# Building Interoperability for European Civil Proceedings Online

## Research Conference - Bologna, 15-16 June 2012

## BUILDING SEMANTIC INTEROPERABILITY FOR EUROPEAN CIVIL PROCEEDINGS ONLINE

*Marta Poblet, Josep Suquet, Antoni Roig, Jorge González-Conejero*

**Institute of Law and Technology (IDT-UAB)**
**Barcelona, Spain**

**Project coordinator**

Research Institute on Judicial Systems, National Research Council of Italy (IRSIG-CNR)

**In partnership with**

Law Faculty, Central European University (CEU), Budapest

ICT Department, Italian Ministry of Justice, Rome

Institute of Law and Technology, (IDT) Autonomous University of Barcelona

Department of Political Science (DSP), University of Bologna

Interdepartmental Research Centre on European Law (CIRDE), University of Bologna

Centre for Social Studies (CES), University of Coimbra

**With the financial support of**

European Commission, Directorate General Justice Freedom and Security,

Specific Programme Civil Justice, Grant JLS/2009/JCIV/AG/0035

IRSiG-CNR
Istituto di Ricerca sui Sistemi Giudiziari
Consiglio Nazionale delle Ricerche

## Executive summary

This report provides an analysis and evaluation of semantic interoperability as regards the European Small Claims Procedure (ESCP) and the European Payment Order Procedure (EPO).

The practical implementation of the ESCP and EPO procedures by European applicants has highlighted some problems in their operation by national courts. E-justice national systems such as Money Claim in England and Wales, CITIUS in Portugal and COVL in Slovenia raises similar semantic issues: (i) claimants may find it difficult to express their will within a limited number of characters; (ii) courts are likely to interpret in legal terms what it was simply conveyed in plain, non-legal language, so that claimants may need legal advice to properly draft their claims; (iii) plaintiffs may also encounter other issues such as filling in certain details that may not be known (e.g., factual aspects such as the defendant's domicile or Postal Code).

The report provides a conceptual framework of semantic interoperability (SI). SI refers to the ability of computer systems to communicate information and have that information properly interpreted by the receiving system in the same sense as intended by the transmitting system. Therefore, SI can contribute to organise and clarify distributed knowledge regarding ESCP/EPO proceedings. In the shift from the current human-readable Web to the machine-readable Semantic Web the use of ontologies is of paramount importance.

The Institute of Law and Technology (IDT-UAB) has been involved in different projects in the field of semantic interoperability, including annotation and search and retrieval, where defining and implementing one or more ontologies was a core task (IURISERVICE, INTEGRA, ONTOMEDIA, and CONSUMEDIA projects).

According to the practical (analysis of ESCP/EPO semantic issues, e-justice national systems, former projects) and theoretical (conceptual framework) knowledge of SI, a toolbox for the EPO and ESCP proceedings is proposed:

• An ontology for the identification of the Court (EPO and ESCP). The ontology would automatically match the domicile and the court that has to deal with the case. This has some clear benefits from existing databases since it is easier to add, modify or reuse the links between domiciles and court jurisdictions.

• A Frequently Asked Questions (FAQ) repository. A FAQ can include expert knowledge and can be tailored to address practical issues. This could be particularly relevant as to the determination of the applicable Law. It could be a guide for the most usual cases, and could help general users and even lawyers from other countries.

• An ontology as regards the grounds for the court's Jurisdiction (ESCP Form A, number 4). The ontology may help users in order to correctly answer the ground of the court's jurisdiction. This may enrich and complete the glossary of the European Judicial Atlas by providing an ontology that would match different villages or towns with a particular court. Moreover, it would be advisable to formalise the expert knowledge of Private International Law professionals as regards rules on jurisdiction.

• An ontology for the determination of what the claim relates to (EPO's application form: number 6, principal). A natural language description could be transformed into legal concepts such as sales contract, contract of service or rental agreement.

• An ontology may be built with a list of most relevant legal concepts and its equivalent in different countries.

• A Semantic annotation of the structure of the documents. A further step may be to extend the XML annotation to the content of the documents.

# Building Semantic Interoperability for European Civil Proceedings Online

Marta Poblet, Josep Suquet, Antoni Roig, Jorge González-Conejero

Institute of Law and Technology (IDT-UAB)

**CONTENTS**



## 1. Introduction. Semantic interoperability issues for the European Small Claims Procedure (ESCP) and European Order for Payment procedure (EPO)

The European Institutions are committed with the objective enshrined by the EU Treaty of maintaining and developing an area of freedom, security and justice that ensures the free movement of persons. To fulfil this objective, those institutions have adopted different legal instruments in the field of judicial cooperation in civil matters having cross-border implications.[1] The Regulation (EC) 861/2007 of the European Parliament and of the Council of 11 July 2007 establishing a European Small Claims Procedure (hereinafter, ESCP) and the Regulation (EC) No 1896/2006 of the European Parliament and of the Council of 12 December 2006 creating a European order for payment procedure (hereinafter, EPO) are two specific measures aiming at eliminating obstacles to the good functioning of civil proceedings.

The practical implementation of those procedures by European applicants has highlighted some problems in their operation by national courts. Indeed, their functioning requires not a European court but the cooperation among national tribunals. Moreover, the

[1] See article 67 and ss. of the Treaty on the Functioning of the European Union. Consolidated version available at http://eur-lex.europa.eu/en/treaties/index.htm

plaintiff and the defender do not have a "direct dialogue"; there are no horizontal mechanisms of cooperation between the parties, but just vertical ones between the Courts and the parties (Ng, G. Y. and Mellone, M. 2011). While both Regulations provide standard forms to support dialogue and avoid misunderstandings, the daily practice reveals that they are not clear enough and both parties and courts may be confused as to how to deal with them.

Article 6 of the ESCP provides that the claimant may need to translate the documents to other languages, such as the language of the defendant or the language of the court seized. In the judicial cooperation area, where different languages and different harmonised legal systems coexist, we face a number of semantic issues. Embedded in different legal traditions, legal terms remain largely a matter of a particular national system. An Italian court may assign a different meaning to what an English claimant has expressed in his claimant form. The European Legal Atlas provides online automatic translation of the application forms, but there is no translation available for the claimant's description of the nature of the issue and the object of the claim.[2] Precisely, Ng and Mellone have recently showed some of those semantic issues in a simulation for a trans-border small claim (ESCP) between the United Kingdom and Italy (Ng and Mellone 2012). Indeed, the translation of the facts has been reported as a possible problem. Likewise, the same applies to attached documents such as invoices, contracts, etc. The option to appoint human translators to fill these significant gaps is exceedingly costly, and yet, state-of-the-art automatic translators do not appear to be accurate enough. In this regard, this paper will later show that we can use XML to annotate the items of a structured text (e.g., in FORM A or FORM B) but not the content therein. XML will not be useful enough when we have to translate a list of facts. Therefore, in this case, perhaps a human translator may be required.

The *In vitro* Report has revealed further problems such as the determination of the addresses of the competent courts, or the individualisation of the correct jurisdiction. The latter issue is addressed within the European Judicial Atlas in Civil Matters, whose website allows an applicant to determine a competent court. [3] In contrast, the former aspect may be more difficult to resolve if the party is not able to provide the correct address.

In the pages that follow we will show some tools that may contribute to solve some of these issues. Ontologies play a relevant role here. However, there are specific problems which

[2] See Part 8 of the Claim form- Annex 1.
[3] http://ec.europa.eu/justice_home/judicialatlascivil/html/sc_courtsjurisd_en.jsp?countrySession=3&txtPostalCode=08008&txtMunicipality=#statePage1

semantic tools may not properly address. Again, the *In Vitro* report shows how the practical issue to determine a legal interest poses difficulties for applicants.

Ontologies have proved useful in modeling legal knowledge, but may fall short of adequately representing complex legal or judicial decision-making as a standalone solution. To address such a task, the use of legal ontologies needs to be combined with sequential modeling of the different steps of the process. Once the successive nodes are identified and represented, then semantic tools can enrich the decision-making model. The report refers to the preliminary assessment to determine whether the court is competent, or whether the court is manifestly unfounded, we must rectify or integrate the claim, or we must immediately transmit it to the defender. As a global decision process, it is clearly too wide to be solved by an ontology. The only way to try to manage such a complex process is to transform it into a succession of more specific problems to be dealt with. We ought to determine clearly the classification and the criteria of the decision to make. Only then we have a chance to build up a semantic tool that is worthwhile.

A number of European countries have already developed national e-justice systems. Notably, Money Claim Online in England and Wales[4], CITIUS in Portugal[5] or COVL in Slovenia[6] aim at the same core objectives: to speed up a judicial process, to decrease pending cases and therefore to reduce the judicial backlog. Moreover, automated processes contribute to reducing costs and reassigning resources to other types of requests. However, similarly to some cross-border systems, e-justice national systems may also raise some semantic issues: (i) claimants may find it difficult to express their will within a limited number of characters; (ii) courts are likely to interpret in legal terms what it was simply conveyed in plain, non-legal language, so that claimants may need legal advise to properly draft their claims; (iii) plaintiffs may also encounter other issues such as filling in certain details that may not be known. (e.g., factual aspects such as the defendant' domicile or Postal Code).

In the following pages we will assess how semantic interoperability can contribute to organise and clarify distributed knowledge regarding ESCP/EPO proceedings. Therefore, the report is organized as follows: Section 2 focuses on the background of Semantic Web Technologies; Section 3 addresses the Semantic Interoperability issue; Section 4 provides an overview on ontologies, including features and capabilities. Section 5 discusses the suitability

[4] https://www.moneyclaim.gov.uk/web/mcol/welcome
[5] http://www.dgpj.mj.pt/sections/english-version/legislative-policy/annexes/legal-projects/citius-dematerialization/citius-electronic4155/
[6] https://covl.sodisce.si/

of the semantic interoperability toolbox which is proposed to address some of the EPO and ESCP semantic issues. Finally, Section 6 points out some conclusions.

## 2. Background: Semantic web technologies.

In the Scientific American foundational article of 2001, BernersLee, Hendler and Lassila offered their vision of the future Semantic Web as "not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation." (Berners-Lee, Hendler and Lassila, 2001). . The Semantic Web has come a long way from there, and even if the vision it is not yet in full display, state-of-the-art Semantic Web Technologies and languages offer today a new approach to managing information and processes, the fundamental principle of which is the creation and use of semantic metadata (Warren, P., Studer, P. and Davies, J. 2005). Since metadata tell us about the content of a document we may say that metadata are semantic tags that help to organize and find information based on meaning, not just text. By applying semantics our systems can understand where words or phrases are equivalent, or they can distinguish where the same word is used with different meanings. Moreover, semantics may improve the way information is presented and, instead of a search providing a linear list of results, the results can be clustered by meaning. In a typical pre-web 2.0, people would perform legal searches based on *keywords or* would make up a concept believed to convey the core meaning of what is looking for. There are also more complex searches such as the “Boolean searches” where several keywords are combined with Boolean operators (AND, OR, etc.). Certain databases allow the definition of several aspects of the search (e.g., date, type of court etc). However, these searches do not offer solutions or help towards the interaction between symbols, terms and concepts. Here is where the Semantic Web may be of use.

By applying metadata, semantics contribute to merge information in a meaningful way, removing redundancy, and summarizing where appropriate (Warren, P., Studer, P. and Davies, J. 2005). The use of semantic metadata enhances the storage, search and retrieval of information together with human- computer interaction. In this perspective, the semantic web is a prolongation of the web 2.0 enriched with meaning.

The World Wide Web consortium (WW3) has been developing interoperable technologies such as specifications, guidelines, software and tools to fully develop the promise of the Semantic Web.[7] Berners-Lee famous semantic web stack represents this growing level of

[7] http://www.w3.org/standards/semanticweb/

complexity (more complex at the top) as higher layers depend on lower layers. This overall idea was to construct something (semantic web) from the current work (web) so the work done before was still of use.

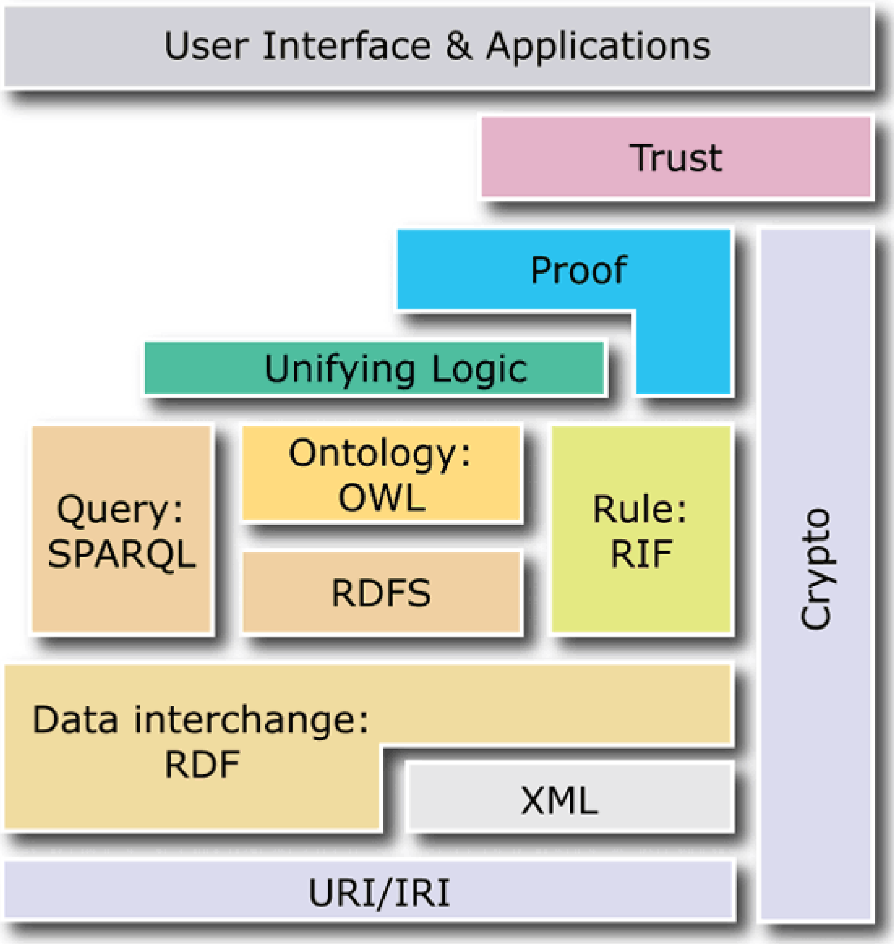



Fig 1. Semantic Web layer cake (T. Berners-Lee), S. Bratt version (2007). Source: W3C http://www.w3.org/2007/03/layerCake.png

## 3. **Semantic interoperability**

Semantic Interoperability addresses the issue of knowledge representation. Alongside with other interoperability concerns, such as organizational interoperability and technical interoperability, Semantic Interoperability (SI) (also, Computable Semantic Interoperability) refers to the ability of computer systems to communicate information and have that information properly interpreted by the receiving system in the same sense as intended by the transmitting system. As Halshofer and Neuhold have recently put it, “the interoperability problem and the representation of semantics have been an active research topic for approximately four decades” (Halshofer and Neuhold 2011: 3). Figure 2 below shows the evolutionary path followed by research on semantics and interoperability from the early database models to the recent developments on Linked Data:

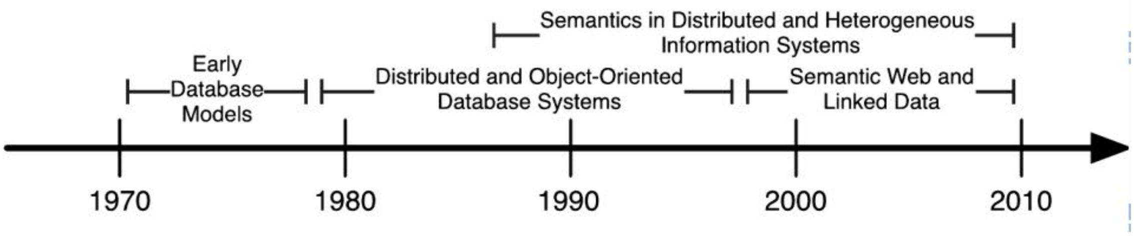



Fig. 2: Semantics and Interoperablity research in Computer Science (Halshofer and Neuhold 2011)

At the EU level, The European interoperability Framework for pan-European e-Government Services establishes that Semantic Interoperability “is concerned with ensuring that the precise meaning of exchanged information is understandable by any other application that was not initially developed for this purpose. Semantic interoperability enables systems to combine received information with other information resources and to process it in a meaningful manner” (European Commission. 2004). SI requires that any two systems will derive the same inferences from the same information. Moreover, the core objective of SI is “not only to allow information resources to be linked up but also to allow information to be automatically understandable, and, consequently, reusable by computer applications that were not involved in its creation” (European Commission. 2004). A further distinction deals with semantic interoperability versus syntactic interoperability. The former is being understood as the meaning of data elements and the relationship between them (including vocabularies to describe data exchanges, and ensuring that data elements are understood in the same way by communicating parties). The latter is understood focusing on the exact format of the information to be exchanged in terms of grammar, format and schemas.

The European Commission has devoted a sustained effort on Semantic Interoperability policies. The Pillar II of the Digital Agenda for Europe (2010-2020) deals with Interoperability and Standards. In this regard, the European Commission recognized in 2010 that action on interoperability is essential to maximise the social and economic potential of information and communication technologies (ICT). Further on, it establishes that Semantic interoperability is jeopardised by different interpretations of the information exchanged between people, applications and administrations. Semantic Interoperability, as well as interoperability at legal, organisational, and technical level "should progressively lead to the creation of a sustainable ecosystem (...) which would facilitate the effective and efficient creation of new European public services" (European Commission. 2010).

The Semantic Interoperability Centre Europe (semic.eu)[8] is a participatory platform and a service by the European Commission that supports the sharing of assets of interoperability to be used in public administration and eGovernment. Moreover, The Interoperability Solutions for European Public Administrations Programme (ISA Programme 2010-2015)[9] – addresses this need by facilitating efficient and effective cross-border electronic collaboration between European public administrations. From the private sector, the activity of the Open Group is also devoted to Semantic Interoperability.[10]

## 4. Information management and Ontologies.

### 4.1. Introduction

Information management has undergone a dramatic transformation in the last decade. Moreover, the Web has become the most important channel to share multimedia contents with the whole world: music, film, television, newspapers or books have been reshaped or redefined in the digital era. Web 2.0 tools and mobile technologies have lowered the barriers not just for people to access the Internet but to create and share content (Poblet and Casanovas 2012). In the social media context, "mash up", "like", "follow", or "tweet" are tinged with new, widely adopted meanings. The legal domain and its huge masses of textual and multimedia contents do not remain aside from this movement.

Indeed, in the World Wide Web, there is a growing amount of unstructured legal information directly available to anyone. This is why there is an urgent need to construct conceptual structures for knowledge representation to share and manage intelligently all this information, whilst making human-machine communication and understanding possible (Casellas Queralt, N. 2008). As regards the legal information domain, the production of legal rules has followed an inflationary path. Today, the main problems are handling the complexity

[8] http://www.semic.eu/semic
[9] http://ec.europa.eu/isa/index_en.htm
[10] http://www.opengroup.org

and types of legal knowledge, and having reasonable ways to store, retrieve and structure a great amount of legal information (Benjamins *et al*. 2005).

Broadly speaking, interoperability is the ability of two or more systems or components to exchange information and to use the information that has been exchanged. It is clear that several aspects related to this topic have to be considered in a previous stage of the exchange information process. When a system is sending information, the receiver must know which type of information is receiving to allow a correct interpretation. If the information is not interpreted correctly, it becomes useless. The Semantic Web has an important application in this field. They could provide the abstraction layer needed to carry out a "negotiation" or "dialog" between the participant systems to put in common concepts, vocabulary, terms, etc. Therefore, all the participants will know the meaning (not necessarily the content) of the exchanged information. Consequently, Semantic Interoperability (SI) is able to meet requirements posed by interoperability affecting the European Payment Order (EPO) and the European Small Claims Procedure (ESCP).

An ontology (Gruber 1993) describes the concepts and relationships that are important in a particular domain, providing a vocabulary for that domain as well as a computerized specification of the meaning of terms used in the vocabulary. These applications include natural language translations, medicine, standardization of product knowledge, and electronic commerce, among others. Ontologies are specific tools to organize and provide a useful description of heterogeneous content. For humans, ontologies enable better access to information and promote shared understanding. For computers, ontologies facilitate comprehension of information and more extensive processing. In addition, there are many tools and applications to facilitate ontology management.[11] For instance, Protégé[12] is a design tool which is specifically devised to develop ontologies from many kinds of fields; and reasoning algorithms such as Pellet (Sirin, E. and others. 2005) provide reasoning capabilities.

In our scenario, we have to consider two main features that ontologies provide: i) expert knowledge acquisition; and ii) resilience to changes. The first feature is related to designers since technical skills are not necessary to develop an efficient model. So, experts in a concrete field will be able to interact directly with this tool. The second feature is related to changes that could affect ontologies over time. The world is continuously changing and

[11] See, i.e., Suárez-Figueroa *et al*. (2011) for an updated overview on languages, methodologies, and tools.
[12] Stanford University, Protégé, available on-line at: http://protege.standford.edu.

ontologies, by definition, are easily adaptable tools and produce minor repercussion to the rest of the system. There are many desired extra features pointed out in Section 4.

In this report, we propose a Semantic Interoperability toolbox to deal with the legal SI issue that is concerning EPO and ESCP. This framework is composed of three different parts: i) ontologies (knowledge representation); ii) Protégé (design tool); and iii) a reasoning algorithms (provide reasoning capabilities to ontologies). The parts of this framework are described in next sections in this report.

### 4.2. Definitions.

The term ontology has been borrowed from philosophy to be used in computer science and artificial intelligence in a technical sense. Nevertheless, there are many definitions of ontology in the computer sciences and AI domains and such definitions have changed and evolved over the years. In 1991, Neches et al. defined ontologies as a "top-level declarative abstraction hierarchies represented with enough information to lay down the ground rules for modelling a domain. An ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary". Moreover, one of the most well known definition of the AI ontology is "a explicit specification of a conceptualization" (Gruber 1993b,a). A more complex definition establishes that "an ontology is a formal, explicit specification of a shared conceptualization. Conceptualization refers to an abstract model of some phenomenon in the world by having identified the relevant concepts of that phenomenon. Explicit means that the type of concepts used, and the constraints on their use are explicitly defined. Formal refers to the fact that the ontology should be machine-readable. Shared reflects the notion that an ontology captures consensual knowledge, that is, it is not private of some individual, but accepted by a group". (Studer et al, 1998). One of the goals of ontologies is the construction of a catalogue of categories that exist in reality, connected with the classification and organisation of knowledge.

The four characteristics present in a general ontology are: 1) "*conceptualization*" which refers to an abstract model of some phenomenon in the world, which identifies the relevant concepts used; 2) "*explicit*" which means that the type of concepts used and constraints on their use are explicitly defined; 3) "*formal*" which refers to the fact that the ontology should be machine readable; and 4) "*shared*" which reflects the notion that an ontology captures consensual knowledge.

Ontologies can be classified, according to the issue of the conceptualization into (Guarino, N. 1998):

• Representation ontologies or meta-ontologies. They capture the representation primitives used to formalize knowledge in a given knowledge representation.

• General or upper-level ontologies. They classify the different categories of entities existing in the world. Very general notions which are independent of a particular problem or domain are represented in these ontologies. The knowledge acquired is applicable across domains and includes vocabulary related to things, events, time and space.

• Domain ontologies. They are more specific ontologies. Knowledge represented is specific to a particular domain. They provide vocabularies about concepts in a domain and their relationships, or about the theories governing the domain.

• Application ontologies. They describe knowledge pieces depending both on a particular domain and task.

Further to ontologies, taxonomies represent a classification of the standardised terminology for all required and involved terms within a knowledge domain. In a taxonomy, all elements are grouped and categorised strictly hierarchical and are usually presented by a tree structure. In a taxonomy, the individual elements are required to reside in the same semantic scope, therefore all elements are semantically related with each other to a certain degree. [13]

### 4.3. Ontology design tools: Protégé

Protégé is specifically devised to develop ontologies from many kinds of fields. Due to its features and properties, it is ideal to acquire and manage knowledge in our scenarios. On the other hand, Pellet is an open source "reasoner" that also provides high features to manage knowledge.

There are several ontology languages available in the literature (like OWL or WSMO family) with different expressiveness and reasoning capabilities.[14] The main criteria for the selection of an ontology language are its knowledge representation mechanism and the inference support needed by an application. The high complexity required by the knowledge modelling requires a representation language with high semantic expressiveness. OWL

[13] See the Semic glossary at: http://www.semic.eu/semic/view/smeta/Glossary.xhtml
[14] See http://www.w3.org/standards/semanticweb/

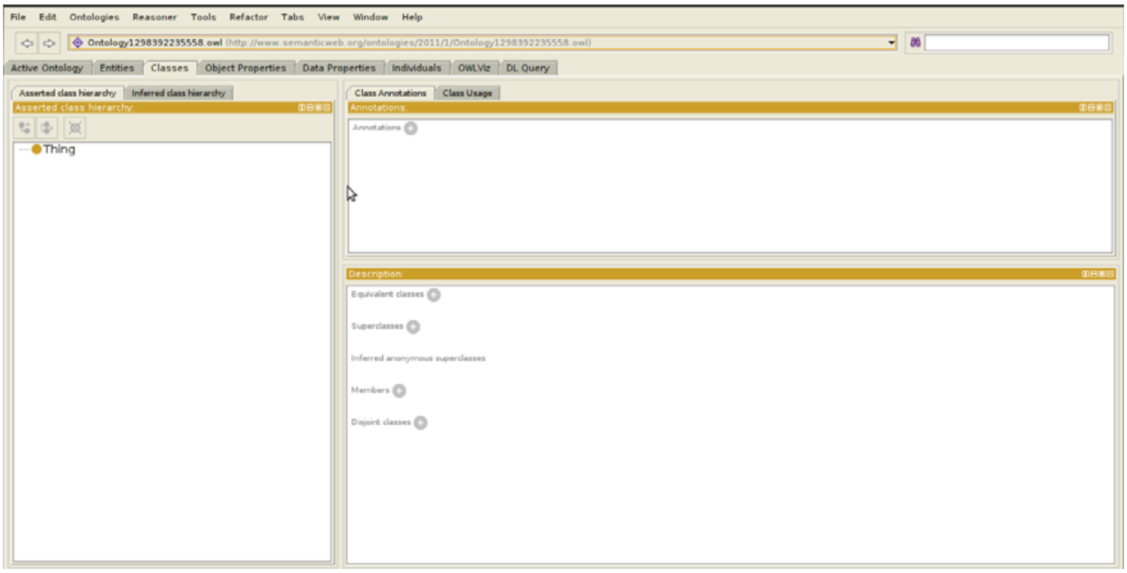
combines the required expressiveness for ontologies and the compliance to W3C standards, which turns it as the most appropriate language.

Protégé is a suite of tools for ontology development and use developed at Stanford University.[15] It is the main framework used in the Institute of Law and Technology at the Universitat Autònoma de Barcelona (IDT-UAB)[16] in those projects implementing ontologies. Its main features are: 1) it is a free, open source platform that provides a suite of tools to construct domain models and knowledge-based applications with ontologies; 2) it also implements a rich set of knowledge-modelling structures and actions that support the creation, visualization, and manipulation of ontologies in various representation formats; 3) the framework supports two main ways of modelling ontologies via the Frames and OWL editors; and 4) ontologies can be exported into a variety of formats including RDF(S), OWL, and XML Schema.

Protégé is extensible, based on Java and provides a plug-and-play environment that makes it a flexible base for rapid prototyping and application development. In the Protégé knowledge model, terminologies and ontologies are represented using "*frames*" (classes, slots and facets). *Classes* are the entities and sometimes named "concepts" in terminologies. *Slots* describe properties or attributes of classes. *Facets* describe characteristics of slots. An ontology in Protégé consists of frames and axioms. Axioms specify additional constraints. An instance is a frame built from at least one class that carries particular values for the slots. A "knowledge base" includes the ontology (classes, slots, facets and axioms) as well as instances of particular classes with specific values for slots.

Figure 3. Protegé Framework in a GNU/Linux Environment.

[15] http://protege.stanford.edu.
[16] http://idt.uab.es




### 4.4. Ontology applications.

Ontologies are of a key importance in order to promote interoperability services. The Institute of Law and Technology (IDT-UAB) has been involved in different projects in the field of semantic interoperability, including annotation and search and retrieval, where defining and implementing one or more ontologies was a core task in all these projects. This has been the case with the IURISERVICE, INTEGRA, ONTOMEDIA, and CONSUMEDIA projects that are featured below.

#### *i) Interoperability*

The INTEGRA project (Research on Technologies for Decision Making in Immigration Policies) aims at developing intelligent systems to manage migration flows in both regulated and non-regulated EU borders, with a global perspective of the problem and an approach to a European solution.[17] The growing differences in the development of the first and third world have caused that migratory movements have multiplied exponentially over the last decade. Migration is one of the most important challenges that face developed countries because of the effect it has on the demographic structure of both sending and receiving countries. The European Union is composed by many different countries with different languages. Furthermore, each country has different legislations and documentation to regulate migration (Poblet and Vallbé, 2010). Ontologies could provide an abstract layer to represent the knowledge acquired from

[17] INTEGRA (Research on Technologies for Decision Making in Immigration Policies); CDTI-Spanish Ministry of Industry, Tourism, and Commerce, Contract 15/02/2008

legislations and documentation from each country, making the exchange of information possible. Indeed, the INTEGRA Project is a good example of interoperability between different databases:

The first database taken into account within the INTEGRA Project is the Schengen Information System of second generation (SIS II)[18]. SIS II has several elements: One main system (SIS II core); one national system («N. SIS II») in each Member State (the national data Systems that communicate with the central SIS II); One Communications infrastructure between the central system and the national systems that provide a SIS II network and the date share between the national services responsible (SIRENE services). Furthermore, the Visa Information System (VIS) was also taken into account in the project. The VIS system is useful for the fulfilment of the common visa policy, the Consulates cooperation and the requirements of the National institutions responsible for the Visa[19]. Moreover, the EURODAC database helps in the management of the requests for Asylum. It is possible to compare the different fingerprints for the proper implementation of the Dublin Convention[20]. Member States can verify whether a person that solicits Asylum in one country has already done the requirement in another member State. There is a central unity coordinated by the European Commission, a central database, and Electronic communication devices between member States and the core institution. The fourth database taken into account was TECS. This is an information system provided by the "Europol Convention"[21]. The Europol Information System (TECS) has three main elements: One indexing system; One information system: Europol Information System (EIS); One analysis System: Europol Analysis System (OASIS).

The main interest of those database structures and systems of coordination is the fact that the central structure does not substitute the National ones, but is added to them. This is

---

[18] Decision 2007/533/JAI of the Council, June 12th 2007, available at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32007D0533:ES:NOT , Regulation (CE) nº 1987/2006 of the European Parliament and of the council, of December 20th 2006, available at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32006R1987:ES:NOT

[19] Regulation (CE) nº 767/2008 of the European Parliament and the Council, of July 9th 2008, on the Visa Information System (VIS) and the short term visa data sharing between member States (Regulation VIS). http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32008R0767:ES:NOT , and Decision of the Commission 2009/377/CE, May 5th 2009. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32009D0377:ES:NOT

[20] Council Regulation (EC) No 2725/2000 of 11 December 2000 concerning the establishment of EURODAC for the comparison of fingerprints for the effective application of the Dublin Convention. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32000R2725:EN:HTML

[21] Available at http://www.europol.europa.eu/legal/Europol_Convention_Consolidated_version.pdf .

perhaps similar to what can be done with EPO and EPSC, where no central procedure is considered, but rather a network and a coordination of National procedures.

The INTEGRA project produces two ontologies aimed at managing interoperability for countries within the Schengen treaty. The first one is focused on document matching to provide interoperability among different kind of documents that are used in the European Union to identify people (e.g. passports, ID cards, etc.). These documents are usually issued in the language of its country. In this situation, the ontology provides a thesaurus that is able to help border agents to identify the document that they are handling. The second one is specifically devised to provide interoperability among different Law Enforcement Agencies within Europe. The main target of this ontology is to identify the permissions granted to agents in function of their country and position within the Law Enforcement Agencies.

### ii) Annotation

Files carry a meaning which can be very versatile. For a human, the meaning of the message is immediate, but for a computer that is far from true. This discrepancy is commonly referred to as the *semantic gap*. Semantic annotation is the process of automatically detecting the presence of a concept in a file. Therefore, the annotation process aims at expressing the semantics of information, improving information seeking, retrieval, classification, understanding and use. With the emergence of the Semantic Web, ontology based document annotation has been the focus of many projects and applications, since the availability of annotated content is one of the key challenges to overcome in order to make the Semantic Web a reality. The ONTOMEDIA[22] projects **are** another example of the application of ontologies to annotate digital documents.

The ONTOMEDIA projects aim at developing and ODR web platform for users and professionals to meet in a community driven portal where contents are provided by users and annotated by the platform. The ODR Web platform is tailored in the Business-to-Consumer (B2C) domain although later on it may be extended to other domains such as family, health care, labour, environment, etc. (Poblet *et al.* 2009a). Citizens (both professionals and users of

[22] ONTOMEDIA (Semantic Web, Ontologies and Online Dispute Resolution); Spanish Ministry of Science and Innovation, CSO2008-005536/SOCI); ONTOMEDIA (Ontologies and web Services for Online Mediation); Spanish Ministry of Industry, Tourism and Commerce) TSI-020501-2008-131.

mediation services) can use any kind of devices to access the portal (computers, smart phones), and in any format suitable to their purposes (text, speech, video, pictures). Ontologies are used to annotate all kind of contents and also to help to analyze multimedia content (Poblet *et al*. 2009b). The multimedia analysis is devoted to enhancing the information a mediator possesses during a mediation session, capturing mood changes of the parties and any other psychological information inputs that can be useful for mediators, just as if they were in a room with the users of the mediation service. All types of metadata will be automatically extracted and stored to be further used within the mediation process. ONTOMEDIA also foresees the application of mediation services as tasks within a mediation process that will be formally described by means of both process ontologies and mediation ontologies.

***iii) Search and Retrieval.***

In the Information Era, the amount of digital documents stored by enterprises and people has been multiplied exponentially. In this scenario, the search and retrieval of this information has become an important challenge. Usually, the search by a keyword or a concrete string is not efficient due to the heterogeneous origin of the documents. In addition, the relevance of a document could be determined by the context and not only by the keyword or the string which performs the search. The use of ontologies to overcome the limitations of keyword-based search has been put forward as one of the motivations of the Semantic Web since its emergence in the late 90's.

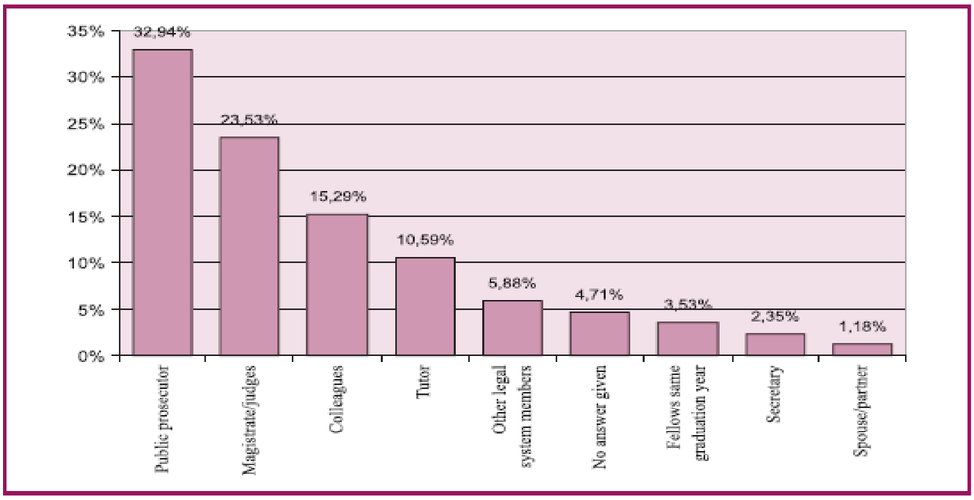
The IURISERVICE application was designed to provide Spanish judges in their first appointment with online access to an intelligent Frequently Asked Questions system (iFAQ), consisting of a repository of practical questions (problems that newly recruited judges were likely to face) with their corresponding answers.[23] The aim of the system was to discover the

[23] IURISERVICE has been developed within the framework of different national and international research projects: SEKT (Semantically Enabled Knowledge Technologies); UE (VI Framework Program, Information Society Technologies); EU-IST 2003-506826; IURISERVICE I (Design of an online network to support newly recruited judges); Spanish Ministry of Science and Technology, FIT-150500-2002-562; IURISERVICE II (Design of an online network to support newly recruited judges); Spanish Ministry of Science and Technology; FIT-150500-2003-198.

best semantic match between the users' input question in natural language and the stored questions. The search engine was enhanced with a legal ontology: the Ontology of Professional Judicial Knowledge (OPJK) and semantic distance calculation. The initial set of practical questions from newly recruited judges had been extracted from previous interviews with incoming judges as part of an extended fieldwork in Spanish Courts. The answers to these questions were left to senior judges from the Spanish School of the Judiciary (Casanovas, P. and others. 2005). Eventually, these pairs of questions and answers composed the initial repository of the system.

Figure 4 shows who do judges discuss their cases with.

**Figure 2** Who do judges discuss their cases with?




Scondly, this list of questions provided the input knowledge for the OPJK ontology, which ought to represent the relevant concepts related to the problems that take place during the on-call period, the knowledge contained in the list of questions. Therefore, the conceptualization process of the Ontology of Professional Judicial Knowledge was based on the previous and careful knowledge acquisition stage. This comprehended the acquisition of the list of questions and the treatment of this corpus in order to obtain relevant terminology related to practical problems faced by judges in their first appointment, through term extraction from the corpus of questions faced by judges.

Figure 5. Overall architecture for IurisService FAQ

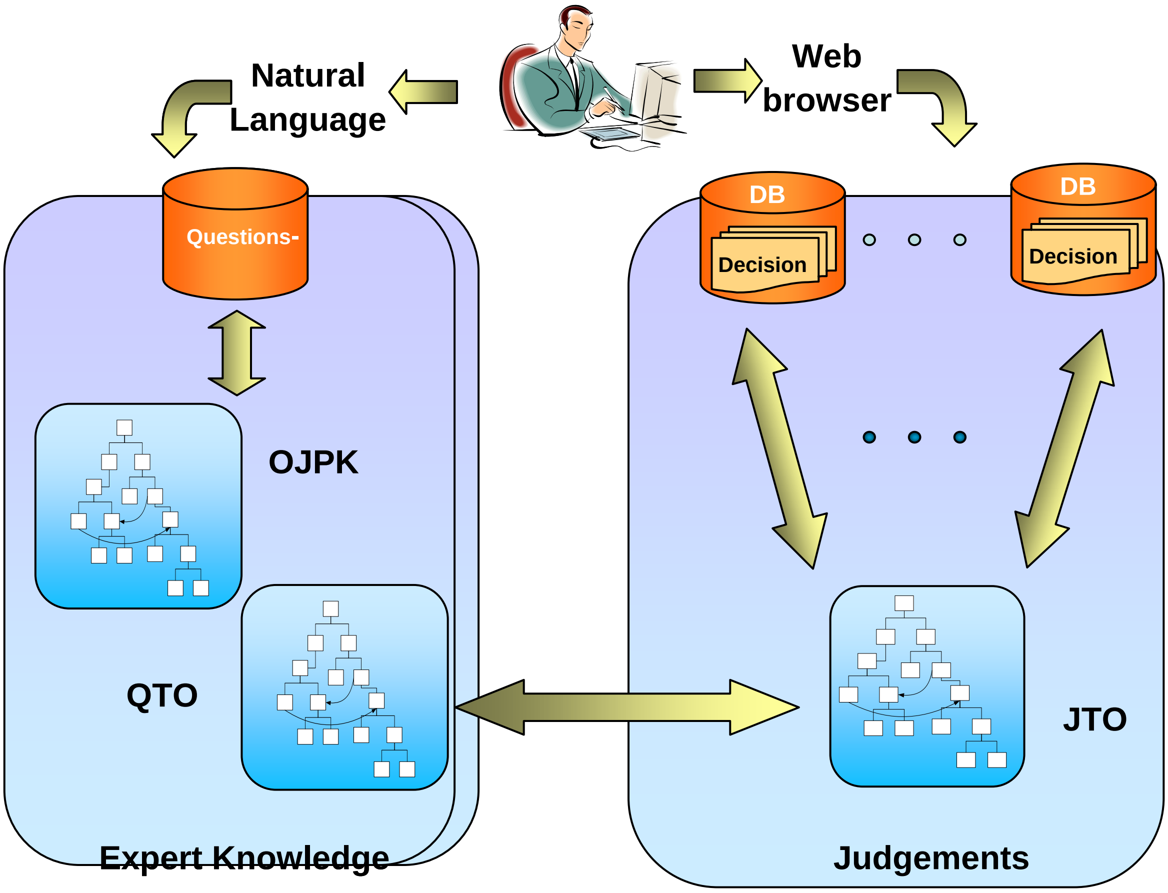
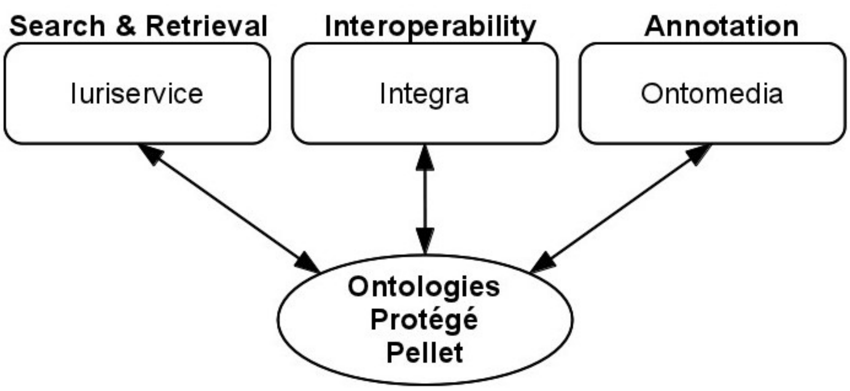



Figure 6. Ontological applications in IDT research projects




### 4.5. Ontology population.

The manual performance of ontology development and population is labour and cost-intensive. If population of ontologies has to be done manually by humans it cannot be taken the most out of ontologies. Therefore, it is of paramount importance to develop a maximum level of automation for those tasks. For this purpose, the identification and extraction of terms that play an important role in the domain under consideration, is a vital first step (Maynard, D., Li, Y. and Peters, W. 2008). Semi automatic knowledge acquisition has relied on the advancement of Natural Language Processing techniques (NLP). This is a field of computer science and linguistics concerned with the interactions between computers and human –natural- languages and it is aimed at identifying the relevant terms of a corpus. They are based on linguistic information, statistical methods, or on hybrid approaches (Fernández-Barrera, M. 2011).

Automatic term recognition (also known as term extraction) is a crucial component of many knowledge-based applications such as automatic indexing, knowledge discovery, terminology mining and monitoring, knowledge management and so on. Term recognition has been performed on the basis of various criteria. The main distinction we can make is between algorithms that only take the distributional properties of terms into account, such as frequency and extraction techniques that use the contextual information associated with terms (Maynard, D., Li, Y. and Peters, W. 2008).

Ontology population is a crucial part of knowledge-based construction and maintenance that enables us to relate text to ontologies, providing on the one hand a customised ontology related to the data and domain with which we are concerned, and on the other hand a richer ontology which can be used for a variety of semantic web-related tasks such as knowledge management, information retrieval, question answering, semantic desktop applications, and so on. Ontology population is generally performed by means of some kind of ontology based information extraction (OBIE). This consists of identifying the key terms in the text (such as named entities and technical terms) and then relating them to concepts in the ontology. Typically, the core information extraction is carried out by linguistic pre-processing (tokenisation, POS tagging etc.), followed by a named entity recognition component, such as a gazetteer and rule-based grammar or machine learning techniques (Maynard, D., Li, Y. and Peters, W. 2008).

## 5. Semantic interoperability toolbox for the European Payment Order and European Small Claims Procedure

### 5.1. FAQs

This report has addressed ontologies as one of the most appropriate tools when it comes to address and solve different semantic issues such as the existence of different legal terms and languages. As they are also hard to design and maintain, they can be combined with other tools. One particular aspect which we assess that it could be of help in EPO/ESCP is a FAQ (Frequently Asked Questions) list of terms. In this regard, the IURISERVICE project led by The Institute of Law and Technology (IDT-UAB) and reviewed above is an example of FAQs and both coordination and management of decentralised expert knowledge. As expressed before, the aim of the project was to create a network to help young judges at their first destination. Similarly to this project, an intelligent tutorship system, based on questions and practical answers could be set up. The goal of the FAQ system in the EPO/ESCP would be to share the professional experience of both judges and lawyers. Moreover, parties could also report and look for similar experiences. A professional Frequently Asked Questions (FAQs) based on expert knowledge is one of the user-friendly alternatives of sharing information. The potential users of this information system are both citizens not experts in law and lawyers alike. The FAQs can be thought in a multiple direction, covering the most common problems a user of EPO or EPSC can encounter during the procedure. Another advantage is that it can be adapted to national legal systems, and take into account specific procedures. We can provide then useful information for particular situations. Moreover, the system can evolve and be upgraded to address additional issues.

### 5.2. Specific tools for the EPO and ESCP

We have identified six general semantic tools that could be useful for solving some of the semantic problems identified so far.

A- Ontology for the identification of the Court (EPO and ESCP)

One important problem indicated by the experts is the identification of the Court. The domicile indicated in the Form should be used to determine the court that is supposed to solve the claim. We can imagine, to solve this problem, an ontology that automatically matches the domicile and the court that has to deal with the case. The ontology should have a list of the cities and their correlated court. Obviously, this could be also done with a general database or with a fixed set of rules. However, to opt for an ontology can be worthwhile in this case. The advantage of using an ontology would be that it is easier to add, modify or reuse the links between domiciles and court jurisdictions. We only need to add, delete or modify the criteria. While a list of logic inferences is better for small challenges, an ontology is more useful in the case we have to deal with different national rules of court competence attribution.

B- A FAQ for the determination of the applicable Law (EPO and ESCP)

Another problem lies in the difficulty to determine the applicable Law. This is too complex to be solved by a current semantic tool that would unsuccessfully try to substitute a lawyer. However, a list or more Frequently Asked Questions (FAQs) could be appropriate in this case. It should be oriented to be a guide for the most usual cases, and could help general users and even lawyers from other countries. As said earlier, a FAQ can include expert knowledge and can be tailored to address practical issues. In this way, we can build a general tutorship for all users requiring quick and precise indications. One of the advantages of a FAQs is that we can customize the application, so as it can assists both general users and legal experts. We can also adapt it to new situations, or set a feedback mechanism from the users and enlarge the FAQ as necessary.

C- Ontology for ESCP FORM A, number 4: Grounds for the court's Jurisdiction (EPO and ESCP)

Both EPO and ESCP clarify that the rules of Council Regulation (EC) 44/2001 on jurisdiction and the recognition and enforcement of judgements in civil and commercial matters apply. However, the application of those rules by citizens is far from being an easy matter. To solve some of these problems there are various instruments which could be taken into account. On the one hand, the European Judicial Atlas[24] provides a database of national courts. Here, the end-user may insert his or her domicile and the database would show which court could be competent. However, this database is not complete since further to big cities the database may not recognise a town that has not any court. Therefore, it could be interesting to enrich such database with an ontology matching different villages or towns with a particular court. Additionally, the European Judicial Network in civil and commercial matters has a glossary that could assist some users.[25] However, the technical terms composing the glossary makes this database primarily intended to experts. In any case, it is difficult to consider how a non-legal expert could answer correctly number 4 of Form A. Private International Law scholars keep on a discussing about the notions of the place of performance, the place of the harmful event and other various connecting factors.

| | | |
|---|---|---|
| 4. | *On what ground do you consider the court/tribunal to have jurisdiction?* | |
| 4.1. | Domicile of the defendant | ☐ |
| 4.2. | Domicile of the consumer | ☐ |
| 4.3. | Domicile of the policyholder, the insured or the beneficiary in insurance matters | ☐ |
| 4.4. | Place of performance of the obligation in question | ☐ |
| 4.5. | Place of the harmful event | ☐ |
| 4.6. | Place where the immovable property is situated | ☐ |
| 4.7. | Choice of court/tribunal agreed by the parties | |
| 4.8. | Other (please specify): ____________________ | |

[24] Available at http://ec.europa.eu/justice_home/judicialatlascivil/index_en.htm
[25] http://ec.europa.eu/civiljustice/glossary/glossary_en.htm

If a user wants to fill in the cross correctly, he needs to know that a legal contract or situation is linked to a particular legal connecting factor: domicile of the defendant, domicile of the consumer, etc. Therefore, it would be advisable to formalise the expert knowledge of a Private International Law expert in a way that most common situations can be managed. Obviously, this ontology will work better for easier cases, and could not give any worthy advice for complex ones. A disclaimer clause should also have to indicate that this tool does not pretend to substitute a lawyer, but merely to offer indications that the user may have to confirm. Additionally, expert information for foreign lawyers would be an asset. They could check whether the application of certain rules in different countries, ascertain whether there is a different procedural regulation on that issue, etc.

D- Ontology for the determination of what the claim relates to (EPO, number 6)

In case we could have a clear description of the different items of EPO's figure 6, then we can try to develop a tool to help non-experts to answer EPO's number 6. For the "additional specifications for claims relating to consumer contracts (if applicable)", we have to answer YES or NO to the following question: ""If yes, the defendant is domiciled within the meaning of Article 59 of Council Regulation (EC) No 44/2001 in the Member State where the court is seized". We could also imagine a tool that gives the correct answer when indicating the domicile.

E- Linguistic issues and semantics (EPO and ESCP)

A general legal translator is out of range. However, a list of most relevant concepts could be built and interoperate for different legal systems, with its equivalent in different countries. An ontology of legal concepts could be built up with the legal equivalent of the more used legal concepts. This semantic tool could help general users, legal advisers and even help in the translation of the facts.

F- Annotation using XML (EPO and ESCP)

One general issue with interoperability in Europe is the language problem. Some legal mandates of translation are indeed provided. However, when there is no translation, the semantic annotation of the structure of the document may be of help. Precisely, the European Judicial Atlas provides for that translation for EPO/ESCP' application forms. The Legal Aid

Application allows for similar translations.[26] According to this, XML permits annotating particular items of the application forms such as names, addresses, etc. After translation, the structure of the application form is being translated into another language. However, a further step may be to extend the XML annotation from the structure of the application form to the content of the same; that is, to the details the applicant writes. There is a general limit, in any case one that we cannot solve with this tool: we cannot translate all the description of the facts done by the parties. We can only say that a text is a list of facts, that's all. A real translator is needed in this case.

Moreover, the European Eurovocs multilingual thesaurus (compilation of comparative multilingual vocabulary) has also a XML version that could be useful.[27] However, it is mainly focussed on formal language and therefore a complement of natural language processing should be added to it.

This is useful for all the following items of FORM A:

1. *Before which court/tribunal are you making your claim?*
1.1. Name:
1.2. Street and number/PO box:
1.3. City and postal code:
1.4. Country:

2. *The claimant's details*
2.1. Surname, first name/name of company or organisation:
2.2. Street and number/PO box:
2.3. City and postal code:
2.4. Country:
2.5. Telephone (*):
2.6. E-mail (*):
2.7. Claimant's representative, if any, and contact details (*):
2.8. Other details (*):

[26] https://e-justice.europa.eu/dynform_intro_form_action.do?idTaxonomy=157&formSelectiondynform_la_a_0_action
[27] Accessible at http://eurovoc.europa.eu/drupal/

## 6. Conclusions

In the shift from the current human-readable Web to the machine-readable Semantic Web the use of ontologies, the use of knowledge representation languages and tools such as ontologies (Casellas, N. 2011) is of paramount importance. Precisely, in the legal field different efforts are made (Francesconi, E. and others. 2010) towards this end. Ontologies and FAQs can be very useful to formalise and manage with expert knowledge in a way general users or expert users can take benefit of it. Yet, ontologies may be improved a great deal. For instance, some efforts are put towards the intelligent processing of non-expert generated content. This will certainly improve the capabilities of existing tools such as in the search and retrieval area. As of today, ontologies are proposed by a community of experts that agree on the representation of a particular domain. However, non-expert content by unknown producers is produced in a distributive way; it delivers content that lacks a conceptual harmonization. Precisely, the notion of emergent semantics questions the autonomy of engineered ontologies and emphasizes the value of meaning emerging from distributed communities working collaboratively through the web (Fernández-Barrera, M. 2011). Therefore, some literature works are focusing on a way to map formal ontologies expressed in RDF or OWL with implicit ontologies emerging from user-generated content. One of the research activities consist in making compatible ontologies (top – down metadata structures) with bottom-up tagging mechanisms such as folksonomies (Fernández-Barrera, M. 2011). There are several possibilities under consideration, from transforming folksonomies into lightly formalised semantic resources to mapping folksonomy tags to the concepts and the instances of available formal ontologies.

Therefore, the approach to create a new tool would be preferable bottom-up, identifying first the problem and then trying to offer a possible solution. We also believe that it might not be necessary to substitute the whole procedure, and therefore creating an e-justice procedure. This is not the case for ESCP and EPO, where National procedures are fully preserved, and there is only a coordination and alternative procedure built upon national rules.

Semantic tools are also evolving, and can wider the range of possibilities in the near future. Indeed, IT may help and assist both parties. However, at the current state of semantic tools, IT does not fully substitute the general advice of an expert. We have to be humble and honest not to create added problems instead of solving them. Once this is settled, we can nonetheless affirm that real benefits can be obtained from semantic tools. We have mentioned some of them, and the list can grow with problems not yet detected.

References


Benjamins R and others, 'Law and the Semantic Web, an Introduction.' in Benjamins, Richard and others (eds), *Law and the Semantic Web. Legal Ontologies, Methodologies, Legal Information Retrieval, and Applications* (Springer-Verlag 2005).

Berners-Lee T, Hendler J and Lassila O, 'The Semantic Web' (May 17, 2001) Scientific American.

Casanovas P and others, 'Supporting newly-appointed judges: a legal knowledge management case study' (2005) 9 Journal of Knowledge Management 7 - 27.

Casellas Queralt N, 'Modelling Legal Knowledge through Ontologies. OPJK: the Ontology of Professional Judicial Knowledge' (Universitat Autònoma de Barcelona 2008).

Casellas N, *Legal Ontology Engineering. Methodologies, Modelling Trends, and the Ontology of Professional Judicial Knowledge* (Law, Governance and Technology Series 3, Springer 2011).

European Commission, 'Communication from the European Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions Towards Interoperability for European Public Services' (2010) COM (2010) 744 Final.

European Commission., *European Interoperability Framework for Pan-European eGovernment Services* (v12004)

Fernández-Barrera M, 'User-generated knowledge through legal ontologies: how to bring the law into the Semantic Web 2.0' (European University Institute 2011).

Francesconi E and others, 'Proceedings of the 4th Workshop on Legal Ontologies and Artificial Intelligence Techniques' (European University Institute, Fiesole, Florence, Italy, European University Institute, Fiesole, Florence, Italy, July 7th, 2010 2010).

Gruber, T.R. «Toward principles for the design of ontologies used for knowledge sharing». En: GUARINO, N. & POLI, R. (Ed.) Formal Ontology in Conceptual Analysis and Knowledge Representation. Kluwer Academic. Original paper presented at the International Workshop on Formal Ontology (1993). Available as Stanford Knowledge Systems Laboratory Report KSL-93-04. <http://ksl-web.stanford.edu/knowledge-sharing/papers/ onto-design.rtf>.

Guarino N, 'Formal Ontology and information systems.' (Proceedings of the International Conference on Formal Ontologies in Information Systmes (FOIS'98) IOS Press, Trento 6-8 June 1998 1998).

Halshofer, B.; Neufold, E. (2011). A retrospective on semantics and interoperability research. In D. Fensel (ed.) *Foundations for the Web of Information and Services: A Review of 20 Years of Semantic Web Research*. Berlin: Springer-Verlag.

Maynard D, Li Y and Peters W, 'NLP Techniques for Term Extraction and Ontology Population' (2008) IOS Press Amsterdam, The Netherlands, The Netherlands Proceeding of the 2008 Conference on Ontology Learning and Population: Bridging the Gap between Text and Knowledge.

Ng GY and Mellone M, 'The Legal Interoperability and the European Small Claims Procedure' (2011) in press.

Poblet, M.; Vallbé, J.J. (coords.). 2010. Normative framework for the border control in Spain and the EU. A deliverable for the INTEGRA Project (CDTI-Spanish Ministry of Industry, Tourism, and Commerce)

Poblet, M.; Casanovas, P.; López-Cobo, J.M. (2009). Linking the Semantic Web to ODR: the Ontomedia project. Proceedings of the Workshop on Legal and Negotiation Support Systems 2009 (LDSS 2009), in conjunction with the 12th International Conference on Artificial Intelligence and Law (ICAIL 2009): 28-38.

Poblet, M., Casellas, N., Torralba, S., Casanovas, P. 2009. Modelling Expert Knowledge in the Mediation Domain: A Mediation Core Ontology, in N. Casellas, E. Francesconi, R. Hoekstra, S. Montemagni, LOAIT 2009, Proceedings of the 3rd Workshop on Legal Ontologies and Artificial Intelligence Techniques, CEUR Workshop Proceedings, Vol. 465

Poblet, M.; Casanovas, P. (2012). Crowdsourced crisis mapping: How it works and Why it Matters. The Conversation, http://theconversation.edu.au/crowdsourced-crisis-mapping-how-it-works-and-why-it-matters-7014

Sirin E and others, 'Pellet: A practical owl-dl reasoner. Technical Report. ' (2005) University of Maryland Institute for Advanced Computer Studies 68 Sirin, E ; Parsia, B ; Grau, B C ; Kalyanpur, A ; and Katz,.

Suárez-Figueroa, M.C.; García Castro, R.; Villazón, B.; Gómez-Pérez, A. (2011) Essentials In Ontology Engineering: Methodologies, Languages, And Tools, available at http://oa.upm.es/9739/1/Essentials_In_Ontology.pdf

Warren P, Studer P and Davies J, 'Introduction' in Davies, John, Rudi Studer and Paul Warren (eds), *Semantic Web Technologies. Trends and Research in Ontology-based Systems* (John Wiley & Sons, Ltd 2005).