# TOWARDS A GENERIC FRAUD ONTOLOGY IN E-GOVERNMENT

Panos Alexopoulos, Kostas Kafentzis
*IMC Research, Fokidos 47, Athens, Greece*

Xanthi Benetou, Tassos Tagaris
*Institute of Communication and Computer Systems, National Technical University of Athens, Athens, Greece*

Panos Georgolios
*IMC Research, Fokidos 47, Athens, Greece*

Keywords: Ontologies, e-Government, Fraud Detection.

Abstract: Fraud detection and prevention systems are based on various technological paradigms but the two prevailing approaches are rule-based reasoning and data mining. In this paper we claim that ontologies, an increasingly popular and widely accepted knowledge representation paradigm, can help both of these approaches be more efficient as far as fraud detection is concerned and we introduce a methodology for building domain specific fraud ontologies in the e-government domain. The main characteristic of this methodology is a generic fraud ontology that serves as a common ontological basis on which the various domain specific fraud ontologies can be built. The methodology along with the generic fraud ontology consist a powerful conceptual tool through which knowledge engineers can easily adapt ontology-based fraud detection systems to virtually any e-government domain.

## 1 INTRODUCTION

Fraud is an issue with psychological, economic and legal ramifications for both the public and private sector spanning geographic regions. The last EHFCN (European Healthcare Fraud and Corruption Network – http://www.efhcn.org) conference produced agreement among members on a common definition of fraud: "Civil fraud is the use or presentation of false, incorrect or incomplete statements and/or documents, or the non-disclosure of information in violation of a legally enforceable obligation to disclose, having as its effect the misappropriation or wrongful retention of funds or property of others, or their misuse of purposes other than those specified".

Other definitions of fraud present it as a type of corrupt conduct and risk for organizations which cannot be eliminated. In broader terms fraud is the deliberate and premeditated act perpetrated to achieve gain on false ground. The effects of fraud are economic (reduced operational effectiveness), legal (depriving resources from rightful claimants) and psychological (damage moral and reduce confidence in government).

The consequences of e-government fraud are numerous. For example, in the healthcare domain fraud causes the raise of the cost of health care benefits for everybody. According to the Deputy Health Minister of Scotland Lewis Macdonald (http://www.scotland.gov.uk) the potential losses to healthcare across Europe from fraud and corruption are estimated to be at least 30 billion euros each year and may be as high as £100 billion. For most employers, fraud increases the cost of providing benefits to their employees and, therefore, their overall cost of doing business. That translates into higher premiums and out-of-pocket expenses as well as reduced benefits or coverage. Healthcare fraud, can also impact the quality of the received care. When dishonest providers put greed ahead of care,



Alexopoulos P., Kafentzis K., Benetou X., Tagaris T. and Georgolios P. (2007).
TOWARDS A GENERIC FRAUD ONTOLOGY IN E-GOVERNMENT.
In *Proceedings of the Second International Conference on e-Business*, pages 269-276
DOI: 10.5220/0002112602690276
Copyright © SciTePress

proper diagnosis and treatment may be ignored and patients may be put at risk solely to generate higher dollar claims.

For all these reasons, a number of fraud-fighting organizations, consortia and networks have been created. Such a network is the European Healthcare Fraud and Corruption Network (EHFCN) which coordinates and advances work to counter healthcare fraud and corruption across Europe. The different approaches EHFCN adopts for fighting fraud are common between the various e-government domains and include:

- The creation of anti-fraud and anti-corruption culture among service providers, healthcare suppliers, healthcare payers, healthcare users and ultimately among citizens.
- The use of all possible presentational and publicity opportunities to act as a deterrent to those who are minded to engage in e-government fraud or corruption
- The use of effective prevention systems so that when fraudulent or corrupt activities are attempted, they will fail.
- The professional investigation of all cases of detected or alleged fraud and corruption.
- The imposition, where fraud and corruption is proven, of appropriate sanctions – namely civil, criminal and/or disciplinary processes. Multiple sanctions should be used where possible;
- The seeking of financial redress in respect of resources lost to fraud and corruption and the return of recovered resources to the area of patient care or services for which they were intended;
- The development of a European common standard of risk measurement (baseline figures), with annual statistically valid follow up exercises to measure progress in reducing losses to fraud and corruption throughout the EU.
- The use of detection systems that will promptly identify occurrences of healthcare fraud and corruption

Our interest towards fraud detection lies into the technological aspect of fraud fighting and in particular in the area of fraud detection systems. In this area organizations and agencies seek multiple layers of fraud detection methods and tools ranging from rule-based systems (Belhadji and Dionne 1997) to predictive modelling (Zukerman and Albrecht 2000) approaches. We believe that in all these methods and approaches, ontologies can play a significant role as they have a lot to offer in terms of interoperability, expressivity and reasoning.

In this paper we intend to illustrate a methodology for building domain specific fraud ontologies that are to be used by various ontology-based fraud detection systems. This methodology is accompanied and supported by a generic fraud ontology which acts as a reference framework and a basis for building such specialized ontologies.

The rest of the paper is organized as follows. The next section discusses the way ontologies can be used for detecting fraud. Section 3 illustrates our proposed methodology for building domain specific fraud ontologies while section 4 provides an analytical description of the structure and architecture of the generic fraud ontology that we propose. Finally, section 5 highlights the applicability of our methodology and fraud ontology to specific case studies that cover a wide range of e-government domains and section 6 summarizes our approach.

# 2 ONTOLOGY BASED FRAUD DETECTION IN THE E-GOVERNMENT DOMAIN

## 2.1 Technological Approaches in the Fraud Detection Domain

In general, the IT fraud detection systems in the e-government domain fall into two main categories: those that detect fraudulent activities the minute these take place and those that identify fraud by discovering suspicious behavioural patterns within batches of data. The first are usually based on rules and prediction models while the latter utilize data mining techniques (Hand et al, 2001). Rules practically contain already known fraud patterns and identify fraudulent activities through comparison to these patterns.

Similarly, in predictive modelling, historical data is used to build profiles of fraudulent behaviour in order to detect future occurrences of the same behaviour based on the similarity to the existing profiles.

However, rule-based systems and predictive modelling can only defend against known (or predicted) fraud types. Data mining systems, on the other hand, utilize large datasets in order to discover unknown patterns of suspicious or fraudulent behaviour. Those systems are used in conjunction with large data warehouses that store information relevant to the fraud detection domain. Additionally, data mining systems provide the foundation of predictive modelling. As data mining reveals anomalous behaviour patterns, those cases are

investigated in greater detail and from those that are found to be fraudulent, new fraud profiles are built.

## 2.2 The Importance of Ontologies

Ontologies can play a vital role in both the rule-based and data mining fraud detection approaches. Apart from the rules, a really important component of a rule-based system is its knowledge base. An important issue in knowledge bases is the knowledge representation paradigm they adopt as the latter influences the type and quality of reasoning that can be made within the knowledge-based system. In the Knowledge Representation literature there can be found a number of different knowledge representation schemas and languages including first-order logic (Hodges, 2001), defeasible logic (Nute, 1994), modal logic (Blackburn et al, 2003) etc.

A family of these languages are Description logics (DL) (Baader et al, 2003) on which in turn ontologies are based. Ontologies are knowledge models that represent a domain and are used to reason about the objects in that domain and the relations between them (Gruber 1993). Thus, a knowledge base may use an ontology to specify its structure (entity types and relationships) and its classification scheme. In such a case, the ontology, together with a set of instances of its classes constitutes the knowledge base.

The use of ontologies and ontology-related technologies for building knowledge bases for rule-based systems is considered quite beneficial for two main reasons:

- Ontologies provide an excellent way of capturing and representing domain knowledge, mainly due to their expressive power.
- A number of well established methodologies, languages and tools (Gomez-Perez et al 2004) developed in the Ontological Engineering area can make the building of the knowledge base easier, more accurate and more efficient, especially in the knowledge acquisition stage which is usually a bottleneck in the whole ontology development process.

Ontologies are also very important to the data mining area as they can be used to select the best data mining method for a new data set (Tadepalli et al 2004). When new data is described in terms of the ontology, one can look for a data set which is most similar to the new one and for which the best data mining method is known, this method is then applied to the new data set. In this way, there is no need for trying out every known method on the new data set, but the one (or few) that is most promising can be directly selected.

## 2.3 The Importance of Existing Ontologies and Standards

Creating a knowledge model for a given domain from scratch is most of the times a very difficult and time/resource consuming task especially as far as the knowledge acquisition process is concerned. Therefore, in any such effort, the existence of already established and commonly accepted standards, classification schemes and ontologies regarding this domain should always be taken in mind. Of course the degree of existence and reusability of such standards depends largely on the given domain.

For example, in the healthcare domain, existing medical classifications, terminologies and taxonomies, which we used for the TSAY case study that we describe in section 5, include the International Classification of Diseases ICD) (http://www.who.int/ classifications/icd), the ATC system (http://www.whocc.no/atcddd) and the SNOMED CT system (http://www.snomed.org). The ICD classification is an international standard diagnostic classification for all general epidemiological and many health management purposes. The Anatomical Therapeutic Chemical (ATC) system is a system for classification of medicinal products according to their primary constituent and to the organ or system on which they act and their chemical, pharmacological and therapeutic properties. Finally, SNOMED (Systematized Nomenclature of Medicine) is a system of standardized medical terminology developed by the College of American Pathologists (CAP).

Apart from such domain specific classifications like ATC or SNOMED, attempts for building fraud ontologies for certain domains and fraud types have also been made. Examples include financial fraud (Leary et al, 2003) and e-mail based fraud (Kerremans et al, 2005).

# 3 METHODOLOGY FOR BUILDING FRAUD ONTOLOGIES

The methodology we propose for building fraud detection ontologies is based on the suggestion that fraud is actually an operational risk for an

organization and as such it should be treated through a risk management process. Risk management (RM) (Crockford, 1986) (Lam, 2003) is the process whereby public organizations may methodically address the risk associated to their activities with the goal of achieving a sustained benefit within each activity and across their portfolio of activities. The focus of RM is to identify, measure and treat these risks in order to reduce their probability of happening.

In a similar fashion, our methodology defines a process for identifying, measuring and treating fraud in the context of e-government services. This process comprises three steps; a) establishment of the fraud context, b) identification of fraud within this context and c) transformation of this information into an ontological model.

Establishment of the fraud context within an organization involves defining the type of fraud the organization wishes to fight and identifying the business processes fraud occurs upon. This is done through a business process modelling procedure which records the fraud susceptible business processes of the organization and their context. On the other hand, fraud identification involves the description of potential fraud cases that could occur within the organization and of corresponding detection methods. This identification is done in two ways, namely by acquiring organizational knowledge regarding fraud from experts and by utilizing data mining methods in order to extract unknown fraud patterns.

The final step of the methodology involves transforming the knowledge derived from the two previous steps into an ontology so that it can be utilized by fraud detection systems. This step usually requires following some formal knowledge engineering procedure.

Obviously, these three steps should be repeated for each different domain or case study meaning that the proposed methodology is an iterative procedure. In order to minimize the effort required in each iteration we created a generic fraud ontology which acts as the basis for building domain specific fraud ontologies.

# 4 FRAUD ONTOLOGY

The fraud ontology is practically a generic framework for defining domain and case specific fraud ontologies which are to be used in ontology-based fraud detection systems. Among others, this framework should be easily adaptable and extendible to different domains and types of fraud. This was made possible through a multi-layer architectural design of the fraud ontology which makes the latter adaptable, extendible and to a significant degree reusable.

## 4.1 Fraud Ontology Layered Architecture

The overall architecture of the fraud ontology consists of three independent but interconnected layers each one defining its own set of ontologies (see Fig. 1).

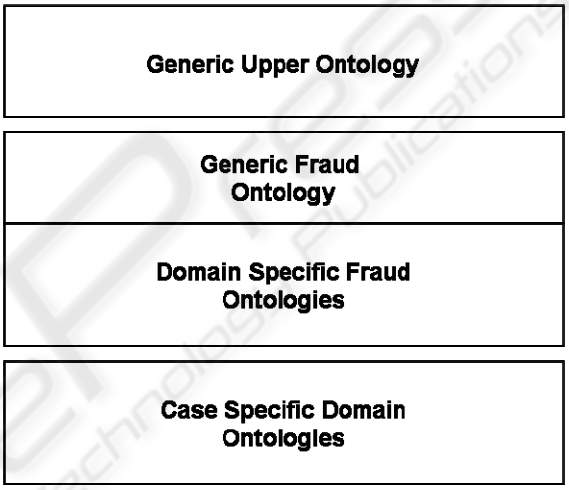


Figure 1: Fraud Ontology Layered Architecture.

The bottom layer (or case specific layer) consists of domain ontologies which model the business processes of the specific cases that are examined for fraud, e.g. a specific organization in social security. The concepts and relations contained in these ontologies are practically derived from the business process analysis of the particular case and from the knowledge of the corresponding domain experts. The main purpose of the case specific layer is to provide the basic knowledge on which fraud detection rules or data mining techniques are going to be based on. Reusability of existing ontologies is applicable nut only in the sense of best practices transfer from one case to another.

The middle layer (or fraud domain layer) comprises of ontologies which model fraud related knowledge such as fraud types and fraud detection processes. The content of these ontologies reflects the knowledge of fraud domain experts and it is primarily used as the basic means for expressing the fraud detection rules that these experts provide.

The middle layer could be considered as having two sublayers, a domain-specific one and a generic one. The domain-specific sublayer models the fraud characteristics of the domain at hand, e.g. social security or public procurement. The generic sublayer provides more abstract and generic knowledge that constitute the basis for applying knowledge-based approaches into virtually any fraud susceptible field. A small fraction of the generic fraud ontology is depicted in figure 2. As it can be seen from this diagram the fraud ontology contains concepts representing fraud actors, fraud cases etc and relations linking actors with motivations and cases with actors.

Finally, the upper layer, namely the Generic Upper Ontology, captures generic and domain-independent knowledge that helps minimize redundancy and duplication of knowledge within the overall ontology.

The most important of the advantages such a layered architecture provides, are the following:

- Modularity: When a large-scale ontology is composed out of smaller ontologies then its development and maintenance are easier and more efficient.
- Reusability: When the independent parts of the ontology are well defined and separated then it is highly possible that these parts can be reused in other similar applications.
- Extensibility: With the layered architecture, and more specifically with the generic ontologies, it is far easier to extend the ontology so that it can cover domains of application other than the existing ones.

# 5 CASE STUDIES

## 5.1 The Case of TSAY, a Greek Social Security Fund

TSAY is the insurance body of all healthcare professionals in Greece and its main focus concerning healthcare fraud is detected in the prescription reimbursement domain. Since TSAY is a health insurance body organization, one of the most common services it offers to its members is the payment of the drugs they consume. This payment has mainly the form of reimbursement meaning that a TSAY's member purchases the drugs s/he needs from a pharmacist paying only a percentage of the actual cost and then the pharmacist claims the rest of the money from TSAY.

However, it is often the case that the prescriptions TSAY is asked to reimburse contain erroneous or deliberately inaccurate data so that larger sums of money can be claimed or inappropriate drugs can be prescribed. Or, it is possible that prescriptions contain data which when viewed isolated do not indicate fraud but when considered along with other prescriptions they form some suspicious pattern of misbehaviour.

Of course, the cases targeted for detection do not necessarily constitute fraud from a legal point of view because it might be that the inaccurate data are due to human error or that the objectionable misbehaviour can be explained by reasons that are not obvious. However, even then, the need for detection remains strong since fraud in this case can be considered to be synonymous to waste in the form of monetary losses from the reimbursement of inappropriate prescription.

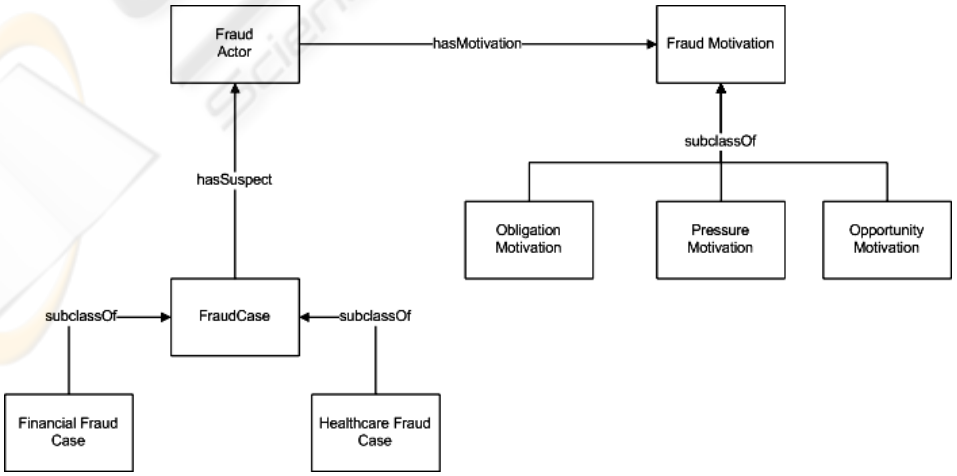


Figure 2: Generic Fraud Ontology.

### 5.1.1 TSAY Fraud Context and Fraud Identity

In the case of TSAY the fraud domain is that of prescriptions. According to our methodology the first required step was the establishment of the fraud context namely the description of the prescription domain. Thus, a business process modelling procedure was performed and a complete business process model of the prescription domain was developed. The high level processes contained in that model were:

- The issuance of prescription booklets to TSAY members by the Fund
- The issuance of prescriptions by doctors to patients that own these booklets
- The inspection of prescriptions by the ministry of health.
- The filling of members' prescriptions by the pharmacists
- The reimbursement process of TSAY for filled prescriptions.

According to the business process analysis, prescription issuance, inspection and filling occur outside the organization and TSAY has no control over the events that take place there. This meant that these processes could not be a part of TSAY' s fraud detection mechanism. On the other hand, the prescription reimbursement process was considered perfect for applying fraud detection methods and rules.

These methods and rules (the TSAY fraud identity or the second step of the methodology) were provided by people involved in the prescription process, namely doctors, pharmacists, TSAY's inspectors (patients could also be included).

The rules identified comprised two main categories, namely auditorial rules and medical rules. Auditorial rules try to detect incomplete prescriptions and invalid or miscalculated data while medical rules try to detect prescriptions in which the data are inconsistent from a medical point of view.

An example of an auditorial rule is when a prescription contains no diagnosis at all for the drugs that it prescribes and an example of a medical rule is when the diagnosis written on the prescription is not included in the indications of the prescribed drugs.

### 5.1.2 TSAY Domain Specific Fraud Ontology and TSAY Case Specific Domain Ontology

The third step of applying our methodology was the actual building of the TSAY specific ontologies. As described in section 4 these ontologies are the TSAY domain specific fraud ontology and the TSAY case specific domain ontology.

The first contains the knowledge regarding the prescription domain and utilizes the business process model created in the previous steps. The second models the fraud types and fraud detection methods and rules for the prescription domain and utilizes the knowledge derived from the domain experts. Both are built under the generic upper and fraud ontologies so that the development effort and knowledge redundancy are minimized. Figures 3 and 4 present fractions of these two ontologies.

Figure 3 depicts the refinement and specialization of a generic fraud case to the social security domain and especially to prescription related fraud. Several fraud cases identified in step 2 of the methodology are represented as concepts in the domain ontology.

Figure 4 presents the representation of a prescription as viewed by TSAY experts. The different concepts – entities, their characteristics and their relationships are depicted in the ontological model. It is clear from the figure that even this particular part of the TSAY case specific ontology can be transferred and applied to another organization that faces a similar increased risk in its prescription process with minor adaptation.

## 5.2 Other cases

In order to illustrate and test the generic character of our approach, we applied our methodology and the generic fraud ontology to three more cases and domains apart from that of TSAY's.

The first one concerned one of the largest cardiothoracic centre in the UK and part of NHS Trust, which provides specialist services for patients of all ages from across the UK, including Scotland and Wales. The centre's interest in fraud detection involved the identification of conflict of interest in the process of procurement of goods and services within the Trust.

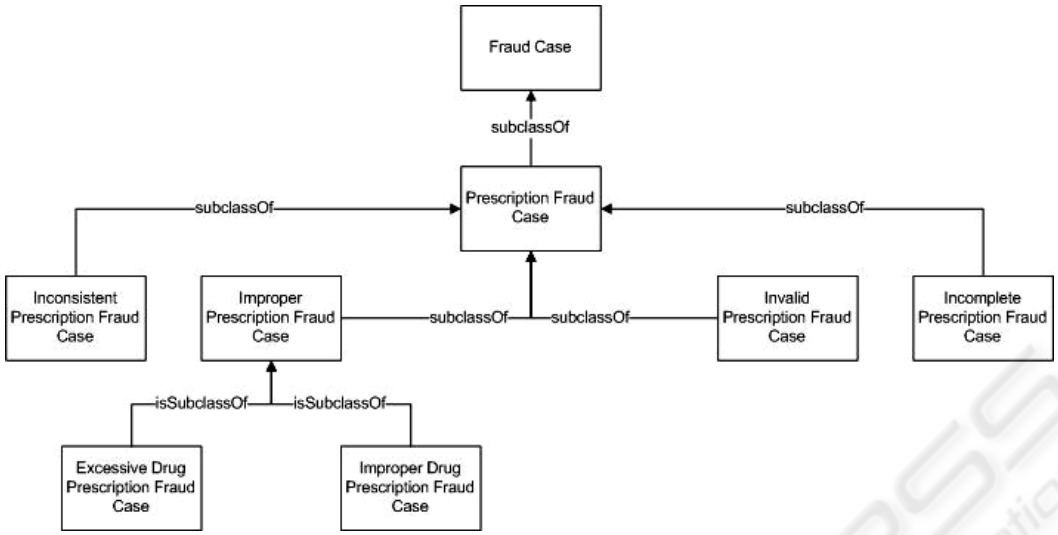


Figure 3: TSAY Domain Specific Fraud Ontology.

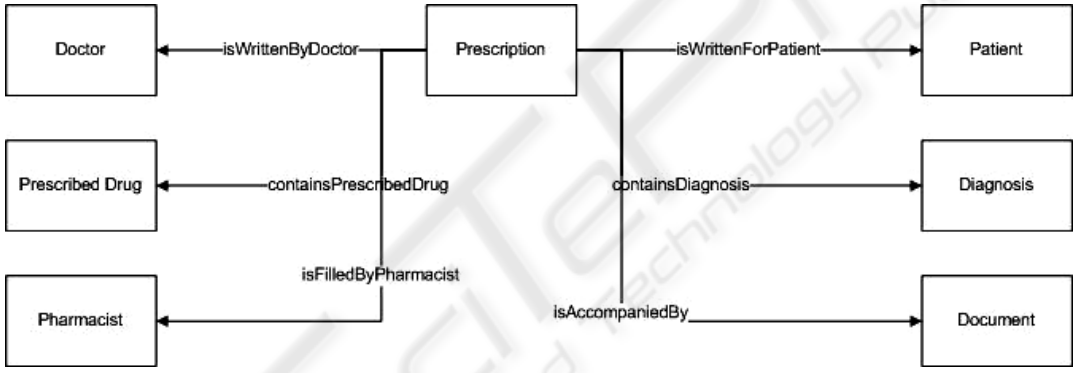


Figure 4: TSAY Case Specific Domain Ontology

Our approach to facilitating the detection of such kind of fraud was similar to the one we followed in the TSAY case. Thus, at first a business process model describing the way the procurement process performed within the centre was created and then a number of potential conflict of interest cases were identified along with corresponding fraud detection rules. All this knowledge was transformed correspondingly into the centre's Domain Specific Fraud Ontology and Case Specific Domain Ontology. The centre's experts evaluated the final ontology and found it adequate to cover the fraud detection process described during the first step of the methodology.

The second case concerned customs control and particularly fraud regarding tax evasion during the movement of goods between countries of the EU which originate from non-EU countries or pass through non-EU countries. Again, we followed the same procedure and we managed to create a complete ontological model of this kind of fraud. Finally, we applied our methodology in the field of Public Administration for assisting the General Inspector Office of Public Administrations to detect corruption and any other potential fraudulent activities that take place within the government. In both cases the final ontology for the particular organizations and domains was developed in a short period of time by applying the methodology and refining the generic ontology. The results were judged as satisfactory by organizations' experts.

## 6 CONCLUSIONS

In this paper we presented a methodology for building fraud ontologies across domains spanning the area of e-government. Fraud ontologies are

usually part of rule-based or predictive modelling fraud detection systems but they can also be utilized in data mining systems that try to discover fraudulent behaviour among seemingly irrelevant data. Our methodology is supported by a generic ontological framework (called fraud ontology) that can be used during the building of the domain specific fraud ontologies for increasing the efficiency of the whole ontology development process.

In essence, our methodology and generic ontology are tools that can be used by any knowledge engineer who needs to build a domain ontology for a fraud detection application in the field of e-government. The methodology provides the engineer a roadmap of how s/he should proceed with acquiring the required knowledge for the application while it leaves him/her free to choose the knowledge engineering tools and methods s/he wishes. On a second level, the fraud ontology provides the engineer useful insights of how the ontology should look like and helps him/her do the knowledge modelling more accurately, efficiently and with less effort.

## ACKNOWLEDGEMENTS

The work presented in this paper is funded by the European Commission under Grant FP6-2004-IST-4-028055.

## REFERENCES

Baader F., Calvanese D., McGuinness D. L., Nardi D., Patel-Schneider P.F.: The Description Logic Handbook: Theory, Implementation, Applications. *Cambridge University Press, Cambridge, UK, 2003.*

Belhadji, B. & Dionne, G., 1997. Development of an Expert System for Automatic Detection of Automobile Insurance Fraud, *Ecole des Hautes Etudes Commerciales de Montreal- 97-06, Ecole des Hautes Etudes Commerciales de Montreal-Chaire de gestion des risques.*

Blackburn, Patrick, Maarten de Rijke, and Yde Venema (2001) Modal Logic. *Cambridge Univ. Press*

Gomez-Perez Asuncion, Oscar Corcho, Mariano Fernandez-Lopez (2004) Ontological Engineering. *Springer-Verlang London Limited*

Crockford, Neil (1986). An Introduction to Risk Management (2nd ed.). *Woodhead-Faulkner. 0-85941-332-2*

Gruber TR (1993) A translation approach to portable ontology specification. *Knowledge Acquisition 5(2):1999-220*

Hand D., Mannila H., Smyth P. (2001). Principles of Data Mining. *MIT Press, Cambridge, MA*

Kerremans, Koen, Tang, Yan, Temmerman, Rita and Zhao, Gang (2005). Towards Ontology-based E-mail Fraud Detection. In: C. Bento, A. Cardoso and G. Dias, (eds.) *Proceedings of EPIA 2005 BAOSW Workshop of 12th Portuguese conference on AI*, Covilha, Portugal, p. 106-111.

Lam, James (2003). Enterprise Risk Management: From Incentives to Controls. *John Wiley. ISBN-13 978-0471430001*

Leary, R. M., VanDenBerghe, W. and Zeleznikow, J. 2003. Towards a financial fraud ontology. A legal modeling Approach. *ICAIL'03 Workshop on Legal Ontologies and Web-based Legal Information Management*

Noy N. F., D. L. McGuinness (2001). Ontology development 101: A Guide to Creating Your First Ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001.

Nute D. (1994). Defeasible logic. In Handbook of logic in artificial intelligence and logic programming, *volume 3: Nonmonotonic reasoning and uncertain reasoning, pages 353-395*. Oxford University Press.

Tadepalli S., A.K. Sinha, N. Ramakrishnnan (2004). Ontology driven data mining for geosciences. *Proceedings of 2004 AAG Annual Meeting, Denver, USA, 2004.*

Zukerman I., D.W. Albrecht (2000). Predictive statistical user models for user modeling. *User Modeling and User-Adapted Interaction 11(1-2), 5-18*

Hodges W, 2001, "Classical Logic I: First Order Logic," in Lou Goble, ed., *The Blackwell Guide to Philosophical Logic*. Blackwell.