitter</td>
<td>drug trafficking organ*</td>
<td>political violence</td>
</tr>
<tr>
<td>facebook</td>
<td>Dto</td>
<td>attacker*</td>
</tr>
<tr>
<td>blogger</td>
<td>drug cartel*</td>
<td>extremis*</td>
</tr>
<tr>
<td>instagram</td>
<td>narco*</td>
<td>insurgen*</td>
</tr>
<tr>
<td>reddit</td>
<td>gang*</td>
<td>radicalised</td>
</tr>
<tr>
<td>myspace</td>
<td>street gang*</td>
<td>radicalized</td>
</tr>
<tr>
<td>youtube</td>
<td>criminal gang*</td>
<td>radicalisation</td>
</tr>
<tr>
<td>wordpress</td>
<td>youth gang*</td>
<td>radicalization</td>
</tr>
<tr>
<td>vine</td>
<td>motorcycle gang*</td>
<td>fundamentalis*</td>
</tr>
<tr>
<td>tencent qq</td>
<td>boss*</td>
<td>separatis*</td>
</tr>
<tr>
<td>renren</td>
<td>smuggler*</td>
<td>radical islam*</td>
</tr>
<tr>
<td>vk</td>
<td></td>
<td>militant islam*</td>
</tr>
<tr>
<td>vkontakte</td>
<td></td>
<td>jihadi*</td>
</tr>
<tr>
<td>internet</td>
<td></td>
<td>islamic state</td>
</tr>
<tr>
<td>website</td>
<td></td>
<td>daesh</td>
</tr>
<tr>
<td>online selling</td>
<td></td>
<td>al qaeda</td>
</tr>
<tr>
<td>internet financing</td>
<td></td>
<td>far right</td>
</tr>
<tr>
<td>internet policing</td>
<td></td>
<td>right wing</td>
</tr>
<tr>
<td>internet crim*</td>
<td></td>
<td>hard right</td>
</tr>
<tr>
<td>online communication strateg*</td>
<td></td>
<td>nazi*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>fasci*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>far left</td>
</tr>
<tr>
<td></td>
<td></td>
<td>left wing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>anarchist terroris*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>communist terroris*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>guerrilla movement*</td>
</tr>
<tr>
<td>Database</td>
<td>Query</td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>----------------------------------------------------------------------</td>
<td></td>
</tr>
</tbody>
</table>
| EBSCO         | AB (cyberspace OR "dark web" OR "deep web" OR "dark net" OR online activit* OR | **Terrorism**
|               | internet activit* OR "online presence" OR "internet presence" OR |                                
|               | "online visibility" OR "internet visibility" OR cybercrime OR "social |                                
|               | networking site" OR "social media" OR "online platform" OR twitter OR |                                
|               | facebook OR blogger OR instagram OR reddit OR myspace OR youtube OR |                                
|               | wordpress OR vine OR "tencent QQ" OR renren OR vk OR VKontakte |                                
|               | OR internet OR "online selling" OR "internet policing" OR |                                
|               | "internet crime" OR "online communication strategies" OR website*) |                                
|               | AND AB (terroris* OR terror OR terrorist group* OR terrorist network* |                                
|               | OR international terrorist group* OR freedom fighter* OR "lone |                                
|               | wolf" OR "lone wolves" OR lone actor* OR suicide bomb* OR |                                
|               | suicide attack* OR bombing OR bomber* OR political violence |                                
|               | OR attacker* OR extremis* OR insurgen* OR radicalised OR |                                
|               | radicalised OR radicalisation OR fundamentalis* OR international |                                
|               | terroris* OR domestic terroris* OR separatis* OR religious |                                
|               | terroris* OR radical islam* OR militant islam* OR jihadi* OR |                                
|               | islamic state OR al qaeda OR daesh OR far right OR |                                
|               | right wing OR hard right OR far left OR left wing OR |                                
|               | anarchist terroris* OR communist terroris* OR guerrilla movement* OR |                                
|               | nazi* OR fasci*) lang:"en"                                          |
| Open Grey     | (cyberspace OR "dark web" OR "deep web" OR "dark net" OR online |                                
|               | activit* OR internet activit* OR "online presence" OR |                                
|               | "internet presence" OR "online visibility" OR |                                
|               | "internet visibility" OR cybercrime OR "social network" OR |                                
|               | "social networking site" OR "social media" OR "online platform" |                                
|               | OR twitter OR facebook OR blogger OR instagram OR reddit OR |                                
|               | myspace OR youtube OR wordpress OR vine OR "tencent QQ" OR |                                
|               | renren OR vk OR VKontakte OR internet OR "online selling" OR |                                
|               | "internet policing" OR "internet crime" OR "online communication |                                
|               | strategies" OR website*) AND (terroris* OR terror OR |                                
|               | terrorist group* OR international terrorist group* OR freedom |                                
|               | fighter* OR "lone wolf" OR "lone wolves" OR lone actor* OR |                                
|               | suicide bomb* OR suicide attack* OR bombing OR bomber* OR |                                
|               | political violence OR attacker* OR extremis* OR insurgen* OR |                                
|               | radicalised OR radicalised OR radicalisation OR |                                
|               | fundamentalis* OR international terroris* OR domestic |                                
|               | terroris* OR separatis* OR religious terroris* OR |                                
|               | radical islam* OR militant islam* OR |                                
|               | jihadi* OR islamic state OR al qaeda OR daesh OR |                                
|               | far right OR right wing OR hard right OR far left OR left wing |                                
|               | OR anarchist terroris* OR communist terroris* OR guerrilla |                                
|               | movement* OR nazi* OR fasci*) lang:"en"                            |
| ProQuest      | AB(cyberspace OR "dark web" OR "deep web" OR "dark net" OR online |                                
|               | activit* OR internet activit* OR "online presence" OR |                                
|               | "internet presence" OR "online visibility" OR |                                
|               | "internet visibility" OR cybercrime OR "social network" OR |                                
|               | "social networking site" OR "social media" OR "online platform" |                                
|               | OR twitter OR facebook OR blogger OR instagram OR reddit OR |                                
|               | myspace OR youtube OR wordpress OR vine OR "tencent QQ" OR |                                
|               | renren OR vk OR VKontakte OR internet OR "online selling" |                                
|               | OR internet financing OR "internet policing" OR |                                
|               | "internet crime" OR online communication strategy OR website*) |                                
|               | AND AB(terroris* OR terror OR terrorist group* OR |                                
|               | international terrorist group* OR freedom fighter* OR |                                
|               | "lone wolf" OR "lone wolves" OR lone actor* OR suicide |                                
|               | bomb* OR suicide attack* OR bombing OR bomber* OR |                                
|               | political violence OR attacker* OR extremis* OR insurgen* OR |                                
|               | radicalised OR radicalised OR radicalisation OR |                                
|               | fundamentalis* OR international terroris* OR domestic |                                
|               | terroris* OR separatis* OR religious terroris* OR |                                
|               | radical islam* OR militant islam* OR jihadi* OR islamic |                                
|               | state OR al qaeda OR daesh OR far right OR right wing OR |                                
|               | hard right OR far left OR left wing OR anarchist terroris* OR |                                
|               | communist terroris* OR guerrilla movement* OR nazi* OR |                                
|               | fasci*)                                                             |
| PubMed        | (online activit[Title/Abstract] OR cyberspace[Title/Abstract] |                                
|               | OR online visibility[Title/Abstract] OR social network[Title/Abstract] |                                
|               | OR social media[Title/Abstract] OR internet policing[Title/Abstract] |                                
|               | OR online selling[Title/Abstract] OR internet crime[Title/Abstract] |                                
|               | ) AND ("terrorism"[Title/Abstract] OR |                                
|               | "terrorist group"[Title/Abstract] OR "freedom fighter"[Title/Abstract] |                                
|               | OR "lone wolf"[Title/Abstract] OR "radicalized"[Title/Abstract] |                                
|               | OR radicalised[Title/Abstract] OR |                                
|               |
"extremist"[Title/Abstract])

**Scopus**

(TITLE-ABS-KEY ((cyberspace OR "online activity" OR "online visibility" OR "social network" OR "social media" OR "internet policing" OR "online selling" OR "internet crime") ) AND TITLE-ABS-KEY ((terrorism OR "terrorist group" OR "freedom fighter" OR "lone wolf" OR radicalized OR radicalised OR extremist ) ) AND (EXCLUDE (SUBJAREA, "MEDI") OR EXCLUDE (SUBJAREA, "PHYS") OR EXCLUDE (SUBJAREA, "BIOC") OR EXCLUDE (SUBJAREA, "EART") OR EXCLUDE (SUBJAREA, "ENVI") OR EXCLUDE (SUBJAREA, "HEAL") OR EXCLUDE (SUBJAREA, "NURS") OR EXCLUDE (SUBJAREA, "CENG") OR EXCLUDE (SUBJAREA, "IMMU") ) AND (LIMIT-TO (LANGUAGE, "English"))

**Web of Science**

(((TI=(cyberspace OR "dark web" OR "deep web" OR "dark net" OR online activit* OR internet activit* OR "online presence" OR "internet presence" OR "online visibility" OR "internet visibility" OR cybercrime OR "social network" OR "social networking site" OR "social media" OR "online platform" OR twitter OR facebook OR blogger OR instagram OR reddit OR myspace OR youtube OR Wordpress OR vine OR "tencent QQ" OR renren OR vk OR Vkontakte OR internet OR "online selling" OR "internet policing" OR "online policing" OR "internet crime" OR "online communication strategies" OR website*) AND TI=(terroris* OR terror OR terrorist group* OR terrorist network* OR international terrorist group* OR freedom fighter* OR "lone wolf" OR "lone wolves" OR lone actor* OR suicide bomb* OR suicide attack* OR bombing OR bomber* OR political violence OR attacker* OR extremist* OR insurgen* OR radicalised OR radicalized OR radicalisation OR radicalization OR fundamentalis* OR international terrorism* OR domestic terrorism* OR separatist* OR religious terrorism* OR radical islam* OR militant islam* OR jihadi* OR islamic state OR al qaeda OR daesh OR far right OR right wing OR hard right OR far left OR left wing OR anarchist terrorism* OR communist terrorism* OR guerrilla movement* OR nazi* OR fasci*)))) AND LANGUAGE: (English)

**Organised crime**

**EBSCO**

Criminal Justice Abstracts Full text

AB ((cyberspace OR "dark web" OR "deep web" OR "dark net" OR online activit* OR internet activit* OR "online presence" OR "internet presence" OR "online visibility" OR "internet visibility" OR cybercrime OR "social network" OR "social networking site" OR "social media" OR "online platform" OR twitter OR facebook OR blogger OR instagram OR reddit OR myspace OR youtube OR Wordpress OR vine OR "tencent QQ" OR renren OR vk OR Vkontakte OR internet OR "online selling" OR "internet policing" OR "online policing" OR "internet crime" OR "online communication strategies" OR website*) AND AB (("criminal organisation" OR "criminal organisation association" OR "organized crime" OR "organised crime" OR mafia* OR mafios* OR crim* OR "organized crime" OR "criminal association" OR "criminal organisation" OR "organized crime" OR crime facil* OR drug traffick* OR drug traffick* OR drug trafficking organ* OR drug cartel* OR criminal group* OR crime facilit* OR drug cartel* OR narco* OR "Italian Mafia" OR "Russian Mafia" OR "Japanese Mafia" OR "Chinese Mafia" OR gang* OR street "gang" OR criminal gang* OR youth gang* OR motorcycle gang" OR "boss" OR smuggler* OR trafficker* OR maritime piracy) )

**Open Grey**

(cyberspace OR "dark web" OR "deep web" OR "dark net" OR online activit* OR internet activit* OR "online presence" OR "internet presence" OR "online visibility" OR "internet visibility" OR cybercrime OR "social network" OR "social networking site" OR "social media" OR "online platform" OR twitter OR facebook OR blogger OR instagram OR reddit OR myspace OR youtube OR Wordpress OR vine OR "tencent QQ" OR renren OR vk OR Vkontakte OR internet OR "online selling" OR "internet policing" OR "online policing" OR "internet crime" OR "online communication strategies" OR website*) AND (terroris* OR terror OR terrorist group* OR terrorist network* OR international terrorist group* OR freedom fighter* OR "lone wolf" OR "lone wolves" OR lone actor* OR suicide bomb* OR suicide attack* OR bombing OR bomber* OR political violence OR attacker* OR extremist* OR insurgen* OR radicalised OR radicalized OR radicalisation OR radicalization OR fundamentalis* OR international terrorism* OR domestic terrorism* OR separatist* OR religious terrorism* OR radical islam* OR militant islam* OR jihadi* OR islamic state OR al qaeda OR daesh OR far right OR right wing OR hard right OR far left OR left wing OR anarchist terrorism* OR communist terrorism* OR guerrilla movement* OR nazi* OR fasci*))

**ProQuest**

AB((cyberspace OR "dark web" OR "deep web" OR "dark net" OR online activit* OR internet activit* OR "online presence" OR "internet presence" OR "online visibility" OR "internet visibility" OR cybercrime OR "social network" OR "social networking site" OR "social media" OR "online platform" OR twitter OR facebook OR blogger OR instagram OR reddit OR myspace OR youtube OR Wordpress OR vine OR "tencent QQ" OR renren OR vk OR Vkontakte OR internet OR "online selling" OR "internet policing" OR "online policing" OR "internet crime" OR "online communication strategies" OR website*) AND (terroris* OR terror OR terrorist group* OR terrorist network* OR international terrorist group* OR freedom fighter* OR "lone wolf" OR "lone wolves" OR lone actor* OR suicide bomb* OR suicide attack* OR bombing OR bomber* OR political violence OR attacker* OR extremist* OR insurgen* OR radicalised OR radicalized OR radicalisation OR radicalization OR fundamentalis* OR international terrorism* OR domestic terrorism* OR separatist* OR religious terrorism* OR radical islam* OR militant islam* OR jihadi* OR islamic state OR al qaeda OR daesh OR far right OR right wing OR hard right OR far left OR left wing OR anarchist terrorism* OR communist terrorism* OR guerrilla movement* OR nazi* OR fasci*))

lang:"en"
**Other literature sources**

Two additional sources contributed to the final list of included studies. The objective was (a) to complement the database of studies, avoiding gaps in the literature, and (b) to assess the consistency of the databases.

1) **Experts’ contribution:** Five experts (Table 46) on the field of organised crime and terrorist networks online (i.e. researchers, professors) have been contacted in the final steps of the systematic review. They
suggested relevant studies, and studies too recent to be catch by the electronic database search.

2) References: The references screened included the lists of references of the relevant academic and grey literature, and the tables of contents of the edited books.

Table 46 – Contributing experts

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuela Caiani</td>
<td>Institute of Scienze Umane e Sociali at Scuola Normale Superiore</td>
</tr>
<tr>
<td>David Décary-Hétu</td>
<td>University of Montreal</td>
</tr>
<tr>
<td>Jytte Klausen</td>
<td>The Wilson Center in Washington D.C.</td>
</tr>
<tr>
<td>Eric Rutger Leukfeldt</td>
<td>NSCR &amp; Director Cybersecurity &amp; SMEs at The Hague University of Applied Sciences</td>
</tr>
<tr>
<td>Rolf van Wegberg</td>
<td>Faculty of Technology, Policy and Management of Delft University of Technology; TNO</td>
</tr>
</tbody>
</table>

Selection of studies

Covidence software was selected for the screening of the studies. Covidence automatically removed the duplicates in the two databases of records. The number of records after removing the duplicates were:

- OC: n=7,758 records
- TR: n=18,373 records

A team of researchers performed a multiple evaluation of the studies to include. At least two researchers evaluated each study.

Title and abstract screening

A double round title and abstract screening allowed for implementing a rigorous method, which could be followed by all the researchers involved in title and abstract screening. The first round title and abstract screening was based on excluding criteria of irrelevance. The researchers excluded all studies not pertaining to terrorism, organised crime, Internet crimes and crime facilitated by the Internet. For example, the researchers considered as irrelevant studies with overlapping keywords from other fields (e.g., medical studies on pathologies; country reports; geopolitical analyses of geographical regions; press releases and news on politicians’ declarations). A senior researcher resolved the conflicts.

Most of the studies have been excluded at this first step. About 10% of studies met the inclusion characteristics both for terrorism and organised crime:

- OC: n=762 records
- TR: n=1866 records

The second round title and abstract screening was based on including criteria. All those studies in which Category A was logically linked to Category B or Category C have been selected.
At this stage, irrelevant studies concerned primarily the general studies on terrorism and organised crime not related to online activities. As with the first round title and abstract screening, a senior researcher resolved the conflicts (OC n=139; TR n=477) between the two screening researchers.

After concluding the title and abstract screening, the number of results to be assessed for full text eligibility were:

- OC: n=105 records
- TR: n=420 records

The first round of title and abstract screening has been a pilot for the research team. It allowed the research team to adopt common strategies and be objective in evaluating the studies according to the selected criteria. This led to a number of conflicts lower than 10%. A senior researcher with experience in the field and in systematic review methodology resolved the conflicts. The next steps of the systematic review also observed the multiple evaluation principle, including the second round title and abstract screening with stricter exclusion criteria, the full text screening and the selection of studies according to inclusion criteria.

**Eligibility for full text screening**

Prior to the full text screening stage, the relevant documents identified through the title and abstract screening have been assessed for eligibility. Among the studies deemed suitable for full text screening, the full text of 25 of them could not be retrieved (OC n=1; TR n=24). These studies had to be excluded from the review, because a) an electronic version was not available, b) a paper-based version could not be apprehended. A number of documents have been excluded as they were not in the scope of the this study, or they were from non-scientific sources (e.g., magazines):

- OC: n=35 documents excluded
- TR: n=152 documents excluded

Before moving forward to the next steps, the two lists of selected studies have been integrated with relevant studies from the screening of (a) the lists of references of the included records and (b) the lists of contents of the included papers, grey literature, and edited books. N=2 have been added to the studies on organised crime, and n=19 studies have been added to the studies on terrorism. After the integration, the number of studies to be assessed for the full text eligibility were:

- OC: n=59 documents included
- TR: n=263 documents included

**Full text screening**

As a precondition for the selection of studies according to inclusion and exclusion criteria, all the studies included have been classified according to their characteristics and their methodologies. Three types of methodologies have been identified, including (a) quantitative empirical studies, (b) mixed quantitative and qualitative empirical studies, (c) qualitative empirical studies and (d) anecdotal studies. As for the characteristics, the studies were classified according to the year of publication, the type of document, the groups and
networks analysed, the activities analysed, the primary/secondary source of data, the method of data collection, and the unit of analysis. The assessment of the studies for the full text eligibility have then based on five excluding criteria. The excluding analysis have been based on the full text documents.

1) *Anecdotal studies*: publications that were theoretical, philosophical, or opinion-based without any mention that a systematic empirical analysis was undertaken.
   - OC: n=16 anecdotal studies excluded
   - TR: n=90 anecdotal studies excluded

2) *Tools*: studies presenting and testing computational or technical methodologies for the study of criminal online activities, which did not elaborate on substantive conclusions and did not present any scientifically generalizable results of the study.
   - OC: n=6 tools excluded
   - TR: n=47 tools excluded

3) *Type of publication*: documents which do not possess the required level of methodological sophistication (e.g., master's dissertations).
   - OC: n=0 publications excluded
   - TR: n=4 publications excluded

4) *Out of scope*: studies not falling within the scope of this systematic review, including the definitions of organised crime and terrorist networks and of online activities considered, and the objectives of the analysis (e.g., results focused on the network analysis of the organised crime and terrorist networks examined).
   - OC: n=5 out of scope excluded
   - TR: n=6 out of scope excluded

5) *Double*: studies published in different journals/edited books, but presenting the same characteristics, including the authors, the samples analysed, the final results. In this case, the criteria applied for including one of the double studies were, in order by relevance, were: (a) the type of publication - papers from peer reviewed journals prevailed; and (b) the year of publication - recent studies prevailed.
   - OC: n=2 doubles excluded
   - TR: n=2 doubles excluded

At this point of the systematic review, experts’ contribution has been included, consisting of a number of suggestions for possible relevant empirical studies.
Five experts on the field of cybercrime, terrorism, and organised crime (i.e. researchers, professors) provided suggestions. Most of the studies suggested were already present in systematic review database, and some of them have been excluded due to the selected criteria of irrelevance/exclusion.

The included studies suggested by the experts were:
- OC: n=4 experts’ contributions
- TR: n=2 experts’ contributions

The final number of studies selected for the full text screening according to their relevance to the topic and their methodological characteristics were:
- OC: n=35 studies included
- TR: n=116 studies included

The final selection of the full text studies to be included in the systematic review has been based on their methodological characteristics. Both quantitative and qualitative methodologies have been considered as appropriate for the study of online activities of organised crime and terrorist networks and the objectives of the this systematic review. Due to the large number of studies analysing data from online sources, the qualitative analysis of web contents has been evaluated as essential for providing insights on the online activities of organised crime and terrorist networks.

Prior to the application of the inclusion criteria the majority of studies selected for full text review adopted mixed-method or qualitative approach. The evaluation of the methodologies followed the guidelines from the Critical Appraisal Skills Programme (CASP). The selected inclusion criteria were based on (a) the complete elaboration of the study design, (b) a systematic data collection method and a rigorous data analysis, and (c) a clear statement of findings. 67 studies could not meet the criteria of complete elaboration of the study design (OC n=3; TR n=62).

The majority of studies that have failed to meet these criteria were following qualitative approach and failed to mention how they have approached their research objectives, why they have opted for the use of certain type of data, how they have collected the data and what methods they have used to analyse it. Thus, the researchers were not able to assess the credibility of their findings. 16 studies did not meet the criteria of a systematic data collection method and a rigorous data analysis (OC n=3; TR n=11). While these studies elaborated on their study design, they provided no information on important methodological issues, as sample size, data analysis method, or coding framework. 2 studies failed to meet the criteria of a clear statement of findings as they were more focused on the appropriateness of the methodology for the study of online activities of organised crime and terrorist networks (OC n=0; TR n=2).

**Studies included**

Out of the initial records (OC: N=10,588; TR: N=23,463), plus the studies suggested by experts or stemming from the bibliographies of relevant studies,

---

only N=25 met all criteria for inclusion in the systematic review for organised crime, and N=41 for terrorist networks. Table 47 shows the final numbers of included studies.

Table 47 – Studies included TN & OC

<table>
<thead>
<tr>
<th></th>
<th>TN</th>
<th>OC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantitative</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>Mixed</td>
<td>20</td>
<td>7</td>
</tr>
<tr>
<td>Qualitative</td>
<td>10</td>
<td>16</td>
</tr>
</tbody>
</table>
Annex B Accessability of different data sources

10 Google Drive

The core functionality of Drive apps is to download and upload files in Google Drive. However, the Drive platform provides a lot more than just storage.

1.1 Search for Files

You can search or filter files with the `files.list` method of the Drive API. These methods accept the q parameter which is a search query combining one or more search clauses. Each search clause is made up of three parts.

- **Field**
  - Attribute of the file that is searched, e.g., the attribute `name` of the file or `fulltext`.
- **Operator**
  - Test that is performed on the data to provide a match, e.g., `contains`.
- **Value**
  - The content of the attribute that is tested, e.g. the name of the file `My cool document`.

Combine clauses with the conjunctions `and` or `or`, and negate the query with `not`.

More information can be found here: https://developers.google.com/drive/v3/web/search-parameters

11 Archive.org

Internet Archive encourages developers to add media to archive.org as well as to consume and repurpose metadata and media.

https://archive.readme.io/docs
1.2 What is an item?

An item is a logical “thing” that we present on one web page on archive.org. An item may be one video file along with scans of the DVD cover, one book, one audio file, or a set of audio files that represent a CD, etc. All archive.org items have this format URL: http://archive.org/details/[identifier] (where [identifier] is unique within our system).

1.3 Traditional Search

The traditional method for API access for search is the Advanced Search API. The advanced search page describes the formats provided, and the query language for searching. We limit the number of sorted paged results returnable to 10,000. Paged sorted results are supported only until the 10,000th result. For example, the search: https://archive.org/advancedsearch.php?q=subject:palm+pilot+software&output=json&rows=100&page=5 should be fine, but requesting page=10000 is rejected.

1.4 Scraping API

To provide the ability to page deeply into the items at the Archive, we are now providing a scraping API. The scraping API uses a cursor approach. One makes a scraping API call, which will return a list of results and a cursor. The cursor can then be used to search again with the search continuing where the cursor left off. The scraping API can be found at: https://archive.org/services/search/v1/scrape. Its parameters are:

- q: the query (using the same query Lucene-like queries supported by Internet Archive Advanced Search).
- fields: Metadata fields to return, comma delimited
- sorts: Fields to sort on, comma delimited (if identifier is specified, it must be last)
- count: Number of results to return (minimum of 100)
- cursor: A cursor, if any (otherwise, search starts at the beginning)
- total_only: if this is set to true, then only the number of results is returned.

A description of these fields (in Swagger/OpenAPI format) can be found at https://archive.org/services/search/v1.
12 WordPress

1.5 REST API

The WordPress REST API is organized around REST, and is designed to have predictable, resource-oriented URLs and to use HTTP response codes to indicate API errors. The API uses built-in HTTP features, like HTTP authentication and HTTP verbs, which can be understood by off-the-shelf HTTP clients, and supports cross-origin resource sharing to allow you to interact securely with the API from a client-side web application.

The REST API provides public data accessible to any client anonymously, as well as private data only available after authentication. Once authenticated the REST API supports most content management actions, allowing you to build alternative dashboards for a site, enhance your plugins with more responsive management tools, or build complex single-page applications.

1.6 REST API Developer Endpoint Reference

<table>
<thead>
<tr>
<th>Resource</th>
<th>Base Route</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posts</td>
<td>/wp/v2/posts</td>
</tr>
<tr>
<td>Post Revisions</td>
<td>/wp/v2/revisions</td>
</tr>
<tr>
<td>Categories</td>
<td>/wp/v2/categories</td>
</tr>
<tr>
<td>Tags</td>
<td>/wp/v2/tags</td>
</tr>
<tr>
<td>Pages</td>
<td>/wp/v2/pages</td>
</tr>
<tr>
<td>Comments</td>
<td>/wp/v2/comments</td>
</tr>
<tr>
<td>Taxonomies</td>
<td>/wp/v2/taxonomies</td>
</tr>
<tr>
<td>Media</td>
<td>/wp/v2/media</td>
</tr>
<tr>
<td>Users</td>
<td>/wp/v2/users</td>
</tr>
<tr>
<td>Post Types</td>
<td>/wp/v2/types</td>
</tr>
<tr>
<td>Post Statuses</td>
<td>/wp/v2/statuses</td>
</tr>
<tr>
<td>Settings</td>
<td>/wp/v2/settings</td>
</tr>
</tbody>
</table>
13 Tumblr

1.7 URI Structure

All Tumblr API requests start with api.tumblr.com. The next segment of the URI path depends on the type of request:

- **Blog**: get blog data or write to a blog
  
  api.tumblr.com/v2/blog/{blog-identifier}/...

- **User**: get user data or perform user actions
  
  api.tumblr.com/v2/user/...

1.8 Response Format

The API returns JSON-encoded objects (content-type: application/json). Responses vary according to the method used, but every response envelope includes these common parts:

| meta    | The meta object matches the HTTP response message:
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• <strong>status</strong>: the 3-digit HTTP Status-Code (e.g., 200)</td>
</tr>
<tr>
<td></td>
<td>• <strong>msg</strong>: the HTTP Reason-Phrase (e.g., OK)</td>
</tr>
</tbody>
</table>

| response | API-specific results |

1.9 Blog Methods

/info — Retrieve Blog Info

This method returns general information about the blog, such as the title, number of posts, and other high-level data.

**Examples**

- [https://api.tumblr.com/v2/blog/scipsy.tumblr.com/info](https://api.tumblr.com/v2/blog/scipsy.tumblr.com/info)
- [https://api.tumblr.com/v2/blog/good.tumblr.com/info](https://api.tumblr.com/v2/blog/good.tumblr.com/info)

<table>
<thead>
<tr>
<th>Method</th>
<th>URI</th>
<th>HTTP Method</th>
<th>Authentication</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>api.tumblr.com/v2/blog/{blog-identifier}/info?api_key={key}</td>
<td>GET</td>
<td>API key</td>
</tr>
</tbody>
</table>

Request Parameters
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Type</th>
<th>Description</th>
<th>Default</th>
<th>Required?</th>
</tr>
</thead>
<tbody>
<tr>
<td>blog-identifier</td>
<td>String</td>
<td>Any blog identifier</td>
<td>N/A</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>See the <a href="#">Overview</a> for more details.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>api_key</td>
<td>String</td>
<td>Your OAuth Consumer Key</td>
<td>N/A</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>See <a href="#">Authentication</a> for more details.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Responses

<table>
<thead>
<tr>
<th>Response Field</th>
<th>Type</th>
<th>Description</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>title</td>
<td>String</td>
<td>The display title of the blog</td>
<td>Compare name</td>
</tr>
<tr>
<td>posts</td>
<td>Number</td>
<td>The total number of posts to this blog</td>
<td></td>
</tr>
<tr>
<td>name</td>
<td>String</td>
<td>The short blog name that appears before tumblr.com in a standard blog hostname</td>
<td>Compare title</td>
</tr>
<tr>
<td>updated</td>
<td>Number</td>
<td>The time of the most recent post, in seconds since the epoch</td>
<td></td>
</tr>
<tr>
<td>description</td>
<td>String</td>
<td>You guessed it! The blog's description</td>
<td></td>
</tr>
<tr>
<td>ask</td>
<td>Boolean</td>
<td>Indicates whether the blog allows questions</td>
<td></td>
</tr>
<tr>
<td>ask_anon</td>
<td>Boolean</td>
<td>Indicates whether the blog allows anonymous questions</td>
<td>Returned only if ask is true</td>
</tr>
<tr>
<td>likes</td>
<td>Number</td>
<td>Number of likes for this user</td>
<td>Returned only if this is the user's primary blog and sharing of likes is</td>
</tr>
</tbody>
</table>
### 1.10 /likes — Retrieve Blog's Likes

This method can be used to retrieve the **publicly** exposed likes from a blog.

#### 1.10.1 Method

<table>
<thead>
<tr>
<th>URI</th>
<th>HTTP Method</th>
<th>Authentication</th>
</tr>
</thead>
<tbody>
<tr>
<td>api.tumblr.com/v2/blog/{blog-identifier}/likes?api_key={key}</td>
<td>GET</td>
<td>API Key</td>
</tr>
</tbody>
</table>

#### 1.10.2 Request Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Type</th>
<th>Description</th>
<th>Default</th>
<th>Required?</th>
</tr>
</thead>
<tbody>
<tr>
<td>limit</td>
<td>Number</td>
<td>The number of results to return: 1–20, inclusive</td>
<td>20</td>
<td>No</td>
</tr>
<tr>
<td>offset</td>
<td>Number</td>
<td>Liked post number to start at</td>
<td>0 (first post)</td>
<td>No</td>
</tr>
<tr>
<td>before</td>
<td>Integer</td>
<td>Retrieve posts liked <strong>before</strong> the specified timestamp</td>
<td>None</td>
<td>No</td>
</tr>
<tr>
<td>after</td>
<td>Integer</td>
<td>Retrieve posts liked <strong>after</strong> the specified timestamp</td>
<td>None</td>
<td>No</td>
</tr>
</tbody>
</table>

More endpoints can be found under [https://www.tumblr.com/docs/en/api/v2](https://www.tumblr.com/docs/en/api/v2)

### 14 Telegram

#### 1.11 Bot API

This API allows you to connect bots to our system. **Telegram Bots** are special accounts that do not require an additional phone number to set up. These accounts serve as an interface for code running somewhere on your server. To use this, you don’t need to know anything about how our MTProto encryption protocol works — our intermediary server will handle all encryption and communication with the Telegram API for you. You communicate with this
server via a simple HTTPS-interface that offers a simplified version of the Telegram API.

### 1.12 Telegram API

This API allows you to build your own customized Telegram clients. It is 100% open for all developers who wish to create Telegram applications on our platform.

#### 1.12.1 Handling Push-Notifications

To subscribe to notifications, the client must invoke the `account.registerDevice` query, passing in `token_type` and `token` as parameters that identify the current device. It is useful to repeat this query at least once every 24 hours or when restarting the application.

### 15 Steam

Valve provides these APIs so website developers can use data from Steam in new and interesting ways. They allow developers to query Steam for information that they can present on their own sites.

#### 1.13 Steam Web APIs available

- **ISteamNews**: Steam provides methods to fetch news feeds for each Steam game.
- **ISteamUserStats**: Steam provides methods to fetch global stat information by game.
- **ISteamUser**: Steam provides API calls to provide information about Steam users.
- **ITFItems_440**: Team Fortress 2 provides API calls to use when accessing player item data.

#### 1.14 Output Formats

All API calls take the form `http://api.steampowered.com/<interface name>/<method name>/v<version>/?key=<api key>&format=<format>`. Format can be any of:

- **json**: The output will be returned in the JSON format
- **xml**: Output is returned as an XML document
- **vdf**: Output is returned as a VDF file.

If you do not specify a format, your results will be returns in the JSON format.
More information can be found under
https://developer.valvesoftware.com/wiki/Steam_Web_API
Annex C – List of studies included in the results of the systematic review

Table 48 – List of studies included on terrorist networks

<table>
<thead>
<tr>
<th>Study</th>
<th>Method</th>
<th>Type of terrorism</th>
<th>Communication &amp; networking</th>
<th>Propaganda</th>
<th>Recruitment &amp; mobilisation</th>
<th>Information gathering, training, planning of attacks</th>
<th>Computer focused crime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study</td>
<td>Method</td>
<td>Type of terrorism</td>
<td>Communication &amp; networking</td>
<td>Propaganda</td>
<td>Recruitment &amp; mobilisation</td>
<td>Information gathering, training, planning of attacks</td>
<td>Computer focused crime</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>-------------------------</td>
<td>----------------------------</td>
<td>----------------------------</td>
<td>------------</td>
<td>----------------------------</td>
<td>----------------------------------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Study</td>
<td>Method</td>
<td>Type of terrorism</td>
<td>Communication &amp; networking</td>
<td>Propaganda</td>
<td>Recruitment &amp; mobilisation</td>
<td>Information gathering, training, planning of attacks</td>
<td>Computer focused crime</td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>-------------------</td>
<td>-----------------------------</td>
<td>------------</td>
<td>----------------------------</td>
<td>--------------------------------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Study</td>
<td>Method</td>
<td>Type of terrorism</td>
<td>Communication &amp; networking</td>
<td>Propaganda</td>
<td>Recruitment &amp; mobilisation</td>
<td>Information gathering, training, planning of attacks</td>
<td>Computer focused crime</td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>-------------------</td>
<td>---------------------------</td>
<td>------------</td>
<td>---------------------------</td>
<td>-----------------------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Scrivens, R. 2017. 'Searching for Signs of Extremism on the Web: An Introduction to Sentiment-Based Identification of Radical Authors'. Behavioral Sciences of Terrorism and Political Aggression: 1–21</td>
<td>Mixed empirical</td>
<td>Islamic terrorism &amp; radicalisation</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Yunus, Z., and R. Ahmad. 2014. 'Evaluating Cyber Terrorism Components in Malaysia'. Kuching, Malaysia</td>
<td>Qualitative</td>
<td>Terrorism &amp; radicalisation</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>
**Table 49 – List of studies included on criminal networks**

<table>
<thead>
<tr>
<th>Study</th>
<th>Method</th>
<th>Type of criminal network</th>
<th>Trade of illicit goods/services</th>
<th>Indirect propaganda &amp; spread of criminal culture</th>
<th>Recruitment</th>
<th>Cybercrime &amp; other computer crimes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Title</td>
<td>Journal</td>
<td>Year</td>
<td>Type</td>
<td>Keywords</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-------</td>
<td>---------</td>
<td>------</td>
<td>------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>Leukfeldt, E. R., Kleemans, E., &amp; Stol, W.</td>
<td>Cybercriminal Networks, Social Ties and Online Forums: Social Ties versus Digital Ties Within Phishing Qualitative and Malware Networks.</td>
<td><em>British Journal of Criminology</em>.</td>
<td>2017</td>
<td>Qualitative</td>
<td>Cybercriminal networks</td>
<td></td>
</tr>
<tr>
<td>Tade, O., &amp; Aliyu, I.</td>
<td>Social Organization of Internet Fraud among University Undergraduates in Nigeria.</td>
<td><em>International Journal of Cyber Criminology</em>, 5(2), 860–875.</td>
<td>2011</td>
<td>Qualitative</td>
<td>Cybercriminal networks</td>
<td></td>
</tr>
</tbody>
</table>
# Annex D – Studies on methodologies and software

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Title</th>
<th>TR/OC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adler, R.M.</td>
<td>2007</td>
<td>A dynamic social network software platform for counter-terrorism decision support</td>
<td>TR</td>
</tr>
<tr>
<td>Agarwal, S.; Sureka, A.</td>
<td>2015</td>
<td>A topical crawler for uncovering hidden communities of extremist micro-bloggers on tumblr</td>
<td>TR</td>
</tr>
<tr>
<td>Agarwal, S.; Sureka, A.</td>
<td>2015</td>
<td>Using KNN and SVM based one-class classifier for detecting online radicalization on twitter</td>
<td>TR</td>
</tr>
<tr>
<td>Agarwal, Swati; Sureka, Ashish</td>
<td>2015</td>
<td>Topic-Specific YouTube Crawling to Detect Online Radicalization</td>
<td>TR</td>
</tr>
<tr>
<td>Akhgar, B.; Tabatabayi, F.; et al.</td>
<td>2013</td>
<td>Investigating Radicalized Individual Profiles through Fuzzy Cognitive Maps</td>
<td>TR</td>
</tr>
<tr>
<td>Al-Khateeb, S.; Agarwal, N.</td>
<td>2015</td>
<td>Analyzing deviant cyber flash mobs of ISIL on twitter</td>
<td>TR</td>
</tr>
<tr>
<td>Aliprandi, C.; De Luca, A.E.; et al.</td>
<td>2014</td>
<td>CAPER: Crawling and analysing Facebook for intelligence purposes</td>
<td>OC</td>
</tr>
<tr>
<td>Ashcroft, M.; Fisher, A.; et al.</td>
<td>2015</td>
<td>Detecting Jihadist Messages on Twitter</td>
<td>TR</td>
</tr>
<tr>
<td>Balasuriya, L.; Wijeratne, S.; et al.</td>
<td>2016</td>
<td>Finding street gang members on Twitter</td>
<td>OC</td>
</tr>
<tr>
<td>Benigni, M.; Carley, K.M.</td>
<td>2016</td>
<td>From tweets to intelligence: Understanding the Islamic Jihad supporting community on twitter</td>
<td>TR</td>
</tr>
<tr>
<td>Berzinji, A.; Abdullah, F.S.; Kakei, A.H.</td>
<td>2013</td>
<td>Analysis of terrorist groups on facebook</td>
<td>TR</td>
</tr>
<tr>
<td>Cesur, R.; Ceyhan, E.B.; et al.</td>
<td>2017</td>
<td>Determination of potential criminals in social network</td>
<td>OC</td>
</tr>
<tr>
<td>Chen, H.; Denning, D.; et al.</td>
<td>2011</td>
<td>The dark web forum portal: From multi-lingual to video</td>
<td>TR</td>
</tr>
<tr>
<td>Chen, H.</td>
<td>2008</td>
<td>Sentiment and affect analysis of dark web forums: Measuring radicalization on the internet</td>
<td>TR</td>
</tr>
<tr>
<td>Chen, H.; Chung, W.; et al.</td>
<td>2008</td>
<td>Uncovering the dark Web: A case study of Jihad on the Web</td>
<td>TR</td>
</tr>
<tr>
<td>Diaz, A.R.; Choi, J.; et al.</td>
<td>2016</td>
<td>Data-driven system identification of the social network dynamics in online postings of an extremist group</td>
<td>TR</td>
</tr>
<tr>
<td>Ferrara, E.; Wang, W.Q.; et al.</td>
<td>2016</td>
<td>Predicting online extremism, content adopters, and interaction reciprocity</td>
<td>TR</td>
</tr>
<tr>
<td>Gordon, A.</td>
<td>1997</td>
<td>Terrorism on the Internet: Discovering the Unsought.</td>
<td>TR</td>
</tr>
<tr>
<td>Kaati, L.; Omer, E.; et al.</td>
<td>2015</td>
<td>Detecting Multipliers of Jihadism on Twitter</td>
<td>TR</td>
</tr>
<tr>
<td>Kaur, K.</td>
<td>2016</td>
<td>Development of a framework for analyzing terrorism actions via twitter lists</td>
<td>TR</td>
</tr>
<tr>
<td>Park, A.J.; Beck, B.; et al.</td>
<td>2016</td>
<td>Temporal analysis of radical dark web forum users</td>
<td>TR</td>
</tr>
<tr>
<td>Author</td>
<td>Year</td>
<td>Title</td>
<td>TR/OC</td>
</tr>
<tr>
<td>----------------------------</td>
<td>------</td>
<td>----------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Piorkowski, J.; Zhou, L.</td>
<td>2013</td>
<td>Content feature enrichment for analyzing trust relationships in web forums</td>
<td>TR</td>
</tr>
<tr>
<td>Qi, X.; Christensen, K.; et al.</td>
<td>2010</td>
<td>A hierarchical algorithm for clustering extremist web pages</td>
<td>TR</td>
</tr>
<tr>
<td>Qin, J. L.; Zhou, Y. L.; et al.</td>
<td>2005</td>
<td>The Dark Web portal project: Collecting and analyzing the presence of terrorist groups on the web</td>
<td>TR</td>
</tr>
<tr>
<td>Richey, Melonie K; Binz, M.</td>
<td>2015</td>
<td>Open Source Collection Methods for Identifying Radical Extremists Using Social Media</td>
<td>TR</td>
</tr>
<tr>
<td>Rico Sulayes, A.</td>
<td>2012</td>
<td>Quantitative Authorship Attribution of Users of Mexican Drug Dealing Related Online Forums</td>
<td>OC</td>
</tr>
<tr>
<td>Sureka, A.; Kumaraguru, et al.</td>
<td>2010</td>
<td>Mining YouTube to discover extremist videos, users and hidden communities</td>
<td>TR</td>
</tr>
<tr>
<td>Wadhwa, P.; Bhatia, M.P.S.</td>
<td>2013</td>
<td>Tracking on-line radicalization using investigative data mining</td>
<td>TR</td>
</tr>
<tr>
<td>Wadhwa, P.; Bhatia, M.P.S.</td>
<td>2015</td>
<td>Measuring Radicalization in Online Social Networks Using Markov Chains</td>
<td>TR</td>
</tr>
<tr>
<td>Wadhwa, P.; Bhatia, M.P.S.</td>
<td>2016</td>
<td>New Metrics for Dynamic Analysis of Online Radicalization</td>
<td>TR</td>
</tr>
<tr>
<td>Wei, Y.; Singh, L.; Martin, S.</td>
<td>2016</td>
<td>Identification of Extremism on Twitter</td>
<td>TR</td>
</tr>
<tr>
<td>Wijeratne, S; Doran, D.; et al.</td>
<td>2015</td>
<td>Analyzing the social media footprint of street gangs</td>
<td>OC</td>
</tr>
<tr>
<td>Xie, U.; Xu, J.; Lu, T.C.</td>
<td>2016</td>
<td>Automated classification of extremist Twitter accounts using content-based and […]</td>
<td>TR</td>
</tr>
<tr>
<td>Yang, C.C.; Ng, T.D.</td>
<td>2007</td>
<td>Terrorism and crime related weblog social network: Link, content analysis and information visualization</td>
<td>TR</td>
</tr>
<tr>
<td>Yang, C.C.; Tang, X.; Gong, X.</td>
<td>2011</td>
<td>Identifying dark web clusters with temporal coherence analysis</td>
<td>TR</td>
</tr>
<tr>
<td>Yunos, Z.; Ahmad, R.; et al.</td>
<td>2012</td>
<td>Illicit activities and terrorism in cyberspace: An exploratory study in the Southeast Asian region</td>
<td>TR</td>
</tr>
<tr>
<td>Zhang, Yulei; Zeng, Shuo; et al.</td>
<td>2009</td>
<td>Dark Web Forums Portal: Searching and Analyzing Jihadist Forums</td>
<td>TR</td>
</tr>
<tr>
<td>Zhou, Yilu; Reid, Edna; et al.</td>
<td>2005</td>
<td>US Domestic Extremist Groups on the Web: Link and Content Analysis</td>
<td>TR</td>
</tr>
<tr>
<td>Zulkarnine, A.T.; Frank, R.; et al.</td>
<td>2016</td>
<td>Surfacing collaborated networks in dark web to find illicit and criminal content</td>
<td>TR</td>
</tr>
</tbody>
</table>

1 Pauwells et-al, 2014