# Leveraging Semantic Web Technologies for Analysis of Crime in Social Science

Luca Pulina[1], Antonietta Mazzette[1], Laura Pandolfo[2], Elena Piga[1], Maria Laura Ruiu[1], and Camillo Tidore[1]

[1] POLCOMING, Università di Sassari, Viale Mancini n. 5 – 07100 Sassari – Italy
{lpulina,mazzette,mlruiu,tidore}@uniss.it
[2] DIBRIS, Università di Genova, Via Opera Pia, 13 – 16145 Genova – Italy
laura.pandolfo@edu.unige.it

**Abstract.** In this paper we present the conceptual level of an ontology-based application aimed to support social scientists in their sociological analysis related on crime. Starting from several concrete issues posed by the research team of the *Social Observatory on Crime* of the University of Sassari, our goal is to build a Semantic Web based tool to collect, organize, and analyze data on crime, as well as for the exploitation of research results by institutions and civil society.

## 1 Introduction

Generally speaking, crime analysis is the activity aimed at finding trends in crimes in order to devise both policies and solutions to crime-related issues. From a sociological point of view, the analysis of crime is addressed to understand how criminal phenomena might influence social assets in the context of urban and rural areas. Moreover, it aims to identify prevention measures and their potential effects on social configurations. In order to do that, data analysis plays a role of paramount importance. Nowadays, information on urban crime is profitably used by a growing number of local governments to identify major risks and make decisions about the safety of their community.

Computer-assisted qualitative data analysis software (CAQDAS) were successfully adopted by the sociological research community in order to both organize and analyze such kind of data – see, e.g., [1]. A number of benefits of CAQDAS have been recognized by sociological literature. On the other hand, several issues are still open in relation to theoretical, methodological issues – see, e.g., [2] – and practical aspects, such as heterogeneity of data sources, data integration, and the exploitation of implicit knowledge related to the collected data. It is well-established that the usage of Semantic Web (SW) technologies can provide a valuable support in order to overcome the practical limits listed above. Moreover, the usage of such technologies in crime and public safety fields is not new – see, e.g., [3].

In this paper we present the conceptual level of an ontology-based application aimed at supporting social scientists in their sociological analysis related to

crime. It represents the first step towards the development of a tool aimed at organize and manage both quantitative and qualitative data related to this application domain. In particular, the need of such a tool has emerged from several concrete issues posed by the research team of the *Social Observatory on Crime* (OSC)[3] of the University of Sassari. On the one hand, the creation of a SW-based crime information platform might allow stakeholders – institutions and civil society – to easily access to data on crime, for instance by mapping crimes and crime-related issues, and identifying where and how they are occurring, where they are concentrated and why. On the other hand, it can facilitate researchers' work in organizing and managing data collected from different sources, such as statistical data and qualitative information from newspapers.

The remainder of the paper is organized as follows. In Section 2 we describe the OSC, while in Section 3 we report the whole process of design and implementation of the presented ontology. We conclude the paper in Section 4 with some final remarks and discussing future works.

## 2 The Social Observatory on Crime

The OSC originated in 2012 thanks to an interdisciplinary team (in particular social scientists such as sociologists, psychologists, economists, jurists) which involves researchers from the University of Sassari. It originated from the Urban Study Center (CSU)[4], that focuses on urban and territory evolution and transformations; coordinates empirical study and promotes the culture of legality by involving students in actively doing research. The CSU contributes to promoting the adoption of governance approaches by involving private and public bodies as both producers and beneficiaries of research outcomes. It also aims to disseminate activities' results through seminars, conferences, educative courses and scientific publications. Since 2004, the CSU has started to focus on crime and insecurity in order to observe their impacts on the Sardinian social context and territory. The analysis and monitoring activities are based on data collected from documents provided by justice officers, newspapers and national statistical reports (e.g., reports of the Italian National Institute of Statistics, ISTAT).

The OSC was built aimed at promoting and generating governance approaches by trying to involve different kinds of stakeholder, such as private and public bodies. In fact, the inclusion of these actors was supposed to be relevant both to collect basic information and data, and to define concerted strategies for reducing criminal behaviors and attitudes. For a long time, both literature and policies have focused on situational crime prevention strategies by creating "defensible spaces", and less focused on the contrast of the motivations that encourage deviance. During the last twenty years governance approaches to crime have been promoted, through collaboration among a multiplicity of actors (also "external" to the control/protection functions). Following this approach, the CSU concentrated its efforts on identifying processes for increasing

[3] *Osservatorio Sociale sulla Criminalità*, `http://polcoming.uniss.it/node/1133`
[4] *Centro Studi Urbani*, `http://www.centrostudiurbani.it`

degree of key-stakeholders' participation and networking rather than defining specific "interventions" and "architectonic fences". Recently [4], OSC identified social indicators in order to create an "Informative System for data collection and analysis". This tool was thought to support policy planning and decision-making oriented at fighting and reducing criminal and illegal activities.

The main goal of OSC is to develop an exhaustive database which includes quantitative data (primary and secondary data from different sources, e.g., ISTAT and regional prosecutors) and qualitative information obtained, e.g., by analyzing local newspapers (in particular *La Nuova Sardegna*[5] and *L'Unione Sarda*[6]). The newspaper consultation has allowed the collection a number of detailed information (e.g., description of places where murders happened, description of authors way of life and their past experiences, connections with other types of crime, etc.) that otherwise would have been difficult to gain and record. However, the consultation refers to those relevant crimes which get newspapers attention such as murders, robberies, attacks, threats, and cannabis cultivation.

A further objective of OSC is to analyze connections between widespread insecurity and crime. In fact, individual and collective behaviors, decision-making and economic activities are often strongly related to the types and the intensity of these connections. Moreover, criminal phenomena should be analyzed in relation to the process of modernization (and its consequences) that has involved the targeted territories by shaping social configuration of urban and rural areas.

## 3 The OCRA Ontology

OCRA aims at being the conceptual layer of a Semantic Web based tool focused on the improvement of the processes related to the collection, organization, management, and analysis of data on criminal phenomena in Sardinia by OSC. In the following, we describe design and implementation of OCRA (Ontology for CRime Analysis). We can summarize as follows the main steps of this process:

1. Definition of the domain.
2. Identification of the key concepts of the domain to be described.
3. Identification of the proper language and Tbox implementation.
4. Ontology population, i.e., filling the Abox with known facts.

Firstly, we reviewed the semi-structured dataset collected by OSC during a 11 years-long research on criminality in Sardinia. Data on criminal phenomena were collected through specific forms to be filled with information obtained, e.g., by local newspapers such as *La Nuova Sardegna* and *L'Unione Sarda*. The criminal phenomena recorded are murder, attacks, robbery and cultivation of cannabis. Data collection forms were mainly composed of the following information:

– Data concerning the newspaper, e.g., name of the newspaper, date, and title of the related article.

[5] `http://lanuovasardegna.it`
[6] `http://unionesarda.it`

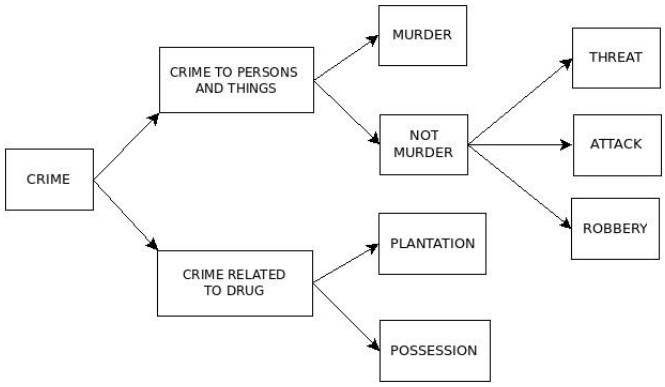


**Fig. 1.** Raw classification of the criminal phenomena.

- Data concerning the crime, e.g., type, place, date, and motive.
- Data concerning the authors of the crime and victims, e.g., name, job, age, and records of criminal offenses.

Regarding the second point, we analyzed collected data in order to highlight common terminology, redundancies, and relationships between different elements, as suggested by the domain experts of the OSC. The results of this process enabled us to compute a taxonomy – depicted in Figure 1 – related to different crimes.

Considering the third point, we proceeded with the choice of the modeling language analyzing the different alternatives offered by OWL 2. To retain most of the practical advantages of OWL 2, but to improve on its applicability, in [5] has been introduced OWL 2 profiles, i.e., a sub-language of OWL 2 featuring limitations on the available language constructs and their usage.

Considering the available profiles, we excluded OWL 2 EL because it does not support inverse object properties, while we discarded both OWL 2 QL and OWL 2 RL because they do not support, e.g., existential quantification to individuals. Thus, OCRA has been developed in OWL 2 DL, and its Tbox is composed of 81 classes, 36 object properties, and 53 data properties. In the following, we describe main classes of the OCRA ontology[7]:

`ArticoloGiornale` *(NewspaperArticle)* represents the class containing the newspapers information about the specific crime. Every instance of this class has data properties such as `Titolo` *(Title)* and `DataArticolo` *(ArticleDate)*.

`Luogo` *(Place)* includes the place where the crime has occurred. Individuals of `Luogo` are also the places in which the victims and offenders were born or live.

[7] The full documentation is available at `http://visionlab.uniss.it/OCRA`.

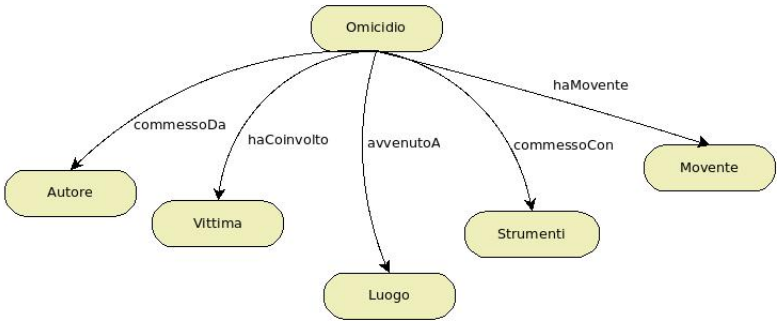


**Fig. 2.** Class `Omicidio` and related classes in the OCRA ontology.

`Movente` *(Motive)* aims to model the motive of the crime. It has different sub-classes, each of which is a specific motive, such as economical, political, revenge, etc.

`PersonaFisica` *(Person)* models people related to a specific crime. It has two sub-classes, namely `Vittima` *(Victim)* and `Autore` *(Offender)*. Every individual belonging to those classes has data properties such as `Età` *(Age)*, `Sesso` *(Gender)*, `StatoCivile` *(MaritalStatus)*, `Precedenti` *(RecordsOfCriminalOffenses)*.

`Reato` *(Crime)* is one of the central classes of OCRA. It has two sub-classes related to the main types of offenses taken into account: crimes involving people or things and crimes related to drug – see below. `Reato` has different data properties such as `DataReato`*(CrimeDate)*, `NumeroVittime`*(NumberOfVictims)*, `NumeroAutori` *(NumberOfOffenders)*.

`ReatoAPersoneECose` *(CrimeToPersonsAndThings)* In this class are included individuals related to crimes that caused material damage or injure people. `ReatoAPersoneECose` has two sub-classes: `Omicidio`*(Murder)*, which includes crimes with homicide, and `NonOmicidio`*(NotMurder)*. The latter covers a large series of crimes which have not led to murder. `NonOmicidio` has three sub-classes, each one modeling different category offenses, namely `Attentato`*(Threat)*, `Minaccia`*(Attack)*, `Rapina`*(Robbery)*.

`ReatoCollegatoAllaDroga` *(CrimeRelatedToDrug)* is the other principal sub-class of `Reato` and is related to all the drug offences. In particular, we modeled the following two sub-classes of drug offences: `Coltivazione`*(Plantation)* and `Detenzione`*(Possession)*. Some of the most relevant data properties of `ReatoCollegatoAllaDroga` are connected to the type and the number of drugs confiscated by the authorities, such as `Semi` *(Seeds)* and `Piante`*(Plants)*.

`Strumenti` *(Weapons)* represents the class containing the instruments used by an offender to commit the crime. It has various sub-classes, such as `ArmiDaFuoco` *(FireArms)*, `Esplosivi` *(Explosives)* and `Veicoli` *(Vehicles)*.

Concerning object properties, we briefly describe the ones related to `Omicidio`, because they enable domain experts to involve in their analysis important

data regarding places in which the crime has occurred, offenders, and victims. Noticeable object properties are:

- `commessoDa` *(CommittedBy)*: connects `Omicidio` to `Autore`.
- `haCoinvolto` *(hasInvolvedIn)*: relationship between `Omicidio` and `Vittima`.
- `avvenutoA` *(takesPlaceIn)*: allows the identification of murder's place (`Luogo`).
- `commessoCon` *(hasWeapon)*: relationship with the murder weapon.
- `haMovente` *(hasMotive)*: it connects the offense with the motive (`Movente`).

In Figure 2 we show a graphical example of these relationships.

Finally, the OCRA Abox has been populated using data provided by the OSC. Actually, the Abox contains more than 15000 individuals, with their related properties, while the whole ontology is composed of about 365000 triples.

## 4 Conclusions

In this paper we described design and development of the OCRA ontology, the conceptual level of the ontology-based application aimed to support people of OSC in their sociological analysis related to crime.

Currently, we are developing a data integration layer in order to exploit information coming from relevant external sources, e.g., open data provided by ISTAT and DBpedia. We are also designing a Graphical User Interface to support the ontology population stage, in order to make this process of knowledge acquisition more interactive and dynamic. More, concerning the ontology population, we are studying automated solutions for data collection and insertion.

Finally, we are planning to perform more detailed experimental analysis on the OCRA ontologies. Some preliminary experiments have shown us that OCRA could be a challenging benchmark for OWL 2 DL reasoners.

*Acknowledgments* The authors wish to thank the anonymous reviewers for their valuable suggestions, which were helpful in improving the final version of the paper.

## References

1. Mangabeira, W.C.: Caqdas and its diffusion across four countries: National specificities and common themes. Current Sociology **44**(3) (1996) 191–205
2. John, W.S., Johnson, P.: The pros and cons of data analysis software for qualitative research. Journal of Nursing Scholarship **32**(4) (2000) 393–397
3. Asaro, C., Biasiotti, M.A., Guidotti, P., Papini, M., Sagri, M.T., Tiscornia, D.: A domain ontology: Italian crime ontology. In: Proceedings of the ICAIL 2003 Workshop on Legal Ontologies & Web based legal information management. (2003)
4. Mazzette, A., ed.: La criminalità in Sardegna, Quarto rapporto di ricerca. EDES, Sassari (2014)
5. Motik, B., Patel-Schneider, P., Parsia, B., Bock, C., Fokoue, A., Haase, P., Hoekstra, R., Horrocks, I., Ruttenberg, A., Sattler, U., et al.: OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax. W3C Recommendation **27** (2009)