PUBLIC

CounteR

Privacy-First Situational Awareness Platform for Violent Terrorism and Crime Prediction, Counter Radicalisation and Citizen Protection

Project Acronym	CounteR
Project Full Title	Privacy-First Situational Awareness Platform for Violent Terrorism and Crime Prediction, Counter Radicalisation and Citizen Protection
Grant Agreement no	101021607
Project Duration	36 months (starting May 1st 2021)

Deliverable number 1.5

Ontology and taxonomy model

Work Package	WP 1 - Systems Specifications & Architecture
Task	Task 1.2 Technical Requirements
Lead Beneficiary	AST
Due Date	30.11.2021
Submission Date	30.11.2021
Deliverable Status	1.0
Deliverable Type	R (Report)
Dissemination Level	PU
Document Name	D1.5 Ontology and taxonomy model

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 101021607

Disclaimer

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 101021607

Editors

First Name	Surname	Beneficiary
Alexandru	Nistor	AST
Adrian	Onu	AST

Contributors

First Name	Surname	Beneficiary
Ionut	Gradinaru	AST
Adrian	Carp	AST
Guillem	Garcia	INS
Dana	Tantu	INS

Reviewers

First Name	Surname	Beneficiary
Guillem	Garcia	INS
Joaquin	Luzon Tuells	INS
Marco	Lombardi	UCSC
Alexandru	Vladuta	SPP

History of Changes

Version	Date	Change	Modified by
0.1	1.11.2021	Added the table of contents	Alexandru Nistor
0.2	12.11.2021	Proposed the ontology diagram	Adrian Carp
0.3	15.11.2021	Added the introduction and the methodology	Adrian Onu
0.4	19.11.2021	Added taxonomy and ontology details	Adrian Carp, Adrian Onu
0.5	21.11.2021	Added process flow diagram	Adrian Onu
0.6	23.11.2021	Added UML class diagram	Adrian Carp
0.7	25.11.2021	Improvements to the introduction chapter	Guillem Garcia, Dana Tantu
0.8	25.11.2021	Review and improvements	Michael Zammit, Alexandru VLADUTA
1.0	25.11.2021	Final review	Adrian Onu, Guillem Garcia, Catalin Trufin

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 101021607

Table of Contents

Executive Summary

Deliverable D1.5 Ontology and Taxonomy model in the report resulted from the T1.2 Technical Requirements task. The report introduces a specific domain language for online radicalisation detection, consisting of explicit formal specifications and the formal system for the naming of entities and processes.

Based on the D1.1 and D1.2 deliverables, this report elaborates the taxonomy and the ontology to be used for the research and implementation phase of the technical CounteR components.

This report also provides the first iteration of the basic concepts related to the radicalisation process and the relationships for the *D2.4 Radicalisation Ontology, Taxonomy and Recommendations*.

List of Abbreviations

Abbreviation	Explanation
AI	Artificial Intelligence
API	Application Programming Interface
CONOPS	Concept of operations
D	Deliverable
DE	Domain Expert
DoA	Description of the Action
DRL	Deep Reinforcement Learning
GA	Grant Agreement
KE	Knowledge Engineer
KPI	Key Performance Indicator
LEA	Law Enforcement Agency
ML	Machine Learning
MS	Milestone
NLP	Natural Language Processing
RL	Reinforcement learning
RSS	Really Simple Syndication
SABiO	Systematic Approach for Building Ontologies
SNA	Social Network Analysis
SRICE	Semantic Reasoning and Insight Correlation Engine
SQL	Structured Query Language
T	Task
UI	User Interface
UML	Unified Model Language
URL	Unified Resource Locator
WP	Work Package

List of Figures

List of Tables

1 Introduction

The *D1.5 Ontology and taxonomy model* will describe the explicit formal specifications (ontologies) and the formal system for the naming of entities and processes (taxonomy) planned to be used for the development of the CounteR platform.

The information related to concepts of the radicalisation software domain is offered in structured and unstructured forms in other deliverables and is made available through different formats and protocols. This can create an issue for a project such as CounteR with multiple partners with different expertise as to extract knowledge all this information needs to be homogenised, categorised, correlated etc.

Therefore, there is a need for a common language and a manner to communicate similar terms in a similar way so that entities on the other side can easily process them. To reach this goal, we need to define the "things" of this specific domain and how they interact with each other.

For that reason, the ontology and taxonomy description from this deliverable will aide in providing the same knowledge and understanding for the multidisciplinary partners.

While this deliverable marks the delivery of the core ontology and taxonomy that will be used throughout the project, by no means does this constitute the freeze of its development. On the contrary, these are subject to research findings and technological needs, and therefore, depending on the upcoming needs of the project, will be updated accordingly.

1.1 Relation with other tasks and deliverables

This deliverable is part of a series of deliverables from WP1: System Specifications & Architecture. In this WP, the LEAs and Internet Service Providers will define the specification requirements of the solution. The aims of this WP are:

- To specify the LEA requirements, use cases and scenarios (Task 1.1).
- To identify the technical specifications of the tool (Task 1.2).
- To determine the legal requirements of the solution and project (Task 1.3).
- To establish the commercial needs of the solution (Task 1.4).

This deliverable is related to **Task 1.2.: Definition of the Technical Requirements of the solution**. This task has the main objective of detailing the general technical requirements necessary for the project, which will be the starting point for the development of the solution.

Therefore, the consortium pays significant attention to this document, as it not only forms a common language for all CounteR partners but also represents a baseline in the development process.

This document uses the previous two deliverables *D1.1 - LEA requirements, use cases and scenarios* and *D1.2 Technical requirements* as prerequisites for building the ontology and the taxonomy models.

The output of D1.5 will provide the definitions of basic concepts in the domain of radicalisation software and the relation between them. Additionally, these results will contribute to the research

and development phase of the CounteR platform components from WP3: Data Acquisition & Management, WP4: Data Analytics for Detecting Radical Content, WP5: Data Modelling Ecosystem and WP6: Backend, Frontend and Infrastructure.

Task 2.4 Radicalisation Taxonomy, Ontology and Recommendations from work package 2 will provide more detailed insights about the specific domain taxonomy and ontology.

Table 1 - Relation to other deliverables

Deliv.#	Deliverable Title	Nature of Relation
D1.1	LEA requirements, use cases and scenarios	Deliverable D1.1 LEA requirements, use cases and scenarios introduce the specifications of the project based on end-users' contributions. It considers the real operational needs and scenarios and adapts them to the scope of the CounteR project. Entities and processes are extracted from the scenarios, functional requirements and CONOPS.
D1.2	Technical Requirements	The Technical Requirements report provided input for the ontology and the detailed description of the concepts.
D2.4	Radicalisation Ontology, Taxonomy and Recommendations	D1.5 report provides the first iteration of the basic concepts related to the radicalisation process and the relationships.
All tasks from the technical WPs: WP3-WP6	Every technical task from WP3 to WP6	D1.5 establishes the taxonomy and the ontology in the project and the output of this report will be used as a prerequisite for the next technical implementation tasks from WP3 to WP6.

1.2 Relationship with project milestones

If we take a close look to the project milestones:

Milestone number[18]	Milestone title	WP number[9]	Lead beneficiary	Due Date (in months)[17]	Means of verification
MS1	Specifications ready	WP1	2 - INS	7	Technical, commercial and regulatory requirements. Platform specifications and architecture.
MS2	Data processing & analytics modules ready	WP2, WP3, WP4, WP5	7 - ELTE	23	Data analysis modules first release ready and tested.
MS3	First prototype	WP6	1 - AST	24	First prototype of the solution completed and deployed, ready for pilots integration and validation.
MS4	Pilots completed	WP10, WP11, WP7, WP8, WP9	2 - INS	31	Training and evaluation report of all the pilots completed. Plan for future exploitation finished.

Figure 1 - Project Milestones

The main milestone related to this deliverable is Milestone 1: Specifications ready. The taxonomy explained in this deliverable is part of the specification phase as it is related to Task 1.2: Definition of the Technical Requirements of the solution.

1.3 Relation with the project objectives

Regarding the main objectives of the project, the deliverable D1.5 is related to the following objective:

- *3.1 Requirements and specifications*. The ontology will help to understand the project technical requirements adequately as it refers to forming a common language between all partners and supports the foundation of the development of the solution.

#	Main project objectives	Main project results (lead)
Objective 1: Integrate and homogenize disparate real-time and static data sources to analyse them for detecting radical and terrorism-related content.		
1.1	Enable real time collection and analysis capabilities for sources such as social media, blogs, forums, websites.	Take down of radical content in real-time.
1.2	Enable integration of static analysis like the prediction of links between suspicious users, early-detection of radical groups.	Early-detection of potential radical content.
1.3	Enable the detection of reuploaded of the removed content, by methodologies of text and image similarity.	Detection of re-uploaded content.
Objective 2: Integrate the analysis technologies with expert knowledge in order to develop models to predict risk and identify hot spots.		
2.1	Automatic detection and classification of terrorism and radical content, including innovative methods on Transfer Learning and Data Augmentation.	NLP Analysis
2.2	Level of radicalisation modelling at individual level, from several socio-psychological dimensions like: level of extreme view; level of support of violence; and psychological analysis. The level of radicalisation will be followed in time to analyse radicalisation processes.	Radicalisation Model
2.3	Development of Social Network analysis methods to deeply understand the role of the group in the radicalization process.	Social Network Analysis
2.4	Creation of an innovative active learning methodology based on Deep Learning techniques to automatically detect new radical content and users.	Deep Reinforcement Learning
Objective 3: Hands-on involvement of a wide range of European LEAS to guarantee high impact in the fight against terrorism.		
3.1	The high number of LEAs (6) and their diversity will ensure that the solution will meet their needs, through an extensive specification phase.	Requirements and specifications
3.2	As we explain in the proposal, the pilots will be one of the core phases of the project. Testing the CounteR solution is essential to adjust it to the needs of the LEAs and Internet Service Providers.	Testing
3.3	Before and during the pilots, all the technical and academic partners will provide training to the LEAs to optimize their feedback and validation of the solution.	Validation
Objective 4: Aggressive dissemination of the project results in order to achieve maximum sector (LEA) uptake and market excitement.		
4.1	Where possible, non-sensitive project deliverables will be shared with the scientific and research communities to increase the impact of the results and further research initiatives.	Academia and Scientific Sector
4.2	Non-sensitive pilot results will be shared with the LEA community in order to increase their understanding of how CounteR can help them.	Market Sector (LEAs and Internet Service Providers)
4.3	The commercial partners will develop individual exploitation plans and a joint business plan for the market-ready solution.	Exploitation

Figure 2 - Project objectives

1.4 Relation with the project KPIs

Figure 3 describes the project KPIs in detail:

Objectives			Project KPI	M12	M24	M36	WP
1	2	3					
X			Real-time performance	partial	full	full	6
X			# social media data sources	4	7	10	3
	X		# analysed languages	4	8	12	4
	X		Accuracy (precision)	70%	75%	80%	5
	X		Scalability	none	partial	full	6
	X		Development integration	partial	full	full	6
		X	Dashboard usability (1-10 scale)	5	7	9	6
		X	Privacy protection	high	high	high	7
		X	Assessment of the solution by the LEAs (scale 1-10)		> 6	> 8	7
	X		# scientific papers	2	4	6	8

Figure 3 - Project KPIs.

The ontology and the taxonomy will facilitate the development integration phase, improving the capability to meet the KPI's goal on time. A significant part of the implementation phase is done in order to prepare the architecture to meet the demands of the data collectors and radicalisation detectors, and this preparation phase needs to be done upfront in order to meet the deadline of Task 6.1 System release v1 in Month 12 of the project.

1.5 Technical risks

In relation to the project's technical risks, The D1.5 deliverable is closely related to risk number 4: *difficulties in integration of all the modules.* The taxonomy will help to reduce this risk, by developing a common schema for all the technical tasks.

#	Description	WP involved	Prob Hi-Med-Lo	Severity Hi-Med-Lo
1	Problems in Data Acquisition because of restrictions in the data sources	WP3	Med	Med
The vast amount of sources in every domain (news, social media, open data, etc.) will compensate for the lack of any of them, by replacing their information with data from another source. For example, if some of the platforms restrict access to information (Facebook), the system overcomes this limitation by indirectly acquiring this information, getting the Facebook messages cross-published on other platforms (Twitter, for example). This method is already used by some of our competitors[83]				
2	Lack of geolocated data	WP3	Med	Med
As far as the solution will analyse the factors with effect on the radicalisation of a community, the data will be, in some way or another, linked to a place. The consequence is that, if the data is not automatically geolocated (via metadata) or explicitly linked to a place, methods will be developed in order to infer a location for every message (using geo-fencing, language inference, user profile information, etc.).				
3	Accuracy of the models	WP5	Low	High
An ensemble of methods will be developed, in order to not depend on the failure/ success of a single methodology. This strategy will reduce the risk to not reach a minimum accuracy of the model for understanding the vulnerability of a community to become radicalized.				
4	Difficult integration because the heterogeneity of the modules	WP6	Low	Med
The integration phase of the project can be difficult because of the multiplicity of modules, coming from different authors and, maybe, involving different technologies and programming languages. For that reason, a powerful tool of continuous integration will be used (Jenkins). This integration environment facilitates the combination of the modules by using a common framework for outputs and inputs of all of them.				
5	Limitations of the tool because of data privacy issues	WP7	Med	Med
Because the data privacy regulations and laws can restrict the type of data the solution can process and store, this can limit the reach of the tool. The great variety of types of data collected (Open data, social media, news, dark web) mitigate the impact of these limitations in the data acquisition, just replacing them by data from other sources.				

Figure 4 - Project technical risks

1.6 Commercial and management risks

In terms of commercial risks presented in Figure 5, some of the outcomes of the current deliverable are relevant for the mitigation measures related to the commercial risks. Specifically on the lack of resources to complete the solution. The ontology will help to improve the time spent in the technical parts.

6	Development of a similar tool by a competitor	WP6	Med	Med
The combination of multidisciplinary expertise plus the participation of LEAs give the project a competitive advantage, compared to companies, even if they are leaders of the security sector. In this respect, the combination of developing the tool from very different perspectives (experts on psychology, sociology, Machine Learning, Big Data) and the constant contact with the end-user is a competitive advantage that makes the difference from other solutions.				
7	Lack of resources to complete the solution successfully	WP6	Low	Med
Given the high complexity of the solution that we will develop, there is a risk of underestimation of the required resources. For this reason, an in-depth analysis of each task has been developed before the writing of the proposal. The methodology used to evaluate the task effort is based on the evaluations from 3 different evaluation teams. This method is useful to avoid underestimating the development time of each module. However, this risk will be mitigated additionally by a continuous monitoring of the technical development carried out by the coordinator, AST.				
8	Insufficient collaboration on the part of the LEAs	WP9	Low	Med
The engagement of the LEAs into the project will be guaranteed using a set of methods to facilitate their participation. These methods will include surveys, collaboration meetings, workshops, and other procedures. The constant information to them about the project progress and outcomes will also be one of the ways to enhance their commitment to the project.				
9	Launch of similar tools	WP9	Med	Med
Via our Innovation Management efforts, the consortium will constantly monitor similar technologies and tools to understand the competitive advantage compared to them. The exploitation plan will be adjusted to this information.				

Figure 5 –Project risks

1.7 Deliverable structure

The D1.5 Ontology and taxonomy model is split into the following parts:

- **Introduction** – describes the need, relation with the other tasks and deliverables and the methodology used to obtain the deliverable results.
- **Taxonomy** – describes the "dictionary" of the project, the terms used to describe the entities, the processes implemented in the platform and the relationship in between.
- **Ontology** – conceptualises the entities and the processes into classes and defines the detailed relationships between the concepts.

2 Methodology

The content of this deliverable is refined based on the scenarios evoked from D1.1 and from the technical requirements from D1.2 with an emphasis on the wireframes, which were previously validated by the LEAs in terms of terminology, processes, and operations.

SABiO (Falbo, 2013) methodology is a systematic approach for building ontologies, and it is considered for the conceptualisation and the development phase. SABiO development process contains five main phases. Each phase includes activities performed by workers with specific expertise and skill set. The main roles are domain experts and end-users, engineers, designers and testers. The main phases are:

- Ontology Purpose Identification and Requirements Elicitation (from D1.1)
- Ontology Capture and Formalization (from D1.2, D1.5)
- Ontology Design (D1.2, UI wireframes, T2.4)
- Ontology Implementation (Technical tasks from WP3 to WP6)
- Ontology Testing (D6.4, WP8 Pilots)

The same approach but with four phases (Inception, Elaboration, Construction, Transition) is described in the paper "A software engineering approach to ontology building" [1] . Figure 6 states that the elaboration and construction phases is critical to ontology elaboration, as this phase provides most of the lexicon, glossary and semantic network. The D1.5 deliverables deals with the Inception phase of the ontology of radicalistion domain. In the WP2-WP6 the ontology will go through the Elaboration and Construction phases, and WP8 will consist in the Transition phase.

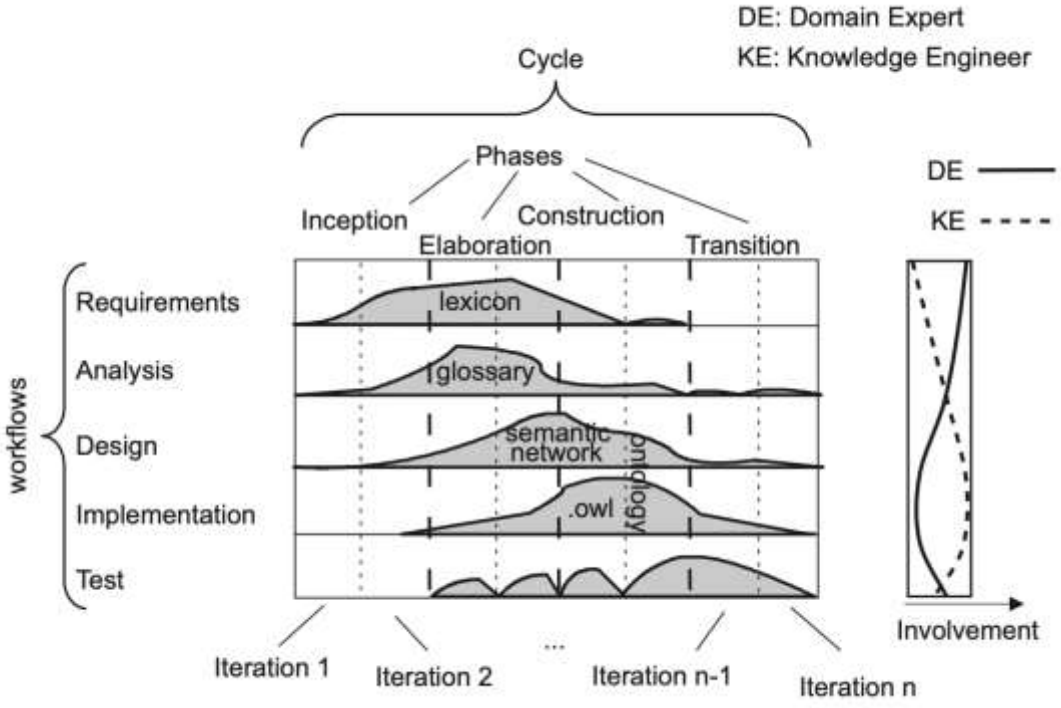

Figure 6 – Domain Experts and Knowledge Engineers involvement in workflows and phases [2]

The process of transforming the user requirements into software ontologies is also described in the paper *Towards an Ontology of Software: A Requirements Engineering Perspective* [1], the Figure 6 describing the process of introducing the abstract software artifacts into the requirements gathering phases.

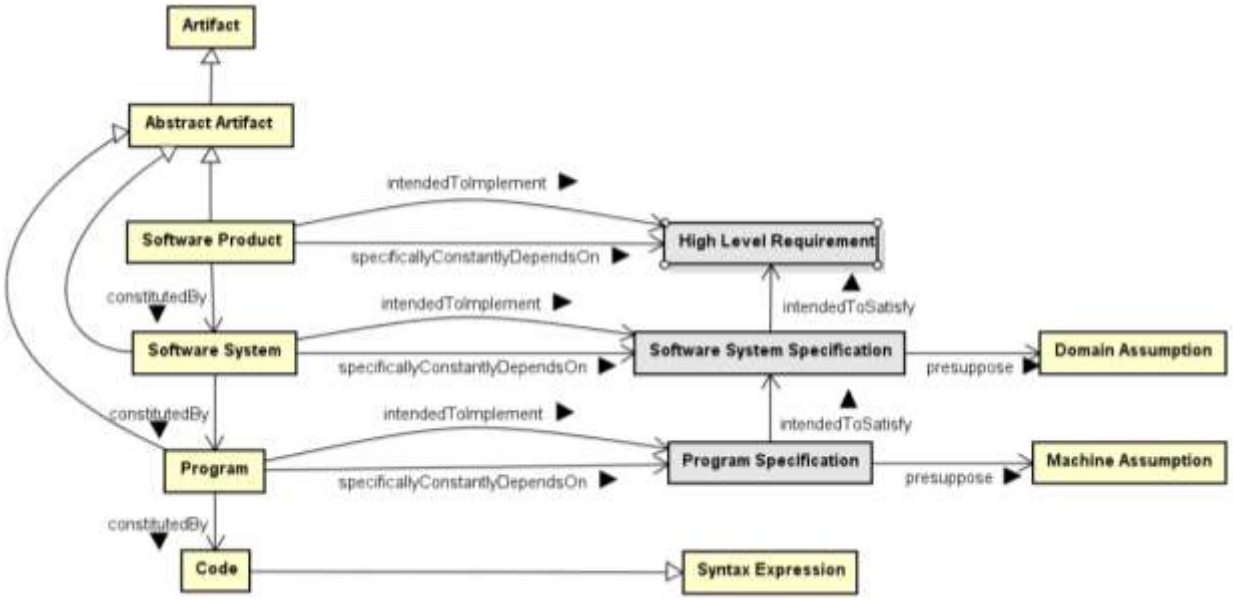

Figure 7 -Relation between the requirements and software [1]

The **ontology** is described based on the components (Figure 1) and overall system architecture, which define the data flow and the processes associated with the integrated CounteR platform. The model is achieved through the conception of classes, UML diagrams and process flow diagrams.

The **taxonomy** is depicted from wireframes and scenarios, with the mention that some of the naming and classification is yet to be discussed, further investigated, and possibly enhanced.

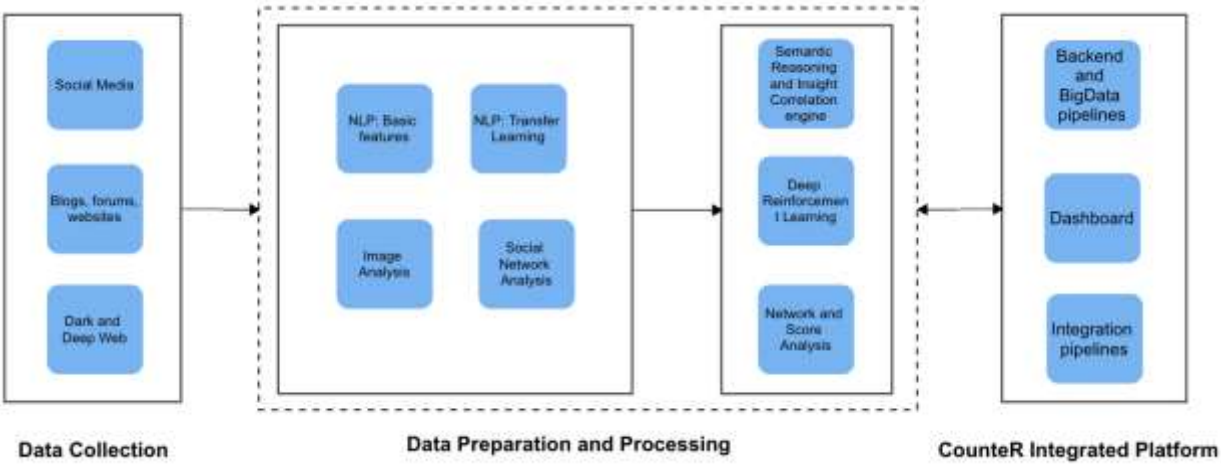

Figure 8 – CounteR Components relation and diagram

3 Taxonomy

The taxonomy of the CounteR software prototype will provide the definition of concepts (entities and processes), the relationship between them and a way faof classification.

3.1 Concept definition: Entities

The below tables describe the entities envisaged to be implemented in the CounteR platform.

Table 2 – Entities concept definition

No	Entity	Definition
1.	LEA Investigator	The LEA Investigator is the representative from the LEA organisation who will use the CounteR platform. In the case of end-users other than LEAs this will be renamed as Investigator.
2.	LEA Group	The LEA Groups can be used by the LEAs to organize the Investigators in departments, allowing the assignment of different sets of permission per each group. In the case of end-users other than LEAs this will be renamed as Group.
3.	LEA Permission	The Permissions consists in a set of rules defined by the LEA organisation in order to restrict or permit different actions and access to specific content from the CounteR platform.
4.	Detection Task	The Detection Task represents the starting point in setting up the CounteR platform to detect specific malicious content.
5.	Detection Alert	The Detection Alert represents a signal which is sent to the LEA Investigators when the CounteR platform detects radical content which is classified as a potential threat.
6.	Detection Result	The Detection Result represents the output of a Detection Task.
7.	Target User	The Target User represents the author of the detected malicious content. By default, this entity is pseudonymised.
8.	Vulnerable User	The Vulnerable User represents the community members potentially affected by the Targeted Users.
9.	Community	The Community represents a group of target users and vulnerable users detected by the SNA module.
10.	Malicious Content	The content that is market as radicalised by the NLP and Deep Learning models.
11.	Post	The collected content from blogs and social media can be categorised as posts.
12.	Article	The collected content from social media or open, dark and deep web can be categorized as articles.
13.	Comment	The collected content from social media or open, dark and deep web can be categorised as comment.
14.	Page	The collected content from open, dark and deep web can be categorised as pages.

15.	Report	The results of a task can be exported in a PDF report.
16.	Text Content	The text content collected from analysis platforms by the CounteR platform with the help of detection tasks.
17.	Multimedia Content	The images, video and audio content collected from analysis platforms by the CounteR platform with the help of detection tasks.
18.	Data Sources / Analysis Platforms	The data sources targeted to be supported by the CounteR platform. For example, Twitter.
19.	Observables	The Observables are the filters that can be used to restrict the search for a specific analysis platform. For example, a specific hashtag in Twitter can be seen as an observable.
20.	Threat Classification	The Threat Classification is represented by threat categories in order of the risk involved.
21.	Radicalisation domains	The radicalization domains consist in the list of targeted areas of radicalisation selected to be researched and implemented in the CounteR project. For example, religious radicalisation.
22.	Keywords	The keywords represent a predefined list of syntagms commonly used in different domains of radicalisation.
23.	Language	The languages targeted by the CounteR platform for the data collection and detection components.
24.	Risk categorisation	The risk categories to be used in describing other entities such as targeted users.
25.	Logs	The logs are representing database records consisting of the details regarding the actions done by LEA Investigators inside the CounteR Platform.
26.	Proactive Mode	The Proactive mode aims to accurately detect radicalisation and terror content and terrorist online communities.
27.	Reactive Mode	The Reactive mode aims to provide accurate content removal (with a human-in-the-loop approach) and automatic prevention of re-upload.
28.	Use cases	A use case is an entity that can be used to organize tasks. For example, a use case can be an event “EU Summit 2023 Paris”, and in this use case, the LEA Investigators can add tasks.

3.2 Concept definition: Processes

The below table described the CounteR platform main processes, excluding the obvious CRUD (Create, Read, Update, Delete) processes for each entity listed in the previous chapter 2.1. The processes are also mapped into the Figure 2 as a process flow diagram.

Table 3 – Processes concept definition

No	Process	Definition
1.	Authentication	The process that enables LEA Investigators to securely access the CounteR platform.
2.	Data collection	The process employed by the CounteR platform to gather data from the targeted data sources. This can be split into Social Media Collector, Web Collector and Dark Web Collector as defined in D1.2 Technical Requirements.
3.	Alerts generation	The process which analyses the malicious content and generates alerts based on LEA's configuration.
4.	Malicious content detection	The process responsible for the analysis of the collected data and classification into malicious content based on the detection task parameters, such as the domain of radicalisation, keywords, language, data source, etc. This process can be split in Image Analysis, NLP Analysis, Social Network Analysis (SNA), Semantic Reasoning and Insight Correlation Egine (SRICE) and the Deep Reinforcement Learning Process (DRL).
5.	Target users' generation	This process enables LEA Investigators to extract target users from the malicious content detected by the CounteR platform. Related processes are the mapping of communities and vulnerable users.
6.	Flag to be deleted process	This is the process to use used by LEA Investigators to remove the content published by target users on third parties' platforms.
7.	Data Pseudonymisation	This is the data privacy process that will be used by the data collection process to pseudonymise the data right after the gathering from data sources. Pseudonymisation allows the continued identification and linkage to one or more datasets without directly identifying the individual [3]. LEA Investigators with special permissions might be able to retrieve the original content for a specific malicious content item, such as a post, article, comment, or a page.
8.	Publish and Unpublish content process	For the Reactive mode, the CounteR platform will enable end-users to automatically unpublish malicious content.

The following high-level process flow diagram (Figure 9) describes how the entities (Table 2) are interacting with each other and the relationships with the main processes (Table 3) conceptualized in the CounteR platform.

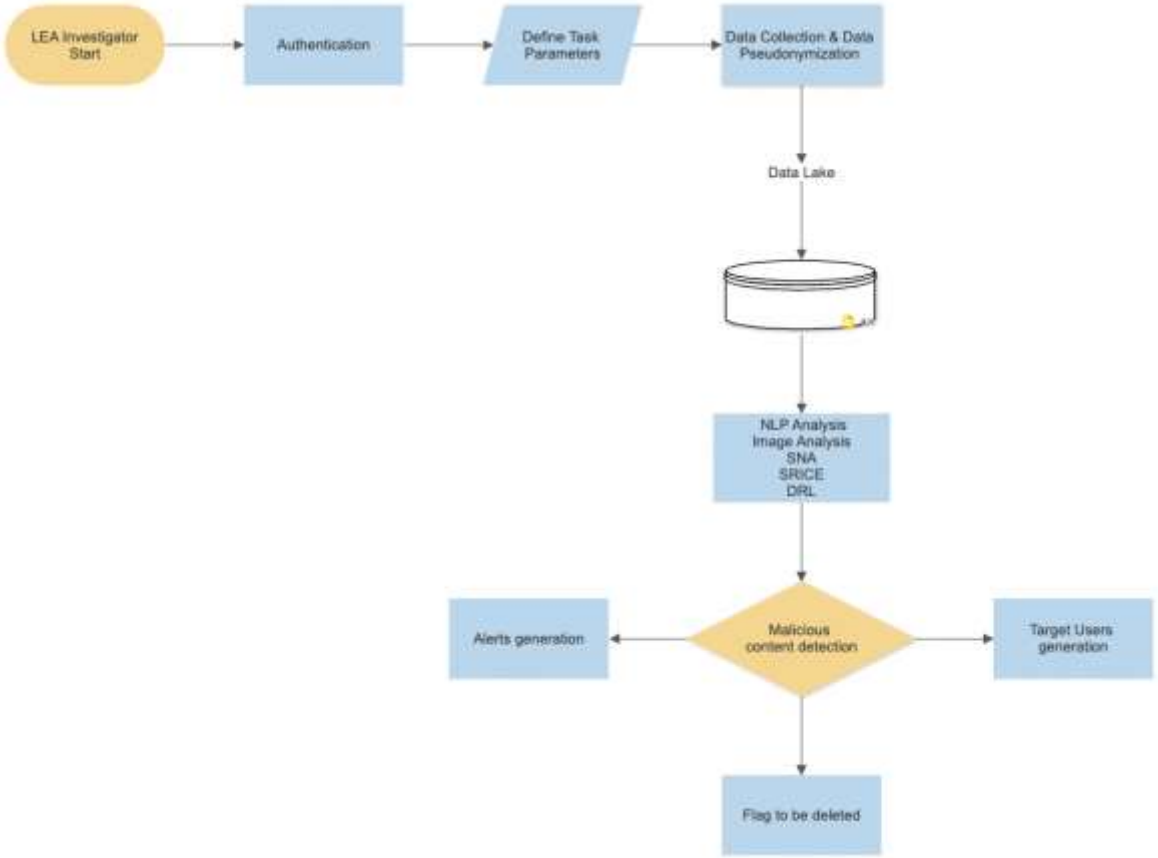

Figure 9 – Process flow diagram

4 Ontology model

The ontology aims at setting a common conceptualization on the radicalisation domain, including entities and the processes established to be implemented in the CounteR project. As a core ontology, Figure 8 provides the general concepts for software processes and the relationship with the entities that will be conceptualized in the overall schema of the CounteR platform.

UML, the Unified Modeling Language, is the most utilized language to the requirements specification. UML is a standardized modeling graphical language that includes an integrated set of diagrams.

A concept that grouped multiple objects that have the same features and share the same behaviors is commonly known as a class in UML. The UML class diagram is a conceptual model which is used for designing the logic model of information system by grouping multiple objects that have the same features and share the same behaviors is commonly known [4].

Class is interpreted as a set of objects. Classes are elements of a terminological knowledge representation, which is known as a class model in UML and an ontology in OWL. Classes can be defined as subclasses of other classes, as equivalent to another class (*subclass of*), for the definition of a class hierarchy.

Domain is a built-in property that links a property to a class description. The domain of a property specifies the set of objects that can be related to other values with the property.

Range is a built-in property that links a property to either a data range or a class description. The range of a property specifies the set of objects or data values that can be related to other objects with the property.

Data property values are defined by using the name of the property and relating it to an object or data value. The range of the data property is datatype. The datatype can be string, int and other datatypes.

The *inverseOf* construct can be used to define an inverse relation between object properties. Properties are unidirectional, that is, their direction goes from domain to range. To state that a certain property is an inverse of another property the *inverseOf* relation can be used.

4.1 Class diagrams

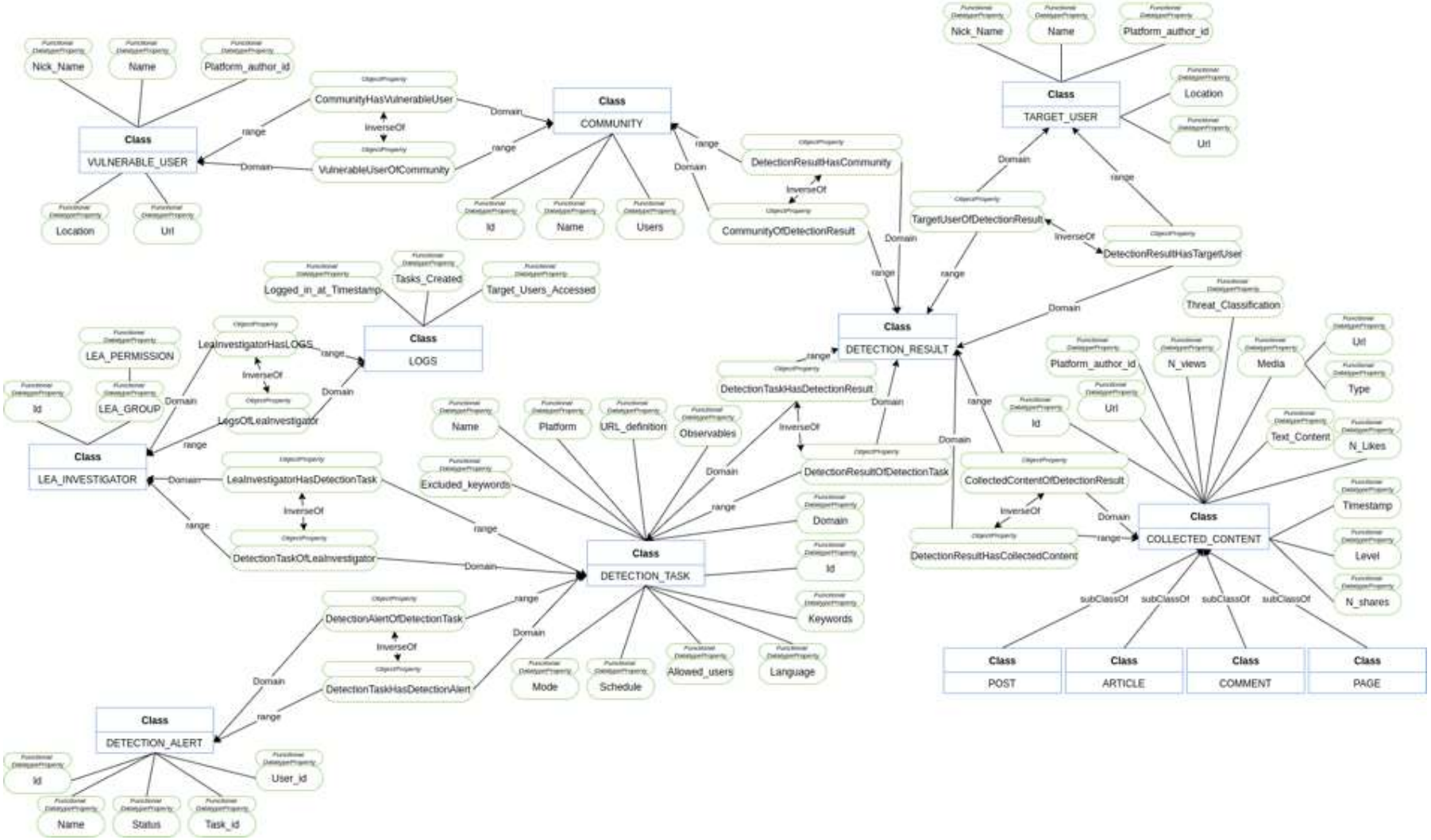

Figure 10 - UML class diagram

4.2 Entities detailed relations

The entities descried in Table 7 are detailed in the below chapter. The database schema of the CounteR project will be built using SQL database relationships “one to many 0..*”, “many to many”, “one to one”.

4.2.1 LEA_INVESTIGATOR

Relations:

LEA_INVESTIGATOR(0..) creates (1..1) DETECTION_TASK*

LEA_INVESTIGATOR(0..) HAS LOGS(1..1)*

LEA_INVESTIGATOR(1..1) part of (1..) LEA_GROUP*

4.2.2 LEA_GROUP

Relations:

LEA_INVESTIGATOR(1..1) part of (1..) LEA_GROUP*

LEA_GROUP(1..1) has (0..) LEA_PERMISSION*

4.2.3 LEA_PERMISSION

Relations:

LEA_GROUP(1..1) has (0..) LEA_PERMISSION*

4.2.4 DETECTION_TASK

Relations:

DETECTION_TASK (0..) creates (1..1) DETECTION_ALERT*

DETECTION_TASK (0..) has (1..*) DETECTION_RESULT*

LEA_INVESTIGATOR(0..) creates (1..1) DETECTION_TASK*

4.2.5 DETECTION_ALERT

Relations:

DETECTION_TASK (0..) creates (1..1) DETECTION_ALERT*

4.2.6 DETECTION_RESULT

Relations*:*

DETECTION_TASK (0..) has (1..*) DETECTION_RESULT*

DETECTION_RESULT (1..1) has (1..) COLLECTED_CONTENT*

TARGET_USER (0..) author of (1..1) DETECTION_RESULT*

DETECTION_RESULT (1..1) author of (0..) COMMUNITY*

4.2.7 TARGET_USER

Relations*:*

TARGET_USER (0..) author of (1..1) DETECTION_RESULT*

4.2.8 VULNERABLE_USER

Relations:

COMMUNITY(2..) HAS (0..*) VULNERABLE_USER*

4.2.9 COMMUNITY

Relations:

DETECTION_RESULT (1..1) author of (0..) COMMUNITY*

COMMUNITY(2..) HAS (0..*) VULNERABLE_USER*

4.2.10 COLLECTED_CONTENT

Relations:

DETECTION_RESULT (1..1) has (1..) COLLECTED_CONTENT*

POST (0..) instance of (1..1) COLLECTED_CONTENT*

ARTICLE (0..) instance of (1..1) COLLECTED_CONTENT*

COMMENT (0..) instance of (1..1) COLLECTED_CONTENT*

PAGE (0..) instance of (1..1) COLLECTED_CONTENT*

4.2.11 LOGS

Relations:

LEA_INVESTIGATOR(0..) HAS LOGS(1..1)*

4.2.12 POST

Relations:

POST (0..) instance of (1..1) COLLECTED_CONTENT*

4.2.13 ARTICLE

Relations:

ARTICLE (0..) instance of (1..1) COLLECTED_CONTENT*

4.2.14 COMMENT

Relations:

COMMENT (0..) instance of (1..1) COLLECTED_CONTENT*

4.2.15 PAGE

Relations:

PAGE (0..) instance of (1..1) COLLECTED_CONTENT*

4.2.16 PLATFORM

Relations:

PLATFORM(0..) part of (0..*) DETECTION_TASK*

4.2.17 OBSERVABLE

Relations:

OBSERVABLE(0..) part of (0..*) DETECTION_TASK*

4.2.18 KEYWORD

Relations:

KEYWORD(0..) part of (0..*) DETECTION_TASK*

4.2.19 LANGUAGE

Relations:

LANGUAGE(0..) part of (1..*) DETECTION_TASK*

4.3 UML Class Diagram

The *Figure 11* represents the high-level visual representation of the database schema with the main tables, properties and the relationships between the tables. The process flow diagram described in the figure above (*Figure 11 – Process flow of diagram*) is a type of interaction diagram where it illustrates how and in what order a set of objects works together. It concentrates on determining the behavioral view of a system.

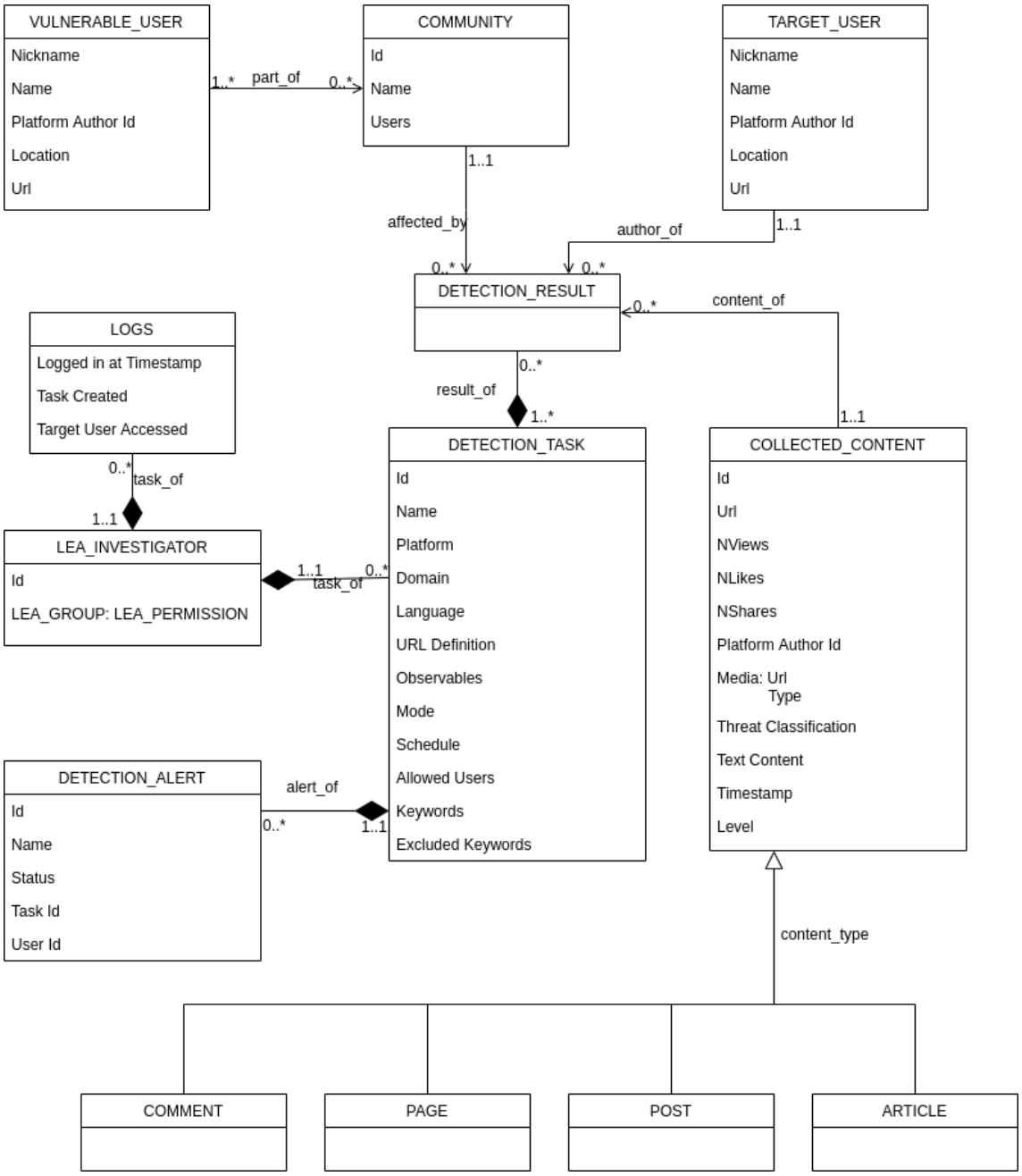

Figure 11– Process flow diagram

The class description for a set of objects has the same structure, behavior, and relationships. The class can have attributes that determine the structure, and the operation determines the behavior of the instance of the class.

The attribute in UML class diagram must be unique in the context of the class. Attributes may have different levels of usage, and they are described as being accessible from other classes.

4.3.1 Behavioral view

The *LEA_INVESTIGATOR* creates one or more tasks of *DETECTION_TASK,* task that is logged as a *LOG* for further investigation and reuse.

The task results into a *DETECTION_RESULT* which analyses the content of *COLLECTED_CONTENT* that content types like *COMMENT, PAGE, POST, ARTICLE.*

A specific *COMMUNITY* that consists of multiple *VULNERABLE_USER* that are part of *COMMUNITY* are affected by *DETECTION_RESULT* that has a *TARGET_USER* who is the author of *DETECTION_RESULT.*

If the *DETECTION_RESULT* has a positive result, then a *DETECTION_ALERT* is created, so *LEA_INVETIGATOR* can act accordingly based on it.

5 Conclusions

The deliverable *D1.5 Ontology and taxonomy model* is a result of the technical analysis done in the previous deliverables of work package 1, the *D1.1 LEA requirements*, use cases and scenarios and the *D1.2 Technical Requirements*. It connects with WP2 - *Social and Psychological Factors in Radicalization Process* as the inception phase completes and moves to the elaboration and construction phases within the technical work packages WP3 to WP6. Towards the final of the project the transition phase will complete the ontology with the help of the WP8 - End-user Training & Knowledge Empowerment & Pilots as defined by [1].

The taxonomy was extracted from wireframes and scenarios, with the note that it will evolve once the projects advance to the next phases. The ontology was described based on the system architecture and scenarios. The model is achieved through the conception of classes, UML diagrams and process flow diagrams.

Having a good understanding of the entities and the processes architecture allowed us to map the classes and relations between the concepts. However, based on the output of work package 2 and specifically the results of Task 2.4 Radicalisation Taxonomy, Ontology and Recommendations, to highlight possible risk alerts of radicalisation and propose measures that help to reduce the vulnerabilities offline and on-line ecosystems to radicalisation, several concepts of the current taxonomy might be changed.

Overall, defining the entities and the processes enables our technical teams to start working at the structure of the components preparing for the first deliverable of WP6, and more specifically: *D6.1 System release v1* with the due date in month 12.

6 Bibliography

[1] X. WANG and G. N. GUARINO, Towards an Ontology of Software: a Requirements Engineering Perspective, Federal University of Espírito Santo (UFES), 2014.

[2] A. D. Nicola, M. Missikoff and R. Navigli, A software engineering approach to ontology building, Istituto di Analisi dei Sistemi ed Informatica, Consiglio Nazionale delle Ricerche, Viale Manzoni, 30-00185 Roma, Italy, 2008.

[3] B. Lubarsky, RE-IDENTIFICATION OF "ANONYMIZED DATA", 2017.

[4] J. O. Grady, System Requirements Analysis, JOG System Engineering, San Diego, CA, USA , 2014.

[5] d. A. F. Ricardo, SABiO: Systematic Approach for Building Ontologies, Federal University of Espírito Santo, 2013.