The Identification and Creation of Ontologies for the Use in Law Enforcement AI Solutions – MAGNETO Platform Use Case

Rafal Kozik[1,4], Michal Choras[1,4], Marek Pawlicki[1,4(✉)], Witold Hołubowicz[4], Dirk Pallmer[2], Wilmuth Mueller[2], Ernst-Josef Behmer[2], Ioannis Loumiotis[3], Konstantinos Demestichas[3], Roxana Horincar[5], Claire Laudy[5], and David Faure[5]

[1] ITTI Sp. z o.o., Poznan, Poland
mpawlicki@itti.com.pl
[2] Fraunhofer IOSB, Karlsruhe, Germany
[3] Institute of Communication and Computer Systems (ICCS), Athens, Greece
[4] University of Science and Technology, UTP, Bydgoszcz, Poland
[5] Thales Research & Technology, Palaiseau, France

Abstract. Every single day more and more organizations face the challenge of finding a way to support their conduct with data. The flooding amounts of data currently available vastly outweigh human capabilities, thus Big Data processing becomes a pressing issue. This problem is especially prevailing for Law Enforcement Agencies (LEAs), where massive amounts of critical data are collected from heterogenous sources, often by various entities in different countries. Ontologies have been developed into a predominant technique for establishing semantic interoperability among heterogeneous systems which transact information. In this paper we propose the Magneto ontology – a solution built as a crucial part of the Magneto project. It has been developed on top of well-established ontologies dealing with people, events and security incidents, bearing in mind the heterogenous nature of the myriad of data sources as the starting point. Examples of the building blocks, a classification of the sources of data, an overview of the application in a specific use scenario and a discussion on the future use of the ontology will be given.

Keywords: Ontology · Artificial intelligence · Common Representational Model · Semantic interoperability · Correlation

1 Introduction

Nowadays, Law Enforcement Agencies (LEAs) - similarly to other entities in different domains operating based on the mass volume of digitalised data - have to face problems related to big data processing, knowledge understanding and interoperability. It is clear that the common understanding of data between cooperating LEAs (or between LEAs and external parties), as well as flexible tools for information sharing and exchange are vital for effective and successful law enforcement and prosecution.

N. T. Nguyen et al. (Eds.): ICCCI 2019, LNAI 11684, pp. 335–345, 2019.
https://doi.org/10.1007/978-3-030-28374-2_29

There is a need to develop a common representational model for data used by LEAs. Therefore, in order to facilitate the law enforcement data representation, we propose our solution – so called MAGNETO ontology. The work presented in this paper is related to the first stage of collaborative research project MAGNETO (Technologies for prevention, investigation, and mitigation in the context of the fight against crime and terrorism), co-funded by the European Commission within Horizon 2020 programme. The main ambition of the project and the MAGNETO consortium is to empower LEAs (Law Enforcement Agencies) with the capability to process, manage, analyse, correlate and reason from large datasets characterized by heterogeneity. In particular, the technical goals defined by the Consortium are: development of solutions enabling the exploration of data from various sources, their indexing, enrichment through meta-information and contextualization, development of tools supporting semantic information fusion and inference based on processed data and development of a human-machine interface (HMI) enriching the situational awareness and operational capabilities of LEAs.

The goal of this paper is to present a flexible and sophisticated representational model for the data utilized by LEAs. The model is supposed to establish a common baseline for cooperation and information exchange for interoperable security systems.

The paper is structured as follows: Sect. 2 presents the current state of the approaches used to delineate and model the security-related knowledge, Sect. 3 addresses classification of the MAGNETO data sources and provides preliminary MAGNETO data model, while Sect. 4 discusses the future use of our ontology and Sect. 5 provides overall summary of our work.

2 Related Work

A multitude of models exist as an expression and representation of both semantic information and knowledge. Natural language, propositional logic, first-order logic, mathematical models, relational model, semantics nets or rules constitute merely several examples. Both in computer science and philosophy, ontologies share the attempt of representing concepts in order to model a specific domain. The representational primitives that are used are typically entities, ideas and events existing in the real world, with all their interdependent properties and relations, similarities and dissimilarities.

Ontologies are considered to be one of the pillars of the Semantic Web that provides standards to identify entities (*URIs*), express facts (*RDF*), express concepts (*RDFS*), share vocabularies, describe constraints (*OWL*), query knowledge (*SPARQL*), link data and publish data (*RDFa*). Ontologies have become a common approach when it comes to the task of ensuring semantic interoperability between heterogeneous systems exchanging information. Ontologies provide an abstraction layer in the Semantic Web to enable dialog and service negotiation between participating systems, ensuring that the participants have the same concept of the information exchanged.

In [1] the published Security Incident Ontology (SIO) is presented. This ontology was built to describe security relevant events on the campus of Ryerson University in Toronto. The light-weight ontology was built upon existing ontologies describing

events, geolocation and timelines, people and their relations, The Event Ontology [2] was developed at Centre for Digital Music in Queen Mary University of London in October, 2004. This ontology is centred around the notion of an event, seen here as the way by which cognitive agents classify arbitrary time/space regions. This ontology has been proven to be suitable in a wide range of contexts, due to its simplicity and usability (i.e. conference talks, concerts, festivals). The ontology reuses Timeline [3] ontology for temporal predicate "event:time" and Geo RDF vocabulary [4] for spatial predicate "event:place". FOAF (Friend of A Friend) [5] is a lightweight ontology modelling persons, their activities and their relations to other persons and objects. It is a widely accepted vocabulary for representing Social Networks. Princeton WordNet (WN) is a lexical database for the English language [6, 7]. The SIO ontology was used to develop an infrastructure transforming textual notification of security incidents to a machine-readable representation. The SIO ontology specification and its Turtle version were made available by registering the dataset to Linked Open Data and adding it to the Linked Open Data Cloud.

The scope of this ontology is limited to security incidents at the university campus and handles only a small section that has yet to be covered by MAGNETO. Also based on the Event Ontology, a Forensic Complex Event Ontology has been developed for the analysis of video material at the Multimedia Vision Research Group of Queen Mary University of London. The ontology framework is a derivative of DOLCE foundational ontology and it is designed to represent events that forensic analysts commonly encounter in the investigation of criminal activities [8]. In the CAPER project, a conceptual structure of the cross-border organized crime has been developed by analysing the work of EUROPOL and its databases, resulting in a Europol Organized Crime Structure (OCS), which is a supranational structure embedding the specific natural structures, overcoming limitations due to non-harmonized criminal law systems [9]. The European LEA Interoperability Ontology (ELIO) aims at ensuring interoperability between the LEAs. ELIO is a lightweight ontology based on the taxonomy of OCS, implementing the concepts as classes and adding object properties for connecting the classes as "hasTechnique", "hasEssentialCondition", "hasCrime" and "hasCountry" [10]. Compared to the SIO ontology, the ELIO crimes concept is missing the embedded event concept of SIO, as a result the crimes are not assigned to agents, geo-location and time. The ELIO and MCO ontologies are not publicly available.

The Global Justice Extensible Markup Language (GJXDM) Data Model is an XML standard designed specifically for criminal justice information exchanges, providing law enforcement, public safety agencies, prosecutors, public defenders, and the judicial branch with a tool to effectively share data and information in a timely manner [11]. The Global Justice XML Data Model was the result of an effort by the justice and public safety communities to produce a set of common, well-defined data elements to be used for data transmissions. In 2005, the successor project National Information Exchange Model (NIEM) started, which was based on GJXDM, broadening the scope to include other federal and state agencies [12].

From our perspective, it is always important to use ontology in practice for reasoning and decisions, so that it can be termed as applied ontology [13].

3 MAGNETO Ontology

3.1 MAGNETO Data Sources Classification

The starting point for the MAGNETO Common representational model specification were the data sources. It is crucial from the perspective of the platform to comprehend the variety of possible data formats that will ingest the analyses. The data is to be sourced from a diverse spectrum of supported formats including structured documents (database exports, structured reports, lists, standardized descriptions), free (unstructured) text, pictures, audio sequences, video sequences, raw data. Those files will be drawn from an assortment of sources, including external LEA's, Police databases, surveillance systems/human sensors, Internet/OSIF (Open Source Information) content, telecon data, other databases.

The files will be drawn from an assortment of sources, including: External LEA's, Police Databases, Surveillance Systems/Human Sensors, standardized descriptions, the Internet/OSIF (Open Source Information) Content, Telecon data, other databases.

The Source Data is to only be stored in its initial location, thus protecting its integrity and assuring efficiency of use. Annotations based on the original data are more fitting to be served. The high-level overview of data source types has been listed in Fig. 1.

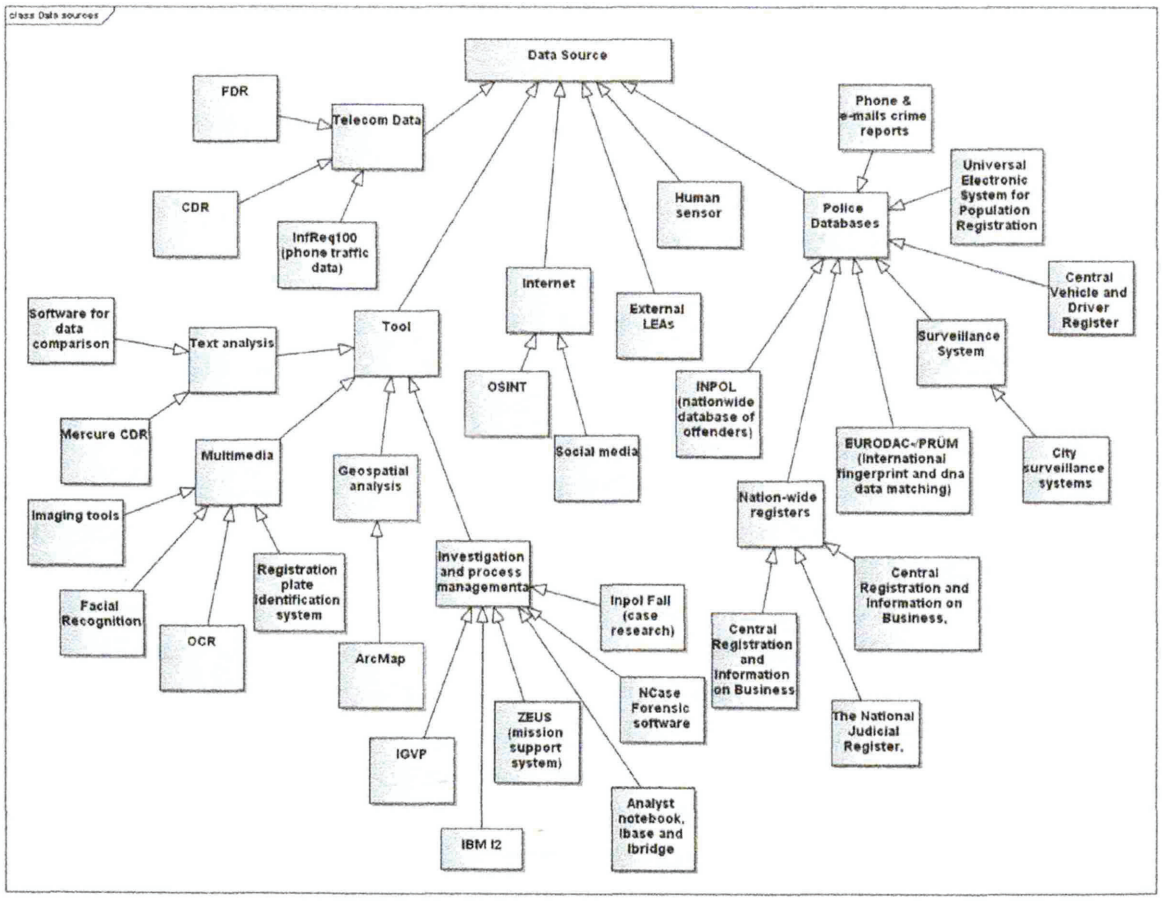

Fig. 1. Data sources identified for MAGNETO project use cases (arrows indicate "generalization" relation).

3.2 Building Blocks of CRM

The Common Representational Model (CRM) will be developed on top of several ontologies, taxonomies and classifications that will facilitate computational and data mining functionalities provided by various MAGNETO components.

Moreover, to bridge the gap between different taxonomies used within the project we propose to adopt several publicly available ontologies providing a common vocabulary and facilitating interoperability.

While developing the model we have decided to build its basis on top of several ontologies identified during the desktop research, namely: SIO (Security Incident Ontology), FOAF (Friend of a Friend) Vocabulary Specification and event ontology.

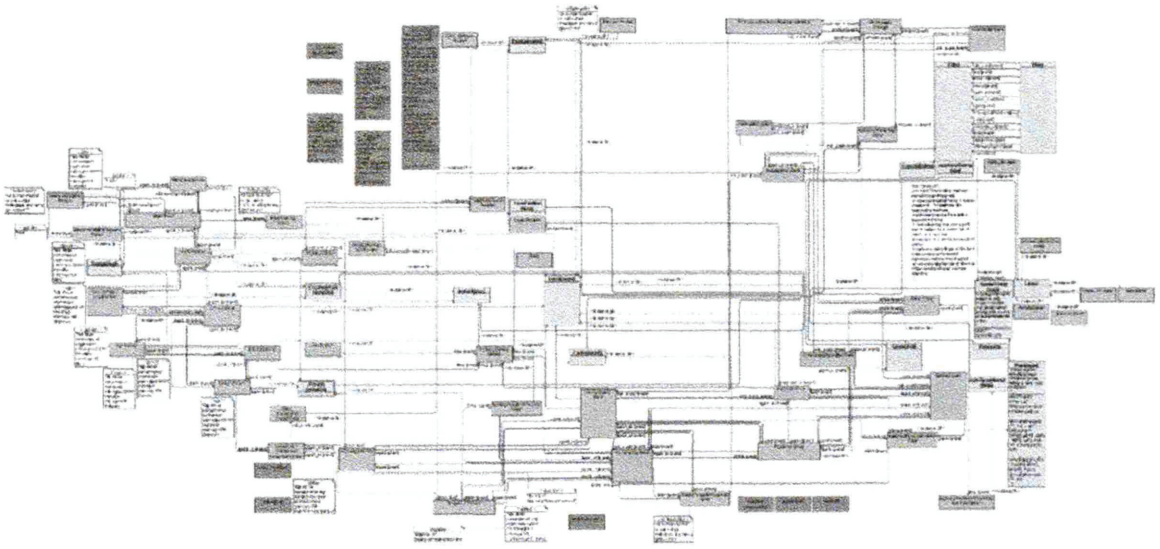

Fig. 2. A general overview of the preliminary version of CRM – intentionally left unreadable to present the scale of the model. It is not possible to fit the whole CRM on a printed page, but legible parts of it are presented in Figs. 6 and 7.

The Event Ontology shown in Fig. 3 describes events with a time and geo-location reference, involved agents/resources and a product generated by the event.

In the centre of the FOAF ("Friend of a Friend")-Ontology (see Fig. 4) is the concept of a Person, that is modelling the static attributes of a person and its relations to other persons, organisations and groups, and their activities in social networks.

The SIO Ontology (see Fig. 5) refines the event concept taken over from the Event Ontology by defining SecurityIncidents, and models persons as victims or subject of SecurityIncidents.

3.3 Validation of the Preliminary Version of the Ontology

To prove the applicability of the approach and test run selected ontologies an example of a use scenario was modelled. The graph (Fig. 2) depicts a situation involving an array of real-life situations and illustrates the relationships between certain agents and events. The graph itself might seem overly intricate, but considering the vast amounts

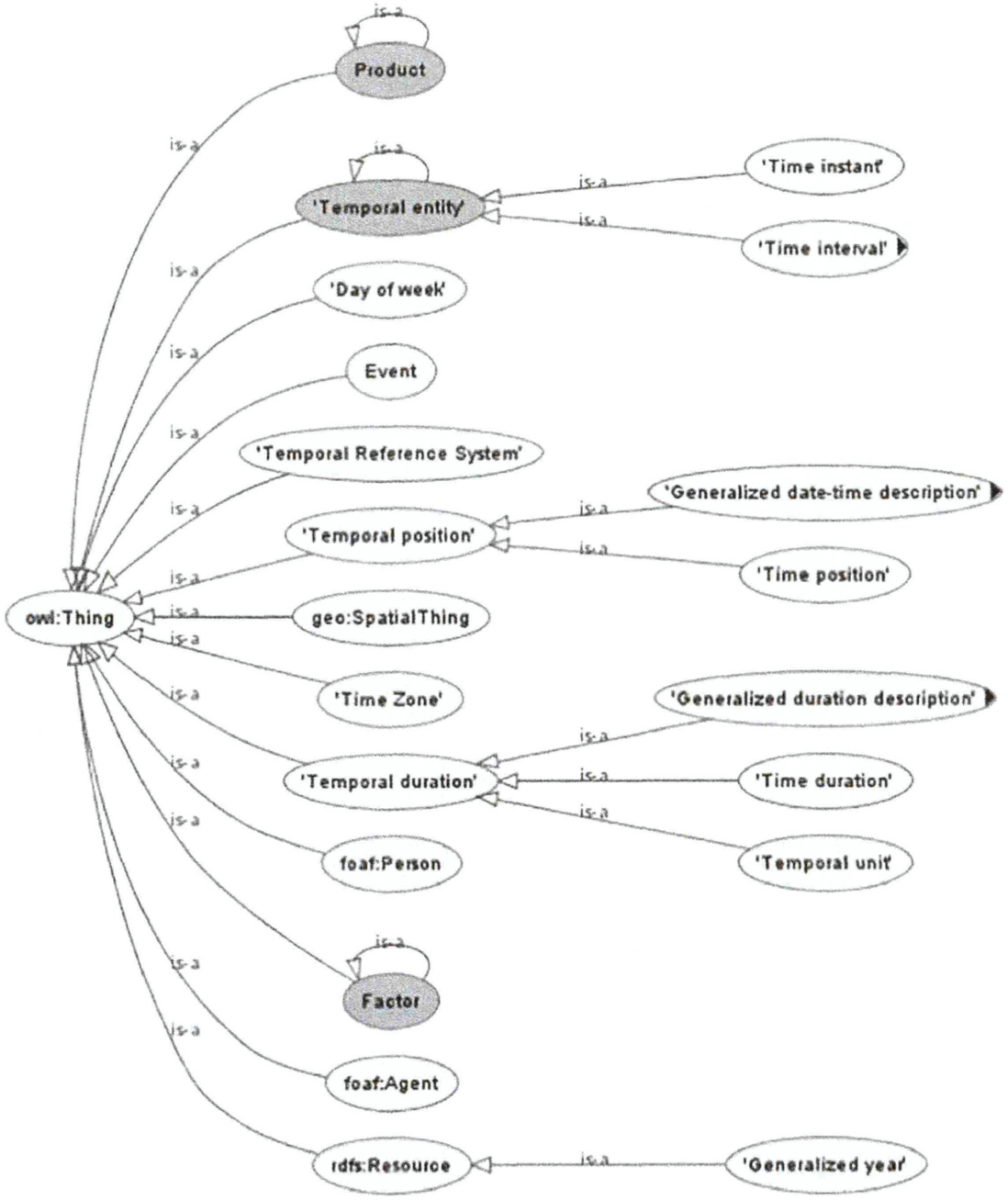

Fig. 3. General abstract classes of the Event Ontology

of information, evidence, detail and events LEA has gone through to finalize the case a certain degree of sophistication seems inevitable.

In order to validate the proposed preliminary version of the ontology, we have selected one of the use case scenarios. Specifically, we intended to examine to what extent it is capable of describing the context, key elements, and relations.

In Fig. 6 one can observe how the Investigation Event has multiple sub-events, namely the *SocialMedia* event, the *Interrogation* event and the *HouseSurvey* event.

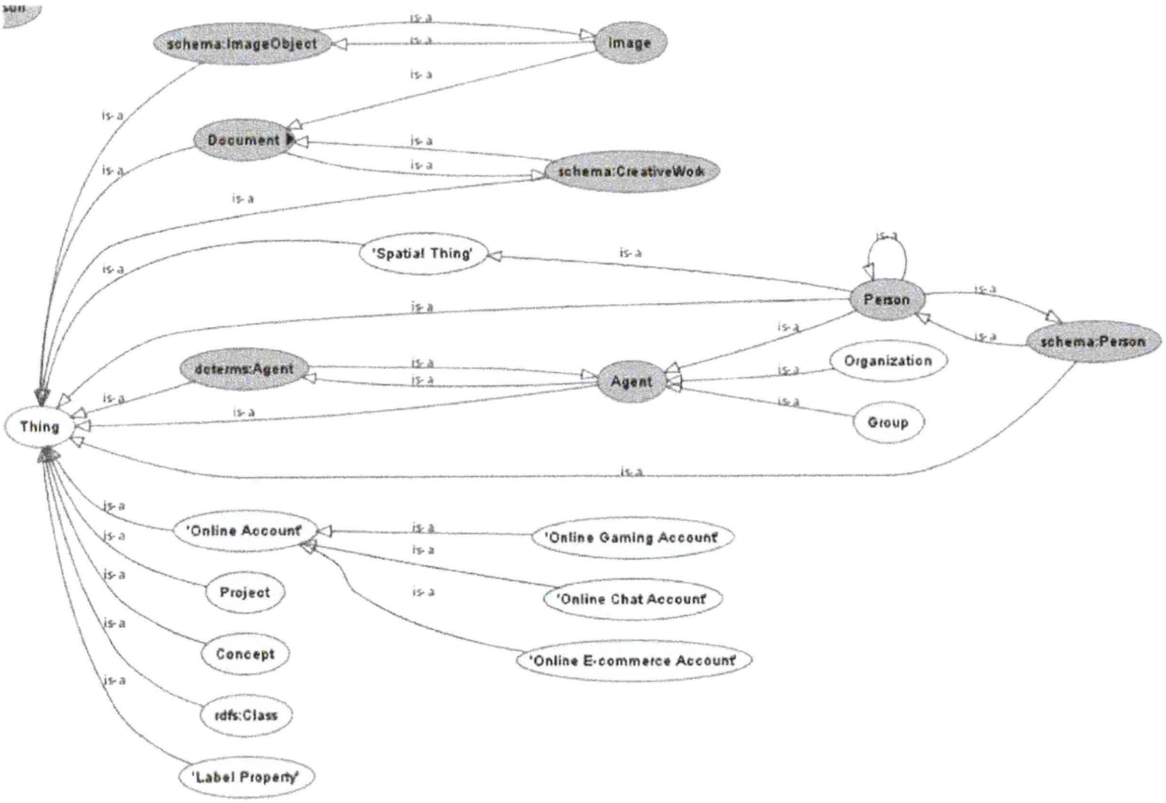

Fig. 4. General abstract classes of the FOAF Ontology

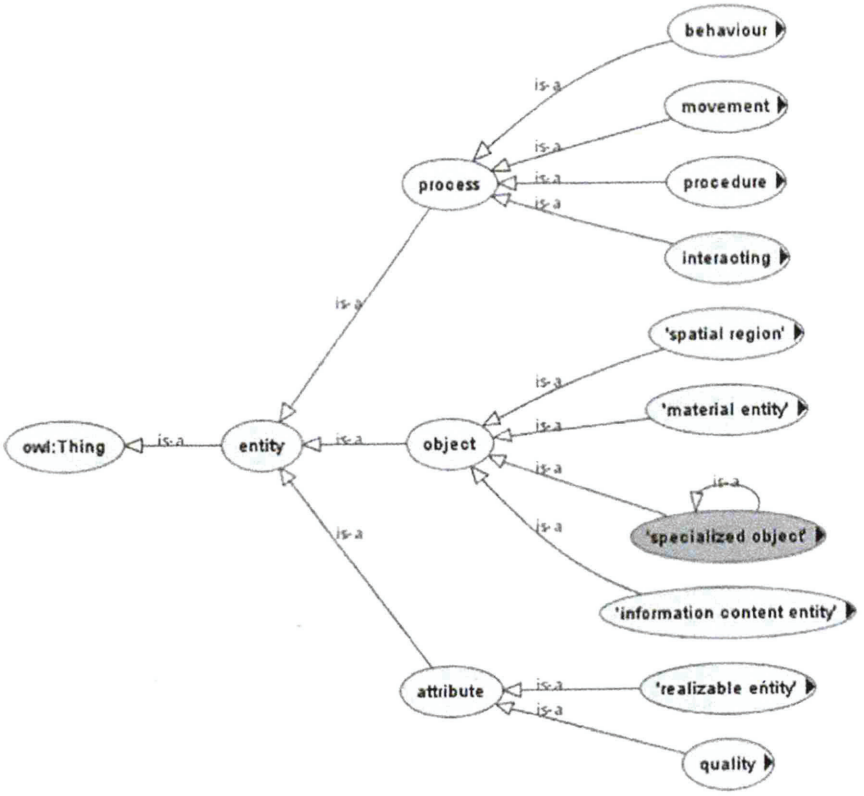

Fig. 5. General abstract classes of the SIO Ontology

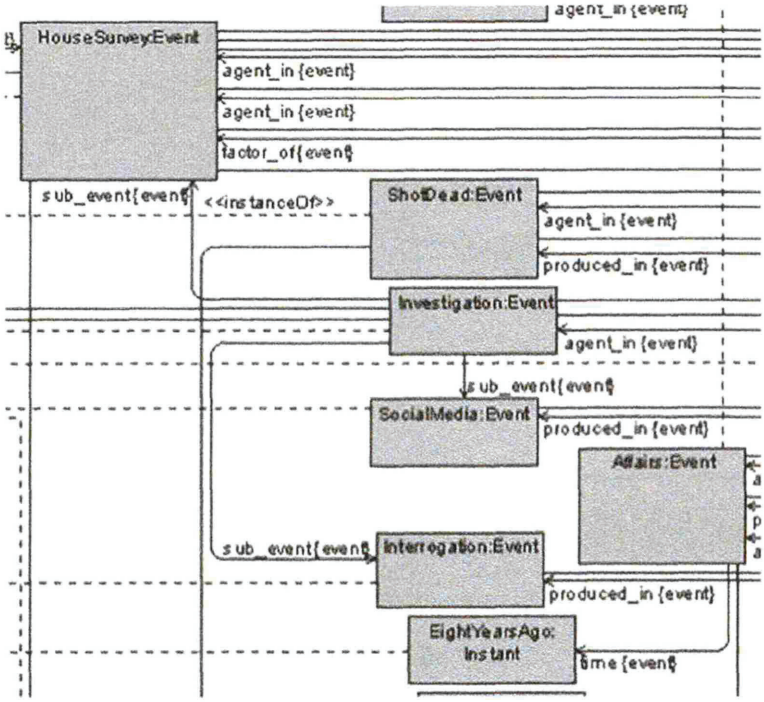

Fig. 6. The investigation event

Figure 7 demonstrates how specific individuals, like *Video_Evidence* or *HeadShotWound* are instances of classes, in this case the *Product* class from the event ontology, or how *Residents* are an instance of the *Group* class from the FriendOfAFriend ontology.

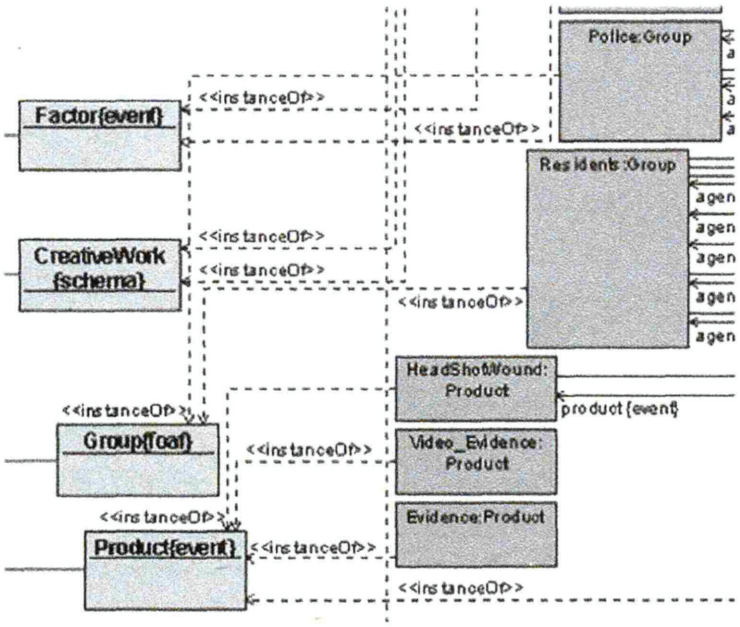

Fig. 7. Instantiated classes.

4 Discussion on the Future Use in Practice

The final product of the MAGNETO project will be a platform integrating all components, services and capabilities developed during the project (and mentioned in the Introduction). To enable communication, the exchange and fusion of information coming from different modules, the MAGNETO platform architecture, defined interfaces and the representational model that was described in the previous section need to be developed. In addition to the ability of representing knowledge and describing security events in a commonly understandable way, the goal of the MAGNETO representational model will be to support both anticipation and prediction of future trends (e.g. security threats) and to establish the common ground for reasoning and cognition to achieve situational awareness.

In addition, MAGNETO has as an ambition to monitor, study and contribute to the activities related to the standards of ISO/TC 292 "Security and Resilience" in the area of security aimed at enhancing the safety and resilience of society and to eventually propose the defined common representational model for standardization.

The ISO 22311 standard, originally prepared by ISO/TC 223 "Societal security" and later taken over by ISO/TC 391 "Societal and Citizen Security", was conceived for societal security purposes and specifies a common output file format that can be extracted from the video-surveillance content to allow different investigators to access digital video-surveillance content and perform its necessary processing.

Since the common data format introduced by the ISO 22311 standard is specific to the video content, it cannot be directly translated to a more general context of the MAGNETO project. Nevertheless, we mention the general principles that were followed during the standardization process that could be adapted to MAGNETO.

A list of minimum technical requirements has been established in order to assure the interoperability of different video-surveillance systems:

- All collected information should be referenced to Coordinated Universal Time (UTC).
- The data format should allow the file export of time slices of data coming from different sources and preserve the time correlation between the contents, whatever export process (removable media or data trans-mission) is used.
- The format should enable compatible, comparable processing of files exported by different systems (covering the same scene) with a common time base.
- The format should facilitate widely available, independent Operating Systems (OS) to allow for minimal processing, ensuring any combination of the following: video data and metadata display, direct access to the metadata without display of the videos, selection of content time slots or access to the sources defined by name or scene-location.

5 Conclusions

The preceding paper aimed to delineate the importance of creating an ontology to provide the semantic context to both the AI and the multitude of entities operating on similar data in the domain of law enforcement. The process of developing a common representational model growing out of several pre-existing ontologies, bridging the gaps between them, and building on top of them has been emphasized.

The main contribution of this document is the presentation and explanation of the proposed MAGNETO common representational model. A thorough analysis of the existing ontologies has been performed, an examination based both on expert knowledge and previous projects. The MAGNETO common representational model is thus created as an ontology, with one of the use cases modelled as a validation of the concept.

Acknowledgement. This work has been performed under the H2020 786629 project MAGNETO, which has received funding from the European Union's Horizon 2020 Programme. This paper reflects only the authors' view, and the European Commission is not liable to any use that may be made of the information contained therein.

References

1. Fani, H., Bagheri, E.: An Ontology for Describing Security Events (2015). https://doi.org/10.18293/seke2015-101
2. Raimond, Y., Abdallah, S.: "The event ontology," Centre for Digital Music, Queen Mary, University of London. http://motools.sourceforge.net/event/event.html
3. Raimond, Y., Abdallah, S.: "The Timeline Ontology," Centre for Digital Music, Queen Mary, University of London. http://motools.sourceforge.net/timeline/timeline.html
4. W3C Semantic Web Interest Group, "Basic Geo (WGS84 lat/long) Vocabulary," W3C Semantic Web Interest Group. http://www.w3.org/2003/01/geo/
5. FOAF Vocabulary Specification 0.99, Namespace Document, 14 January 2014. http://xmlns.com/foaf/spec/
6. Fellbaum, C.: WordNet: An Electronic Lexical Database. MIT Press (1998). http://wordnet.princeton.edu/
7. W3C WordNet RDF/OWL Files. https://www.w3.org/2006/03/wn/wn20/#wnjune06
8. Sobhani, F., Izquierdo, E., Piatrik, T.: Ontology-based forensic event detection using inference rules. Published in 2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC) (2017)
9. CAPER WebSite: http://www.fp7-caper.eu/
10. González-Conejero, J., Varela Figueroa, R., Muñoz-Gomez, J., Teodoro, E.: Organized crime structure modelling for european law enforcement agencies interoperability through ontologies. In: Casanovas, P., Pagallo, U., Palmirani, M., Sartor, G. (eds.) AICOL-2013. LNCS (LNAI), vol. 8929, pp. 217–231. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-45960-7_16
11. Justice Information Sharing - Global Justice XML (Archive). https://it.ojp.gov/initiatives/gjxdm

12. NIEM - National Information Exchange Model, the official NIEM web site. https://www.niem.gov/
13. Choraś, M., Kozik, R., Flizikowski, A., Hołubowicz, W.: Ontology applied in decision support system for critical infrastructures protection. In: García-Pedrajas, N., Herrera, F., Fyfe, C., Benítez, J.M., Ali, M. (eds.) IEA/AIE 2010. LNCS (LNAI), vol. 6096, pp. 671–680. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-13022-9_67