



**TOWARDS A FINANCIAL FRAUD ONTOLOGY
A LEGAL MODELLING APPROACH**

*Richard M Leary, Wim Vandenberghe and John Zeleznikow
Joseph Bell Centre for Forensic Statistics and Legal Reasoning
School of Law, University of Edinburgh*

Abstract: This document discusses the status of research on detection and prevention of financial fraud undertaken as part of the European Commission funded FF POIROT (financial fraud prevention oriented information resources using ontology technology) project. A first task has been the specification of the user requirements that define the functionality of the financial fraud ontology to be designed by the FF POIROT partners. It is claimed here that modeling fraudulent activity involves a mixture of law and facts as well as inferences about facts present, facts presumed or facts missing. The purpose of this paper is to explain this abstract model and to specify the set of user requirements.

Keywords: legal ontology, knowledge modeling, financial fraud analysis, user requirement analysis, facts, evidence, law, WWW.

CONTENTS

Abbreviations and Definitions	p. 4
INTRODUCTION	p. 5
1. FF POIROT PARTNERS	p. 6
2. THE FRAUD CHALLENGE	p. 7
3. SITUATING FFPOIROT IN LEGAL AND FINANCIAL ONTOLOGY RESEARCH	p. 8
4. SITUATING FF POIROT IN COMPUTER-LINGUISTIC RESEARCH AND DEVELOPMENT	p. 9
5. SITUATING FF POIROT IN ONTOLOGY RESEARCH AND ONTOLOGY TOOLS DEVELOPMENT IN GENERAL	p. 9
6. MODELING ONLINE INVESTMENT FRAUD – A LEGAL APPROACH	p. 11
<i>6.1 Abstract Model</i>	p. 11
<i>6.2 Complexity of the Evidence Layer</i>	p. 13
<i>6.3 The Role of Evidence in Argumentation</i>	p. 14
<i>6.4 Anatomy of the Case as a Whole – A Case Study</i>	p. 16
<i>6.4.1 Facts</i>	p. 16
<i>6.4.2 Abstract Model – Low Level of Detail</i>	p. 17
<i>6.4.3 Abstract Model – Higher Level of Detail</i>	p. 18
<i>6.4.4 Macro and Micro Analysis</i>	p. 19
<i>6.5 Evaluation of the Abstract Model</i>	p. 19
7. USER REQUIREMENTS ANALYSIS p. 20	
<i>7.1 Method and Breakdown of Users</i>	p. 20
<i>7.2 Strategic Requirements</i>	p. 21
<i>7.2.1 General</i>	p. 21
<i>7.2.2 Top Level Requirements</i>	p. 23

<i>7.3 User Requirements for CONSOB Type Fraud</i>	p. 25
<i>7.4 User Requirements for VAT Fraud</i>	p. 27
<i>7.5 Meeting the Requirements</i>	p. 29
8. CONCLUSION	p. 30
Annex 1	p. 31

Abbreviations and Definitions

Actors	Users and external systems
AI	Artificial Intelligence
GUI	Graphical User Interface
IR	Information Retrieval
KIF	Knowledge Interchange Format
NLP	Natural Language Processing
OET	Ontology Extraction Tool
UML	Unified Modeling Language
User requirement	A clear statement which describes what is expected

INTRODUCTION

It is estimated that the EU loses several millions euro per year due to financial fraud. Therefore it should not come as a surprise that prevention and early detection of fraudulent activity is an increasingly important goal for the EU and its Member States.¹ The impetus for building financial fraud ontology results from the need to supplement the efforts of EU Member States to combat financial fraud, especially from a more less-obvious angle.

To this end, the FF POIROT project follows in the detective star Hercule Poirot's footsteps to provide *inter alia* law enforcement agencies with a novel approach to solve the financial fraud puzzle.² FF POIROT evaluates fraud control assumptions, policies and systems in terms of their effectiveness in deterring, preventing and detecting criminal fraud.

The goal of the project is to build a detailed ontology of European Law, preventive practices and knowledge of the processes of financial fraud within the European Union. It aims at compiling for several languages (Dutch, Italian, French and English) a computationally tractable and sharable knowledge repository (a formally described combination of concepts and their meaningful relationships) for the financial fraud domain.

The objective of this paper is to set out the user's requirements for the ontology to be developed. It sets out the user's needs in the form of key functions to be performed by the ontology. The paper starts off with a brief introduction to the partners in the project (section 1). Section 2 provides a formulation of the fraud problem and defines the parameters of the research. Section 3, 4 and 5 situate FF POIROT in the ongoing ontology research and development.³ Section 6 gives specific attention to describing the legal model and applies this model to the case of investment fraud online. Section 7 specifies the user requirements and makes it clear to which extent the users of the ontology will be supported. Section 8 concludes.

¹ See Communication from the Commission, Protecting the Communities' Financial Interests. Fight Against Fraud. Action Plan for 2001-2003, COM(2001) 254 final. A recent effort by the UK government is the initiative by HM Customs & Excise, titled "Protecting Indirect Tax Revenues", designed to save £2 bn a year.

² See <http://www.ffpoirot.org>

³ See ffpoirot consortium agreement.

1. FF POIROT PARTNERS

The partners in the FF POIROT projects include legal academics, computer science academics, linguists, software houses, and two user partners, CONSOB and VAT Applications who wish to commercialize the consortium's results.

CONSOB is the public authority responsible for regulating the Italian securities market (a similar body to the UK's Financial Service Authority).⁴ It is the competent authority for ensuring: transparency and correct behavior by securities market participants; disclosure of complete and accurate information to the investing public by listed companies; accuracy of the facts represented in the prospectuses related to offerings of transferable securities to the investing public; and compliance with regulations by auditors entered in the Special Register. The purpose of securities regulatory authorities is to protect investors from unfair, abusive or fraudulent practices, and fostering fair, efficient and competitive capital markets that will provide investment opportunities and access to capital. To this end CONSOB is *inter alia* analyzing investment scams on the World Wide Web and develops appropriate software.

VAT Applications NV is a Belgian software company developing automated software to deal with issues surrounding Value Added Tax at a European and international level.⁵ It has packages for all 15 EU Member States and in eleven languages. The recent decision to add ten new members to the European Union in 2004, will put further pressure on to develop software packages to help compliance with VAT requirements across the European Union and the identification, prevention and reduction of fraud across jurisdictions.

The University of Edinburgh's Joseph Bell Centre for Forensic Statistics and Legal Reasoning is performing the following tasks:⁶

- Prepare for the construction and testing of the financial forensics repository using macro and micro analytical techniques. Developing 'formal' methods for CONSOB's and VAT@'s approach;

⁴ Commissione Nazionale per le Società e la Borsa. See <http://www.consob.it>

⁵ [Http://www.vatat.com](http://www.vatat.com)

⁶ [Http://www.cfslr.ed.ac.uk](http://www.cfslr.ed.ac.uk)

- Gather information on how relevant authorities accumulate and analyse evidence of financial fraud, and analyse the tools auditors and accountants use to maintain up-to-date awareness of financial services and VAT regulations;
- Collect requirements for the retained data, its validation and the applications needed to optimise the use of the information.

Thus the Joseph Bell Centre will not build the FF POIROT ontology – JBC will just conduct the user requirements analysis and the knowledge modeling. However, understanding financial fraud itself is the foundation for any successful software development. As JBC is a partner in the consortium of institutions who will eventually build the ontology, this paper will nevertheless reflect on the construction of the forensic ontology.

2. THE FRAUD CHALLENGE

Financial fraud is growing faster than international trade. Frauds are *prima facie* more complex and involve larger sums than ever before. More than any other wrongdoing, fraud may involve both civil and criminal legal action. The focus of the FF POIROT project is quite deliberately on criminal fraud. Criminal fraud is clearly enough defined, requiring a deliberate misrepresentation or deception leading to some kind of improper pecuniary advantage.

As financial fraud is a very broad field, we have to delineate it to very concrete sub-domains that exist in the fraud area. Our initial focus is to examine cross-border Value Added Tax fraud within the EU and Investment fraud on the World Wide Web. This corresponds respectively with the domain expertise of VAT@ and CONSOB.⁷ This paper will only discuss investment fraud on the World Wide Web. The fraud may be committed or attempted in a number of ways and these will be described in the following section.

⁷ Eventually we might investigate other fraud domains, such as insurance fraud, money laundering, etc.

3. SITUATING FFPOIROT IN LEGAL AND FINANCIAL ONTOLOGY RESEARCH

In developing computer resources such as forensic ontologies, in particular as evidence support for transnational issues in Europe such as financial fraud, we need to be aware of the different legal systems in the European Union. Clearly such resources will also fulfill a strong documentary need for many bona fide organizations that depend on meaningful insight in Europe's complex multi-national regulations.

There is an urgent need to examine the various European legal systems when trying to build forensic ontologies in Europe. A major reason is that crimes are regularly being committed in transnational domains. This is certainly the case for the two sub-domains of financial fraud under FF POIROT scrutiny.

Forensic evidence, by definition, can be used in criminal or civil courts. However, in the FF POIROT project, we confine ourselves to an examination of ontologies for criminal law. A major reason for taking this decision is the different burdens of proof in civil law when compared to criminal law.

We are not aware of the existence of any forensic ontologies, or indeed of any decision support systems related to forensics. There has been more research in the area of forensic statistics, and in one of the work packages of the project we will in fact investigate new possible links with ontology research.⁸

One of the resources to be built is a - partial – ontology of financial fraud evidence. To illustrate the role and importance of an evidence ontology, consider that in the process of fact investigation many things are to be discovered including hypotheses (or possible conclusions), evidence, and arguments linking hypotheses and evidence. These arguments are generated in defence of the relevance and credibility of evidence and form the basis for subsequent assessments of the probative force of evidence. During fact investigation, of each episode of which is unique in law, we have hypotheses in search of evidence at the same time we have evidence in search

⁸ AITKIN has shown how statistical data bearing on the significance of tangible trace evidence in criminal investigation can be useful to forensic scientists. See C. AITKIN, *Statistics and the Evaluation of Evidence for Forensic Scientists*, 1995.

of hypotheses.⁹ Also to be generated or discovered are arguments linking the evidence and hypotheses. While FF POIROT will not cover these aspects of legal reasoning itself, the ontology developed will be a valuable resource to experiment with computer implementations of such reasoning in other ongoing research projects.

4. SITUATING FF POIROT IN COMPUTER-LINGUISTIC RESEARCH AND DEVELOPMENT

An important aspect of FF POIROT is the mining of (formal) ontology elements from unstructured or semi-unstructured resources such as lexicons, text databases, XML documents, RDF schemes, law texts, and of course the Internet. This involves the processing of natural language, which however in the context of FF POIROT will be assumed to be an *a priori* limited to the above mentioned component domains. Language understanding is a process that traditionally is recognized to be the result of various kinds of knowledge: phonological, morphological, syntactic, semantic, pragmatic and world knowledge.¹⁰

For the purposes of FF POIROT, it is possible to simplify the picture and to adopt a somewhat reduction view. Firstly, we can make abstract from the discourse level. Authors of legal documents or descriptive reports on forensic issues in general merely want to convey facts, and not to invoke emotions or to initiate actions by the reader. As such, we can limit our analysis to what in the speech-act literature is known as constative inscriptions, sentences uttered in a descriptive context, however without being too narrow as is the case in the traditional formal linguistic semantics scene where sentence-meaning is viewed as being exhausted by propositional content and is truth-conditionally explicable.¹¹ Since multilingual resources are one of our main objectives, we can however hardly ignore morphology.

5. SITUATING FF POIROT IN ONTOLOGY RESEARCH AND ONTOLOGY TOOLS DEVELOPMENT IN GENERAL

⁹ D. SCHUM and P. TILLERS, "Marshalling Evidence in Adversary Litigation", *Cardozo Law Review* 1991.

¹⁰ J.F. ALLEN, *Natural language Understanding*, 1994.

¹¹ J. SEARLE, *Minds, Brains and Programs. Behavioral and Brain Sciences*, 1980.

The current interest in ontology technology as limited to the context and terms of the FF POIROT project has its roots in artificial intelligence on knowledge representation. Basic research has concentrated on formal aspects trying to determine the underlying fundamental notions of the way we view the world and its organization, to the point of involving and formalizing central ideas of philosophy.¹² A notable practical effort is the attempted distributed development by the IEEE of a Standard Upper Ontology (SUO).¹³ On the other hand, a lot of attention has been focused on the construction of ontologies from a software engineering perspective. Products of this research were the first partial methodologies for the specific development of ontologies such as METHONTOLOGY and techniques for semi-automatic ontology acquisition.¹⁴

Many applied results have been achieved in the field. CYC and Ontolingua are best known.¹⁵ The latter is based on knowledge interchange format (KIF) and considered an important and influential formalism with Lisp-like representation of ontologies, and has been proposed as an ANSI standard and is also being used in the IEEE SUO effort cited above. We will however not adopt this latter formalism in FF POIROT for its suspected lack of scalability to the size of the terminology databases and corpora we expect to mine, or the number of concepts we will need to align to in the project. Instead, we shall adopt a more "layered" approach for the FF POIROT ontology and its tool set inspired by the way "classical" large database systems are typically set up. According to the well tried ontology base, a set of "language independent" domain specific atomic facts, which we will call *lexons*, and instances of their explicit interpretations.¹⁶ The latter in this way form a separate layer mediating between the ontology base and the application instances committing to the ontology. It is precisely this aforementioned separation in relational database systems that allows for the high efficiency and scalability of its management tools (DBMS), the sometimes huge size of the "models", and high volume transaction processing. Also, a number of well-known techniques from the database view and schema integration will be evaluated and tested for alignment and merging of the different parts of the

¹² J.F. SOWA, *Knowledge Representation: Logical, Philosophical and Computational Foundations*, 2000; N. GUARINO, "Formal Ontology in Information Systems", in N. GUARINO (ed.), *Formal Ontology in Information Systems. Proceedings of FOIS'98*, 1998.

¹³ [Http://suo.ieee.org](http://suo.ieee.org)

¹⁴ M. FERNANDEZ, A. GOMEZ-PEREZ and N. JURISTO, "METHONTOLOGY: From Ontological Art Towards Ontological Engineering", in *Workshop on Ontological Engineering, AAAI97*, 1997.

¹⁵ [Http://logic.stanford.edu/kif/kif.html](http://logic.stanford.edu/kif/kif.html)

¹⁶ STARLAB, Vrije Universiteit Brussel.

ontology.¹⁷ Since the ontology will be built up from components for the different sub-domains (evidence, law and finance) the issue of scalability for ontology modeling will be encountered in each of these three dimensions separately as well as across different domains when these sub-domains are aligned and/or merged to create the financial forensics ontology.

6. MODELING ONLINE INVESTMENT FRAUD – A LEGAL APPROACH

6.1 Abstract Model

In order to identify signals useful to the reduction of uncertainty associated with the presence of suspect solicitation agents on WWW, it is useful to be able to analyse the fraud in the form of an abstract model. This section will assess the potential of the fraud model and explain the methodology which supports the model. The objective of our legal model is that it should be helpful in the process of designing the financial fraud ontology.

In the tradition of Wigmore and Twining, the proposed method of knowledge modeling is by using inference networks of law.¹⁸ These are representations for complex probabilistic reasoning tasks often based on masses of evidence. This is highly useful as online investment fraud cases are notorious for huge data files to be investigated. The model is a directed acyclic graph; it is directed because it shows the direction of reasoning, or the direction of probabilistic influence among nodes on the network. It is acyclic because, following any reasoning path, you are never led back to where you started.

Further the model is an integrated logic based model.¹⁹ It's essential that crime investigators work within a guiding frame so that certain items become evidence and certain other items can be discarded. The better they get at the art of solid reasoning, the more efficient they'll be in solving their cases. Using logic to untangle

¹⁷ S. NAVATHE and S. GADGIL, *A Methodology for View Integration in Logical Database Design*, 1982; C. BATINI, M. LENZERINI and S. NAVATHE, *A Comparative Analysis of Methodologies for Database Schema Intergration*, 1987.

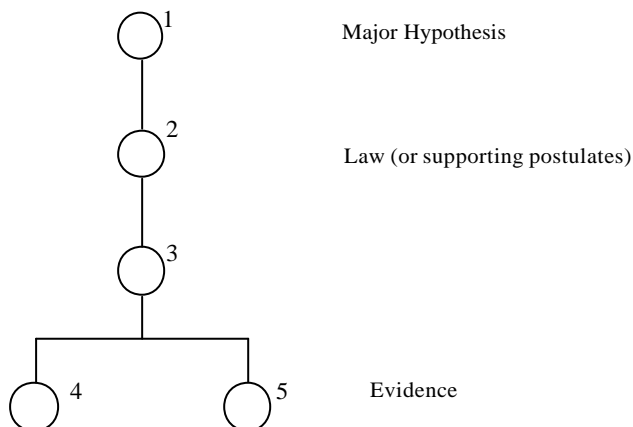
¹⁸ J.H. WIGMORE, *Principles of Judicial Proof as Given by Logic, Psychology and General Experience and Illustrated in Judicial Trials*; W. TWINING, *Rethinking Evidence*, 1990.

¹⁹ For the purposes of reasoning about evidence in fraud cases it seems to us there are three types of logic to take account of: logic of deduction, logic of induction; and logic of abduction. Depending upon the problem faced, any one or any combination of these forms of logic can be used. It is not a binary issue of logic or no logic.

complex scams and solve the complicated puzzles of crime means the difference between dead reckoning that can steer an investigator in the right direction and random guessing that can make things hopelessly confused.

The main idea of the model is that fraud cases can be broken down into three distinct layers of information. Firstly, a proposition (hypothesis) layer, secondly a law layer and thirdly the evidence layer. Any case will therefore have a layer of information about the hypothesis or case theory, for example, (X Defrauded Y), a layer of information about specific elements of law that need to be satisfied if a case of fraud is to be proven and, thirdly, there is an evidence layer comprising all the material facts and evidence that go to make up the facts of the case.

The following is an abstract model of the different layers involved in constructing or assessing a legal case. Note that the chart provides the abstract model and the Key List describes the component parts of the model:



Key List:

1. What is the ultimate intended aim (major hypothesis or proposition)
2. What is the substantive law that will be breached if the ultimate intended aim is reached
3. What are the acts or omissions that need to be undertaken (or not undertaken) if the ultimate intended aim is to be achieved?
4. What acts or omissions are generally seen if the ultimate intended aim is to be achieved?

5. What acts or omissions are generally not seen if the ultimate intended aim is to be achieved?

The abstract model is arrived at by asking a series of questions aimed at exposing the relationships between propositions of law and propositions of evidence. Although these questions appear in the form of a hierarchy above, when being used to discover the presence of a fraud within a tangible or intangible environment, they can be asked within any sequence including 'top down' and 'bottom up.' Logic can therefore be both *ex ante* or *ex post*. In other words, we can move from facts to conclusions and from conclusions to facts. An investigation may be a waste of time and money, however valid its legal conclusion/hypothesis, if evidence is not gathered to support that conclusion.

However, it should be borne in mind that in terms of completeness, no matter how thorough the query and search process is, there will always be unanswered questions and, no conclusion, no matter how well formulated, can ever account for all the facts we may potentially encounter. The evidence layer can become extremely complex for a number of reasons which will be explained in the next section.

6.2 Complexity of the Evidence Layer

Firstly, evidence or facts are always context specific. That is, the relevance of the evidence will be determined by the circumstances in which it is under consideration. Any item of evidence can be used for more than one purpose. It is not unusual for evidence to be used by different sides in a case for different purposes. A prosecutor may seek to use evidence of previous bad character to show that a suspect has a propensity for particular types of behaviour whereas the defence may seek to use the same evidence of previous bad character to demonstrate that the suspect has not offended for a considerable amount of time and is therefore reformed. In other circumstances, the defence may choose to use the same evidence to demonstrate that the suspect could not have fully participated in the crime because he was in prison during the preparatory stages of the offence.

Secondly, evidence never arrives in the hands of the user with its credentials made out. The relevance, credibility and weight of the evidence always has to be assessed

and declared.

Thirdly, the user of evidence is always biased to some degree in the interpretation of evidence. Each user should be aware of their 'standpoint' in using evidence and be prepared to declare it. Different persons have different standpoints each of which may result in a different interpretation of the evidence.

Fourthly, evidence is a word of association and therefore it can only be assessed by comparing, contrasting and juxtaposing it with other evidence and hypotheses. Hence, it is not possible to have a 'single item of stand alone evidence.' There is always other evidence.

Fifthly, all evidence can be broken down into smaller component parts. For example, an item of documentary evidence may be made up of paper, writing, ink, type face, a post mark, glue. The document may even have a fingerprint, a DNA stain or a discarded hair stuck to it. This atomistic view of evidence results in a situation where all evidence can be infinitely broken down into smaller and smaller parts which means that it can always be seen in the light of other evidence. This ancillary evidence about evidence can provide important insights into the relevance, credibility and weight of the evidence as a whole. A question for the user is always going to be "at what level of detail and at what level of granularity should the abstract model be considered to be complete?" This question is most important in fields such as criminal law. The reason is that it bears directly on the forensic standard "beyond a reasonable doubt". Regardless of how well a particular model appears to be formulated, there is always room for doubt. Facts are based on evidence and evidence always falls short of certainty.

6.3 The Role of Evidence in Argumentation

Arguments are made up of hypotheses, sometimes called propositions or case theory's, chains of directly relevant or indirectly relevant evidence and generalizations. Generalizations are generated by humans from perceived signals and stimulus in the environment. Generalizations may be presented by one person to another in circumstances where they become simply 'accepted facts.' These interactions take place between humans in the normal course of communication.

Others are formulated by direct perception of new signals by a single person. In reality, the process of reasoning from evidence (signals) to hypothesis (case theory) inevitably involves both. A useful way to think about generalizations is that they are clusters of signals assembled into forms that resemble an explanation or a story.

Because generalizations are constructed by the clustering of different forms of signals and stimulus from different sources of information, in the pursuit of different objectives and by different people, they exhibit highly subjective characteristics. This means that generalizations need to be managed with care. Understanding the fundamental steps in the construction of explanations and stories from mixtures of hypotheses, evidence and generalizations provides valuable insights into human decision-making. Furthermore, assessing the reliability of the grounds upon which an explanation is constructed provides a means to grade validity. Assessments about validity of explanations are inevitably uncertain and therefore can only be used as inferences towards or away from the hypothesis under consideration. That does not mean they are of limited use. If maximizing the frequency of desired outcomes as opposed to undesired outcomes is important, systematic methods have much to offer humans engaged in processes like investigation, decision making or the assessment of risk.

Generalizations are formulated by first and second hand exposure to information about events in the environment. We receive and process signals generated by these events and experiences and our sensory receptors process the signals into scenarios we can store and recover from memory. Views are generated about event types and often about causality. The purpose of this process is in ways that can assist us in dealing with future. These views or We formulate views of event are supported or negated by ancillary evidence. Evidence is therefore but one component used in the construction of legal arguments.

Inferences flow from items of evidence, generalizations and ancillary evidence towards or away from the hypothesis or proposition. Evidence therefore tends to support or tends to negate the hypothesis under consideration. Handling evidence in cases is therefore complex by virtue of the number of different ways the hypothesis, evidence, generalizations and ancillary evidence can be brought together.

6.4 Anatomy of the Case as a Whole – A Case Study

6.4.1 Facts

The following model was extracted from an actual case file of unauthorized online solicitation that occurred within the jurisdiction of the Italian financial market regulated by CONSOB.²⁰ The company Smallxchange, headquartered in the British Virgin Islands, aims to become an unofficial 24-hour stock exchange on which any company in the world can be listed at no charge. In return the investors are asked to tender shares in exchange for a stake in the venture. The shares will then be traded between partners in this unofficial stock exchange. Investors were solicited by a WWW Page advertising financial investment services (public offering of participation certificates).

In order to establish the efficacy of regulation of fraud on the Internet, it is necessary to consider whether CONSOB has appropriate jurisdiction.²¹ CONSOB considered that its jurisdiction was asserted as Smallxchange targeted the national investment market of Italy.²²

The soliciting agent was not licensed to trade as required under Directive 93/22/EEC (and its implementation in Italy: Legislative Decree 58) and false statements were made on the web page.

The illustration is an abstract model of the fraud. Note that the model is comprised of true claims (signals) as well as false claims. Node 10 is a false statement aggravating the fact that the company, although properly constituted in law in UK, was not licensed to solicit investment services. This is a simple model but a more detailed model follows later.

The chart depicts a simplified version of an investment scam online. In fact it could

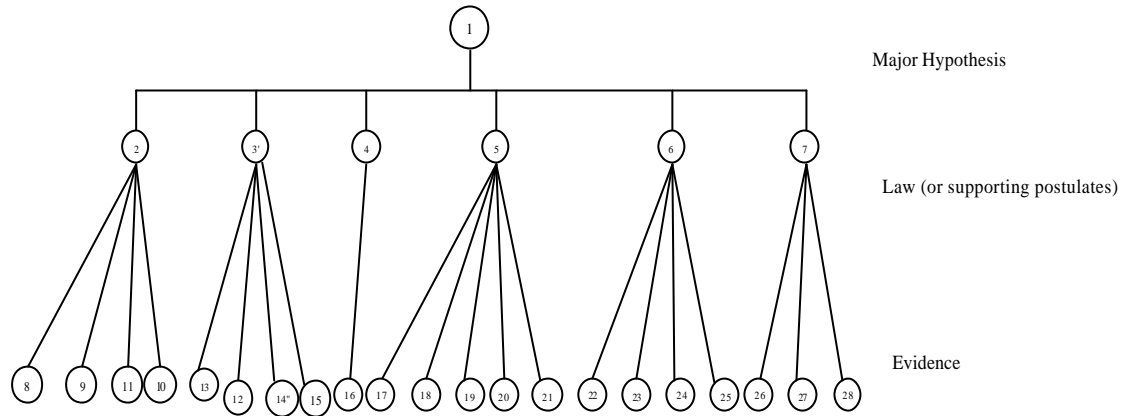
²⁰ See CONSOB's weekly newsletter (October 2001).

²¹ This must be considered separately from the regulator's ability to enforce its powers within such jurisdiction. For example, a securities regulator would first need to consider whether a security was being advertised or sold within its geographical jurisdiction and secondly, whether or not the person advertising the product was subject to their regulation.

²² For an analysis of the jurisdictional aspects of the case, see, J.A. GRAHAM, "The Cybersecurities' Notion of Targeting in General Private International Law", *Cyberbanking & Law* 2003.

involve many more individuals, etc.

6.4.2 Abstract Model - Low Level of Detail



Major Hypothesis

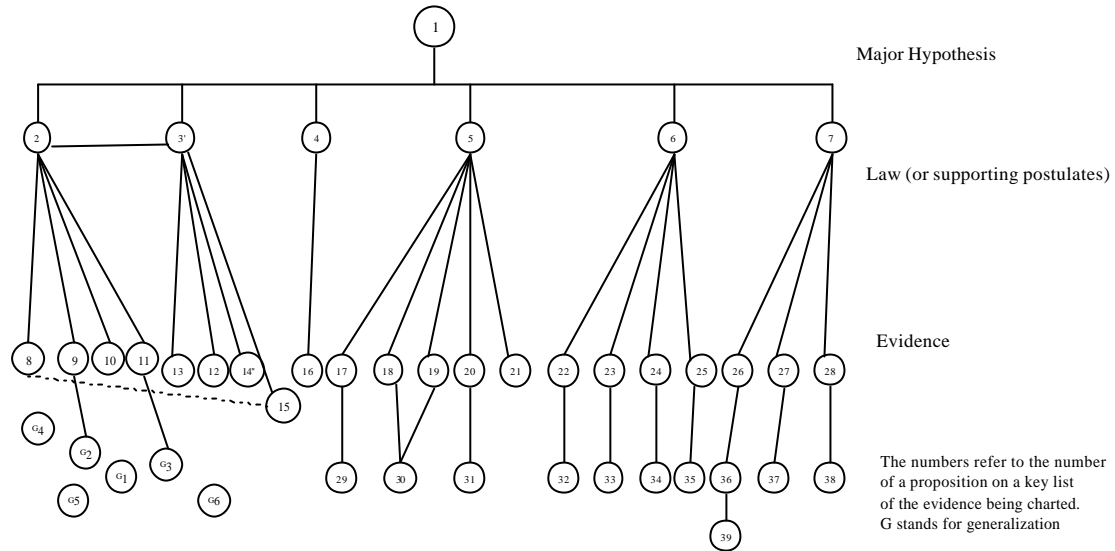
Law (or supporting postulates)

Evidence

The numbers refer to the number of a proposition on a key list of the evidence being charted.

1. Smallxchange.Com Ltd is fraudulently solicited investment services on the WWW
- 2.Targets the national market (i.e. the Italian public/investor market) (assertion of jurisdiction)
3. Smallxchange solicited securities services on WWW for investors
4. Requirement to inform Consob of its existence and to comply with minimal disclosure rules on technical details
5. Unauthorised solicitation of investors (contra article 94 of Decree 58)
6. Unauthorised alternative trading system (contra article 102 of Decree 58)
7. Placing of unauthorised funds
8. An Italian ISP hosted the site
9. Most of the advertising was in Italian
10. Company is run by Italian executives
11. Possibility of paying in lira/euro
12. Non fulfilment
13. Web site lists shares for sale
14. Gianni Alterie was offered investment services by Smallxchange via email dated August 29 2000.
15. Http://www.smallXchange.com is a Web Site managed by Smallxchange Ltd.
16. Http://www.smallXchange.com is hosted by Smallxchange Italia s.r.l.
17. Public offering of financial products
18. Shares of the companies listed on the stock exchange
19. Shares of the stock exchange in exchange for shares of the listed companies
20. Mutual fund shares
- 21'.Unauthorised public offering
22. Organization of a stock exchange
23. Quoted companies
24. Clearing house
25. Trading book
26. Placing of a fund
27. Asset management company
28. Fund name

6.4.3 Abstract Model - Higher Level of Detail



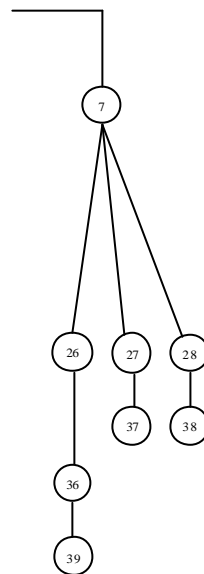
1. Smallxchange.Com Ltd is fraudulently solicited investment services on the WWW
 - 2.Targets the national market (i.e. the Italian public/investor market) (assertion of jurisdiction)
 3. Smallxchange solicited securities services on WWW for investors
 4. Requirement to inform Consob of its existence and to comply with minimal disclosure rules on technical details
 5. Unauthorised solicitation of investors (contra article 94 of Decree 58)
 6. Unauthorised alternative trading system (contra article 102 of Decree 58)
 7. Placing of unauthorised funds
 8. An Italian ISP hosted the site
 9. Most of the advertising was in Italian
 10. Company is run by Italian executives
 11. Possibility of paying in lira/euro
 12. Non fulfilment
 13. Web site lists shares for sale
 14. Gianni Alterie was offered investment services by Smallxchange via email dated August 29 2000.
 15. [Http://www.smallXchange.com](http://www.smallXchange.com) is a Web Site managed by Smallxchange Ltd.
 16. [Http://www.smallXchange.com](http://www.smallXchange.com) is hosted by Smallxchange Italia s.r.l.
 17. Public offering of financial products
 18. Shares of the companies listed on the stock exchange
 19. Shares of the stock exchange in exchange for shares of the listed companies
 20. Mutual fund shares
 - 21'.Unauthorised public offering
 22. Organization of a stock exchange
 23. Quoted companies
 24. Clearing house
 25. Trading book
 26. Placing of a fund
 27. Asset management company
 28. Fund name
 - 29.-36. Etc.
 37. Tricalpa Investment Inc.
 38. Millenium Bug International
 39. Etc
- G₁ The presence of references to facts and circumstances concerning Italy in the site
G₂ The Employment of the Italian language
G₃ The indication of prices or amounts in Italian lire or Euro
G₄ The operations in Italy of intermediaries through which it is possible to carry out or agree to the promotion or placement executed through the Internet
G₅ The spreading of information in Italy; it is included the execution of individualized or mass advertising or information campaigns in Italy, with the object similar to the site contents
G₆ The availability of the site through search motors specialized in Italy or Italian

6.4.4 Macro and Micro Analysis

The abstract model lends itself to both macro and micro analysis. The model is a network of links and relationships between different items of evidence (signals). This allows finer analysis to be undertaken including the introduction of new signals. The analyzer of the case must however at some stage decide when to stop. The nature and characteristics of evidence signals means that the analyst could continue to analyse to ever smaller levels of detail.

The following abstract model is a micro analysis of elements 7, 26, 27, 28, 36, 37 and 38 from the key list. Element 7 in the key list refers to a section of the law which demands that if the placing of unauthorized funds is to be proven, then an inducement to acquire securities must be shown to exist.

Different sections of the abstract model can be broken down in this way to provide ever more detailed analysis.



6.5 Evaluation of the Abstract Model

The abstract model provides a number of benefits in the construction of fraud templates:

- It uses rational principles for the combination of the key logical components in

- fraud: A proposition, the law, evidence and generalizations;
- It facilitates a method for combining key logic. Same 'logic' in detection, prevention and prosecution;
 - Provision of both holistic and atomistic (macro and micro) analysis of the key components in fraudulent activity. Consequently it facilitates to imagine all the sources of doubts that may lurk between the evidence;
 - It ensures that the analysis of fraud cases is undertaken in a rational, repeatable way which can be conveyed to others for analysis and use (checking the coherence of your argument; convincing others of the relevance of an item of evidence);
 - It provides the basic abstract model from which computational models can be constructed;
 - It facilitates both 'top down' and 'bottom up' analysis of the key components in fraud.

7. USER REQUIREMENTS ANALYSIS

What key features and requirements of the financial fraud ontology are important from the user's point of view? This section will set out a systematic requirements analysis by identifying and documenting the main needs of fraud investigators.

7.1 Method and Breakdown of Users

Requirements gathering can be done using a number of different methods separately or in combination. We decided to organise the activities along three lines. First, we consulted the literature on financial fraud. To complement the literature review of the phenomenon of financial fraud to include the practice of fraud with VAT and fraud with securities, a number of structured interviews were conducted with representatives from several institutions in all three countries. Also consortium expertise was used to accumulate necessary information for the construction and testing of a financial fraud ontology. Finally we had brainstorming sessions.

The ontology should at least be useful to three different and EU-relevant types of user communities:

- financial professionals : Accountants, auditors, banks, insurance agencies, government departments, regulators and financial experts
- Police and other law enforcement agencies
- Investigative and monitoring bodies

The emphasis is on defining what is required in terms of the information requirements rather than how the system should be physically implemented. Step one was to figure out what information would help investigators do their job better. Thus we start of with explaining the ontology functions. Then we will scrutinize the specific requirements for each fraud sub-domain. Analogue with this is an analysis of the system attributes.

7.2 Strategic Requirements

7.2.1 General

The system has to be effective and efficient in deterring and preventing financial fraud. Central to achieving this is the identification of fraudulent activity from what is often a vast array of data. Eliminating legitimate activity from illegitimate activity is at the core of the problem, a fact that fraudsters know and use to hide illegal operations.

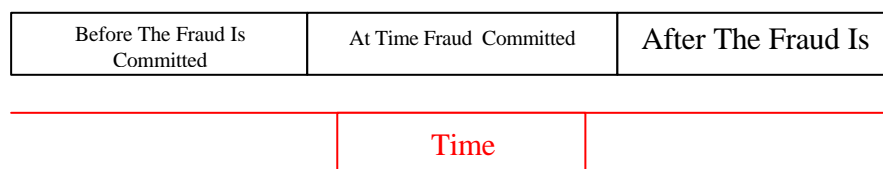
Illegitimate activity ranges from complex, organised and well thought through fraud to simple, disorganised and opportunist fraud. However, care needs to be taken not to assume that because a fraud is simple, disorganised and not particularly well planned that it will not have serious consequences. The www provides many examples of simple single frauds perpetrated for what may appear small financial gains. However, when these are added together, large amounts are often involved. This is used by fraudsters in both VAT fraud and Investment and Securities fraud. Single fraudulent transactions may involve small sums but when added together in the form of a 'continuing fraudulent operation,' massive sums of money can be involved and often over small timescales.

All fraud, complex or simple, involves at some stage of the process a breach of trust, confidence and fiduciary duty between a victim and a fraudster. Gaining and harnessing an understanding of this relationship and building it into the ontology, provides many opportunities for the identification of fraud.

VICTIM	FRAUDSTER
Buyer	Seller
User	Supplier
Employee	Employer
Investor	Investment Advisor
Principal	Agent
Beneficiary	Trustee
Manufacturer	Vendor
Stockholder	Executive
Customer	Broker

Temporal Awareness of Fraudulent Actions

Identifying different types of *Patterns* of fraudulent activity over time is a useful inquisitive technique for focusing the process of enquiry. A useful methodology is to develop a classification of events and activities that take place at different stages of fraud as the process unfolds. For example, events and activities that take place in before the fraud is committed (preparatory acts or omissions) events and activities that take place at the time the fraud is committed (*actus reus*) and finally, events and activities that take place after the commission of the fraud (consequential acts).



A major task in fraud detection is constructing models of fraudulent behavior. This identifies key characteristics of fraud which can be used to prevent future frauds (online fraud detection) and for detection of past frauds (a posteriori fraud detection). It can also be used to identify ongoing fraud.

Fraud cases are notorious for their complexity. This complexity is usually caused by the complex nature and extent of the information involved. Said briefly, the law is often simple but the facts and evidence is often complex. Therefore, the ontology has to be able to manage and control the masses of data gathered during financial fraud investigations. This can aid the investigator to in focusing on relevant areas of law, the relevant facts in issue and the links and associations inherent in the evidence. Some associations between hypotheses, law and facts in a fraud model may be obvious, but others may be less so obvious. Simply modeling these relations is an important part of investigative methodology. Identifying what may be obvious and what may not be so obvious, lies at the heart of effective modeling and investigation. The way in which simply modeling and visualizing the relationships between the hypothesis under investigation, the law and the evidence, should not be under-estimated.

7.2.2 Top Level Requirements

The following top level strategic requirements have been identified:

- The user needs to be able to identify the species of fraud involved. This may be in terms of a legal definition but it will also be in terms of the methodology adopted to commit or even plan the fraud.
- The user needs to be able to identify and express a hypothesis. The hypothesis will be in the form of some tentative explanation, a theory that requires explanation or some exposition - this exposition may or may not be able to account for the law and facts presented. In brief, the user must be able to identify associations between legal rules, facts and explanations gathered during investigations.
- Pre-condition: to automate pattern searching to reveal previously unknown relationships.
- The user must be able to streamline and standardize data capture, storage and analysis.
- The user must be able to engage in the synthesis of probable and even possible models of fraud.
- The user must have access to an information infrastructure for investigations

- The user must be able to have access to information that is geographically-specific.
- The user must be able to have access temporal classifications and associations.
- The system must provide clear audit trails.
- The system must be sensitive to privacy and digital rights management.
- The system has to work interoperable: the system has to take into account the different regulatory requirements, i.e. it has to work cross-jurisdictional; across the UK, Belgium and Italy. A system that will be deployed across multiple jurisdictions, faces the fact that no two law enforcement agencies store their incident data in exactly the same way. Thus it is important to have a data organization design that is flexible enough to be applied to any underlying data set
- A system should use standard and non-standard querying techniques so that it can be used to identify standard patterns of fraud and non-standard patterns of fraud.
- The system has to incorporate knowledge from different domains.
- The user has to be able to share information amongst regulators in the EU.
- The system must be able to conduct querying on the basis of incomplete information.
- The system needs to be able to be interoperable between agencies.
- The system and users need to be able to use 'Red Flags' and 'Alerts' sensitive to fraud signals.
- The system must be sensitive to both fuzzy and linear associations. Associations between facts and law are often fuzzy rather than linear.
- The system has to look at multiple factors in a potential fraud case and select only those where it assumes a certain degree of likelihood of fraud for manual review.
- The system has to include a kind of electronic case management system; to store and work on new cases. This includes a case chart; interests harmed; estimated losses; target; geography.
- Investigators often need to be able to justify and document the manner in which they draw a conclusion. This is used in legal proceedings to justify subsequent actions. A search history should be designed to address this need.

- The General User Interface (architecture) should be both simple but adequate to achieve the requirements of the user.

Following points have to be taken into consideration :

- Confidentiality requirements;
- Privacy rights;
- Digital rights;
- Priority issues;
- Entities to be investigated;
- Periods to be covered in investigation;
- Authority to obtain information and access to premises and records;
- Identify key issues: consider: business activities; operating locations; trading record; management; audit reports; cash flow and financing;
- Decide on documentation to be seized: consider: evidential requirements;
- Range and location of documents to be seized;
- System scalability;
- GUI Ergonomics.

7.3 User Requirements for CONSOB Type Fraud²³

The main objective for CONSOB is a systematic and scalable 'web crawler procedure.' That is, closing the gap between inspection and enforcement in such a way that enables the user to detect more fraud with lower false positives.

CONSOB's current procedure,²⁴ which is fundamentally based on keyword-search, consists of using different Internet search engines (such as Altavista, Googly, Yahoo,

²³ See Annex 1 for a description of CONSOB's use case.

etc) as well as several meta-search engines. The search result is a list of web sites whose content is investigated by CONSOB's inspection officers in order to analyse and identify market abuse phenomena, abuse provisions of investment services and investment solicitation. The keywords are selected and combined to manually create complex queries on basis of the experience acquired during the ordinary supervision activity of the CONSOB's operative units. The use of the FF POIROT ontology in CONSOB's business case is related with the use of tools able to automate the query launching and to optimise the web information retrieval results. That is;

- Ability to examine possible fraudulent websites, and the links included on that website;
- System has to be attached with several subject-specific thesauri, databases of term phrases with respect to the specific crime of fraudulent on-line investment solicitation
- Finding suspected information on the World Wide Web: A web crawler (a proprietary search engine as opposed to general-purpose search engines) with a twofold search task: Which sites are selling securities AND which securities are being sold unlawful. Identification of traders in investment funds on the web. Identification of those traders who are not licensed. Prove that point above are engaged in trading in investment funds
- A match program to compare the found pages with the search intention to filter out the irrelevant pages (lexicographically)
- Data manager; is responsible for the management of search results
- A program to compare the search intention with web pages lexicographically to filter out irrelevant pages.
- Application of image processing technology in the search task.
- Facilities for organizing and managing search results should be provided
- Semantic analysis of the selected pages shall be done to identify the pages containing crime information. First by a natural language processor and, then, by human experts
- A central repository to store the relevant web sites after the semantic analysis. The system should keep a detailed history of the fraudulent website and changes to it to make the job easier for prosecutors (Collection of suspected information). A centralized database with a fraudster's name, method of operation, email

²⁴ See FF POIROT contract; Maria Vittoria's internal paper on Consob user requirements.

address, URL, screen names, or other pertinent data would serve as a national repository for these crimes and criminals. As financial crimes conducted on the World Wide Web are particularly difficult to solve, but investigators linked through such a system could connect clues from various jurisdictions.

- If a page is verified as containing crime information, it will be processed automatically to abstract new concept terms that are to be added to the database for supporting further search.

7.4 User Requirements for VAT Fraud

The main objective in fighting cross border VAT in the EU is to establish an effective system of mutual assistance and information exchange in order to ensure the proper functioning of the VAT system. A possible use case is an automatic and preferably a spontaneous exchange of information to help in the detection and prevention of fraud in intra-Community trade.

The system needs to enable two-way co-operation. This is in terms of maintaining and respecting legal authority of a EU Member State but at the same time co-operating with other national authority's of other EU Member States. The Ontology would provide a solid base for the monitoring and enforcement of noncompliance of VAT laws. This should include:

- data integration within the same agency
- data integration between different (national) agencies
- data integration between two or more EU Member State agencies

Fuzzy logic is central to the effective investigation and identification of tenuous and non-descript facts within a suspected or potential fraud. Distinguishing between probable, potential, accepted facts and frauds will enable and support better decision making and resource management.

The average life span of a VAT carousel fraud scheme is 4 months. After 4 months, the fraud organizer will make changes to the fraud scheme (adding companies, taking companies out, etc). Using fuzzy logic allows the system to adjust the profile dynamically as data are being analyzed. The output of the fuzzy logic system is

twofold. First, a degree of likelihood of fraud is assessed by the fuzzy logic system. A second output variable gives an indication why a certain invoice claim was considered to be possibly fraudulent by the fuzzy logic system.

- Active logic inference engine (VAT fraud → Customs & Excise have power of criminal investigation. While their counterpart in Belgium do not have this power. Only the police have the power to investigate in Belgium. You want to have the system to know this).
- Contextual computing (building in the ability to learn from data should enable systems to apply individual context to decision making)
- Recognise trends in VAT Fraud. Follow up developments in used techniques. Fraud control is played against opponents who think creatively, adapt continuously, and relish devising complex strategies. So a set of fraud controls that is perfectly satisfactory today may be of no use tomorrow, once the game has progressed a little. Maintaining effective fraud controls demands continuous assessment of emerging fraud trends and constant, rapid revision of trends.
- System has to be flexible: VAT fraudsters need only a few days or weeks at the most to change tactics once they find out a particular method is thwarted. E.g. new buffer companies. Because fraud control is dynamic and continuously evolving, a static set of filters has only short-term value.
- The investigator will analyse apparently random data such as invoices files to determine if some external agent (fraudster) is distorting the random nature of the data and leaving a noticeable pattern. Model has to make it feasible to distinct between simple irregularities and actual fraud.
- System needs a typology of VAT fraud.
- Multiple data sources (VIES, ICT listing, etc) are often used, each having different functions and user interfaces. This adds another dimension of difficulty for the end-user. One easy-to-use interface that integrates these different data sources is needed
- The system has to recognise all national rules.
- An ontology is needed to enable a correct and rapid analysis of the VAT regulations in different member states and to keep analytical activity up to date.
- This cross linking to the various national statutes is highly relevant for VAT officers because VAT fraud often has a cross border element. It helps the VAT

officer who does not understand the different languages concepts and their meaning.

- Data mining: to identify deliberate falsification of data (invoices) held within external database sources (VIES, etc).
- Graphical/illustrative presentations of key issues assist users in understanding them. In the case of the UK, visualization will not only assist users. During a trial, the evidence of a fraud investigation is inevitably complex and mountainous. Presenting this information in court to a lay jury is frequently seen as one of the biggest hurdles in any case. Applying simple, clear graphics to illustrate complex commercial data or financial transactions can help a jury to understand highly intricate cases.
- In fraud detection and investigation it is essential to develop a fraud risk profile in order to identify those areas that are vulnerable to fraud and to establish applicable and appropriate red flags.
- Presenting the data in a way that it is easily understood by the judge, client, etc.
- Running of multiple profiles. Indicators: rarely can fraudulent activity be detected through the use of a single profile. Similarly, the running of multiple profiles is normally time consuming and a drain on resources. For example, other profiles in VAT transactions may include post box service, ... As a result, the trend is increasingly towards fully automated systems that can repeatedly run all the known profiles.
- VAT: a lot are repetitive tasks. E.g. checks to ensure VAT numbers are valid. This should be automated so the investigator does not spend time running these tests.

7.5 Meeting the Requirements

When the requirements are met the resulting environment will be useful to the above mentioned target audience in a way that:

- Investigative and monitoring bodies will benefit from the strongly enriched information retrieval made possible by linking e.g. internet or database search facilities to the FF POIROT ontology in order to detect or investigate instances of attempted or actual financial fraud. Species of fraud (typologies) have been identified so that macro and micro analysis can be undertaken then used as

'templates of fraud'. These templates can be stored, accessed and used to mine for new frauds across linguistic and jurisdictional boundaries. In addition, they can use partial templates (bits of the model) to act as 'attractors' or 'magnets' which they can use to mine for data that might (when drawn together) amount to a fraud.

- Financial professionals will benefit from an "FF POIROT style" ontology using it as an authoritative concept base, extensively cross-linked (to other domains, systems and languages) and available for customized applications. Exploitation in this area could be as a high-tech service extending similar services and products (viz. on European VAT resp. accounting rules) currently already commercialized by at least two of FF POIROT's users.
- Law enforcement: benefit by the availability of relevant parts of the FF POIROT ontology e.g. as an RDF-mapped Semantic Web resource, to support *future police-oriented query systems*, in a non-technical user-friendly, attractive, and comprehensive manner. Additionally, sharing of information with investigative bodies and understanding of related documents will be substantially enhanced if such communication and documents are hyperlinked to a shared ontology. Optimizing the investigation, discovery, prevention and reduction of complex frauds is being made routine and efficient.

8. CONCLUSION

The requirements set out in this document will guide the development of the financial fraud ontology prototype. Whilst the initial prototype will be designed to fulfill the requirements expressed here, it will also be designed for flexibility to allow easy modification and iterations based on use cases, user feedback and user-testing results. As this is an *initial* analysis of user requirements there should always be scope for discussion on new necessities of the system.

Annex 1. Use Case Decomposition²⁵

The UML use case diagram shown in figure 1 below represents the CONSOB showcase in terms of actors, use cases and interactions amongst them.

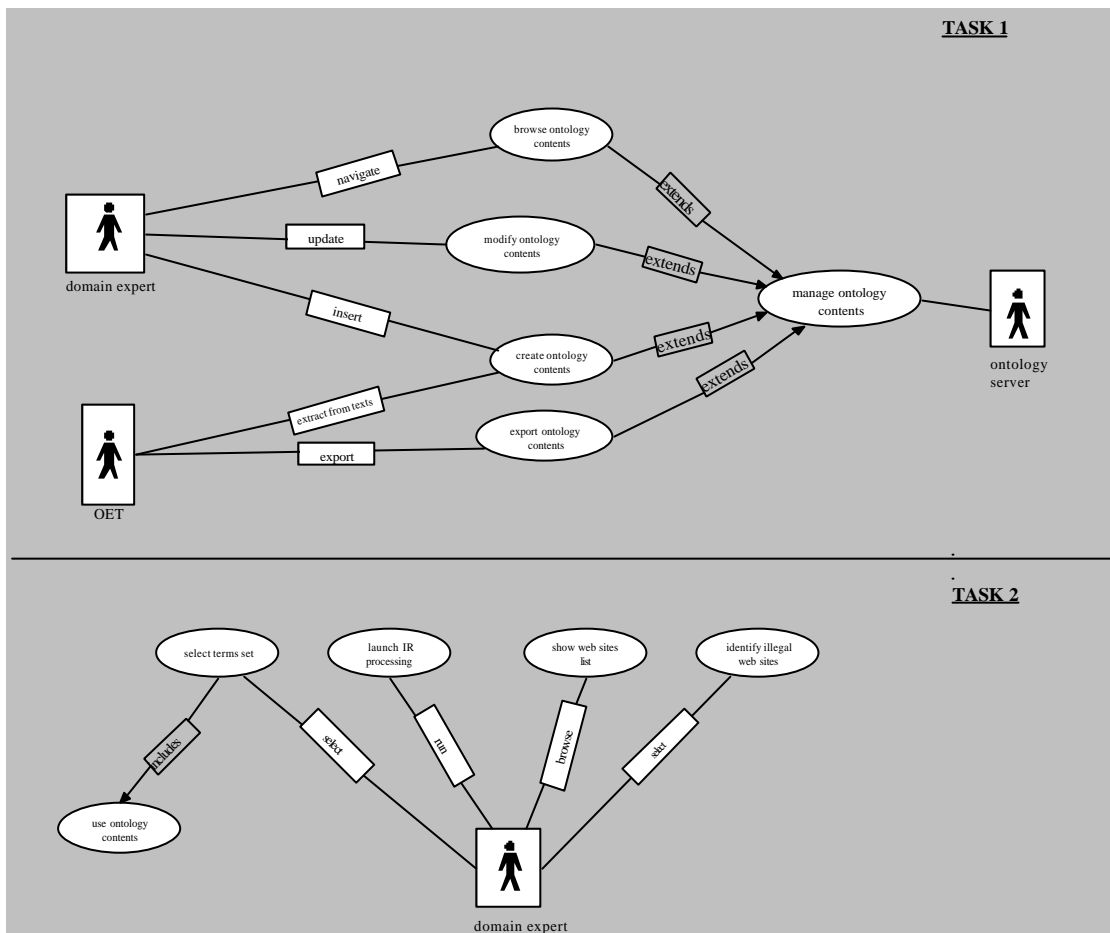


Figure 1: CONSOB use case diagram

In order to understand the considered diagram, a short description of the scenario relevant to the use case identified is given hereafter:

4.3.1 Task 1: Accessing and Editing the Ontology Contents

4.3.1.1 Manage Ontology Contents

²⁵ See document Maria Vittoria Marabello, Knowledge Stones S.p.A.

This use case occurs each time the user wants to access and manage the contents of the domain specific ontology, extracted from texts by the ontology extraction tool (OET), integrated and made available by the project's ontology server.

4.3.1.2 Browse Ontology Contents

This use case occurs when the user wants to navigate the ontology contents.

4.3.1.3 Create Ontology Contents

This use case occurs when the user wants to update the ontology contents by manually creating new elements.

4.3.1.4 Modify Ontology Contents

This use case occurs when the user wants to update the ontology contents by changing the available elements.

4.3.1.5 Export Ontology Contents

This use case occurs when an upload of the domain specific ontology contents into the project's ontology server is required.

4.3.2 Task 2: Supporting the Web Investigation Activity

4.3.2.1 Select Terms Set

This use case occurs each time the user wants to select a terms set to feed the information retrieval process, by means of which it is possible to identify financial frauds carried out through Web Sites.

4.3.2.2 Use Ontology Contents

This use case occurs when the user wants to select a terms set to feed the IR process. The terms set has to be part of the project's ontology validated contents.

4.3.2.3 Launch IR Processing

This use case occurs whenever the user wants to start the IR process, in order to select a group of sites potentially carrying out a financial fraud.

4.3.2.4 Show Web Sites List

This use case occurs each time the user wants to look at the list of sites produced by the IR process.

4.3.2.5 Identify Illegal Web Sites

This use case occurs each time the user, by looking at the sites list produced by the IR process, can carry out a Web Site inspection to identify the illegal services there proposed.